IntechOpen

Genetic Diversity in Plants

Edited by Mahmut Çalışkan

WEB OF SCIENCE™

GENETIC DIVERSITY IN PLANTS

Edited by **Mahmut Çalişkan**

Genetic Diversity in Plants
http://dx.doi.org/10.5772/2640
Edited by Mahmut Çalışkan

Contributors

Ibrahim Turna, Zeki Yahyaoglu, Hakan Sevik, Aydin Uzun, Turgut Yesiloglu, Olawale Aliyu, Innocenzo Muzzalupo, Adriana Ada Ceverista Chiappetta, Enzo Perri, Sónia Gomes, Henrique Guedes-Pinto, Paula Martins-Lopes, Braulio Jorge Soto-Cerda, Sylvie Cloutier, Mathilde Causse, Guillaume Bauchet, Peter Osobase Aikpokpodion, Bryan Kindiger, Kadir Ugurtan Yilmaz, Kahraman Gurcan, Jin Quan Li, Peng Zhang, Héctor Silos-Espino, Flora San Juan Hernández, Olivio Hernández, Dario Silva Bautista, Alan Roy Macias Avila, Francisco Nieto Muñoz, Luis L. Valera Montero, Silvia Flores Benites, Luis Martín Macías Valdéz, Mario Leonel Quezada Parga, Tarsicio Cororna Torres, Clebia Maria Alves De Almeida, Alexandre Silva, Maria Tereza Correia, Luciane Resende, Marcia Silva, Guo Lu Liang, Qiao He, Weixing Wang, Qi Gao Guo, Su Qiong Xiang, Xiao Lin Li, Dagmar Janovská, Petra Hlásná Čepková, Mária Džunková, Steven Rogstad, Stephan Pelikan, Melanie Simkins, Yamini Kashimshetty, Dobrinka Kostadinova Atanasova, Nikolay Tsenov, Ivan Todorov, Charity Aremu, Elena V. Artyukova, Marina Kozyrenko, Olga Koren, Alla Kholina, Olga Nakonechnaya, Yuri Zhuravlev, Tran Thanh, Lai Van Lam, Le Thi Thuy Trang, Vu Van Truong, Le Mau Tuy, Huynh Bao Lam, Rosa Paula Cuevas, Melissa Fitzgerald, Ibrokhim Y. Abdurakhmonov, Danielle Silva, Danielle Cristina Gregório Do Silva Silva, Honglang Xiao, Xiaohua Wang, Xin Zhao, Caizhi Li, Juan Ren, Fang Wang, Lei Pang

First published in Croatia, 2012 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Genetic Diversity in Plants
Edited by Mahmut Çalışkan

p. cm.

ISBN 978-953-51-0185-7

eBook (PDF) ISBN 978-953-51-4325-3

Meet the editor

Mahmut Çalışkan is a Professor of genetics and molecular biology in the Department of Biology at Mustafa Kemal University in Turkey. He obtained his BSc at Middle East Technical University in Ankara, and in the same year was appointed as Research Assistant to the Biology Department of the same University. During his MSc study he was awarded a grant from the Turkish High Education Council to continue his MSc and PhD studies abroad. He attended the University of Leeds Department of Genetics in England for his PhD study, focusing on analysis of germin gene products. After he received his PhD degree, he was appointed as Assistant Professor at Mustafa Kemal University, Hatay. In 2000 he obtained the title of Associate Professor in genetics from the Turkish High Education Council. Following five years of extensive research and lecturing he was appointed as Professor of Molecular Biology in 2006. Prof. Çalışkan was awarded for his postdoctoral research in England with grants from NATO and University. In addition to acting as a reviewer and advisor to numerous journals and grant awarding institutions, he has been involved in the establishment of the Biology Department at Mustafa Kemal University.

Contents

Preface

Genetic diversity is the fundamental source of biodiversity. In 1989, the World Wildlife Fund defined biodiversity as "the richness of life on Earth – millions of plants, animals and microorganisms, including the genes which they carry, and complex ecosystems that create the environment". Currently, the issue of maintaining genetic diversity as a component of the conservation of biodiversity has been accepted at an international level. FAO has included the issue of conservation, evaluation, and the use of animal genetic resources in its fields of interest since the 1970s. In this context, one of the main concerns of scientific research activities is conserving the genetic diversity of local breeds, especially those of economic interest. Genetic diversity among individuals reflects the presence of different alleles in the gene pool, and hence, different genotypes within populations. Genetic diversity should be distinguished from genetic variability, which describes the tendency of genetic traits found within populations to vary. There is a considerable genetic variability within or between natural populations. Population geneticists attempt to determine the extent of this variability by identifying the alleles at each locus and measuring their respective frequencies. This variability provides a genomic flexibility that can be used as a raw material for adaptation. On the other hand, one of the consequences of low genetic variability could be the inability to cope with abiotic and biotic stresses. From the growing knowledge on the genome sequences of organisms it becomes evident that all forms of diversity have their origin at genetic level. In this context, genetic diversity analysis provides vital and powerful data that helps for a better understanding of genetic variation and improved conservation strategies.

Plants are a distinct kingdom of organisms that possess unique properties of reproduction, development, physiology, and adaptation. Plant diversity refers to the variety of plants that exist on the Earth. Plants, in order to survive, have to compete with other plants and organisms in an ecosystem. Over time, they have developed various characteristics to help them survive, which leads to plant diversity. It is essential to have regular assessments of the conservation status of all plant species, in order to prioritize those in need of conservation action and to provide a measure of the success of actions being taken. The improvement of cultivated plants considerably depends on the extent of genetic variability available within the species. The genetic variation that exists among plant populations is a basic requirement for efficient

development and improvement of such populations. It also indicates whether a population can withstand and live with changes in the environment, which are mostly altered in an unpredictable way. Molecular studies shed light on relationships and diversity among plant breeds. The extent and nature of genetic diversity of plants from all around the world has been investigated by typing DNA markers in a set of individuals belonging to several breeds. The most useful techniques for these studies have been micro-satellites, RFLP, RAPD, SSR, AFLP, SCAR and ISSR.

The purpose of *Plant Genetic Diversity* is to provide a glimpse into the dynamic process of genetic variation by presenting the thoughts of some of the scientists who are engaged in the development of new tools and ideas used to reveal genetic variation, often from very different perspectives. I would like to express my deepest gratitude to all the Authors who contributed to this book by sharing their valuable works with us. This book should prove useful to students, researchers, and experts in the area of conservation biology, genetic diversity and molecular biology. The year 2010 has been celebrated as the international year of biodiversity by the United Nations and it has been a unique opportunity to realize the vital role that biodiversity plays in sustaining the life on Earth. Let us all wish much success to all projects and initiatives dealing with the conservation of diversity of life because rich genetic resources are a prerequisite for future generations to be able to breed crop varieties and face new challenges.

Prof. Dr. Mahmut Caliskan
Mustafa Kemal University, Department of Biology
Hatay, Turkey

Part 1

Molecular Approaches for the Assesment of Genetic Diversity

1

Isolation of High-Quality DNA from a Desert Plant *Reaumuria soongorica*

Xiaohua Wang[1], Honglang Xiao[1], Xin Zhao[2], Caizhi Li[1], Juan Ren[1], Fang Wang[1] and Lei Pang[1]
[1]*Key Laboratory of Ecohydrology and of Inland River Basin Cold and Arid Regions Environmental and Engineering Research Institute Chinese Academy of Sciences, Lanzhou*
[2]*Extreme Stress Resistance and Biotechnology Laboratory Cold and Arid Regions Environmental and Engineering Research Institute Chinese Academy of Sciences, Lanzhou PR China*

1. Introduction

The desert plants, widely distributed in arid and semi-arid regions, have the significant ecological functions of combating desertification and maintenance of the ecosystem stability in the desert ecological system, and also have an irreplaceable role in restoration and reconstruction of vegetation in the sand. With the development of modern molecular biology, molecular markers analysis such as amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), simple sequence repeat (ISSR), and especially full genome sequencing technique are widely used in the study of biological diversity, genetic transformation and the conservation of genetic resources of desert plants. The isolation of high molecular weight and high purity genomic DNA is a pre-requisite for these molecular techniques. Now there are many extraction methods of genomic DNA from biomaterials. The commonly employed DNA isolation methods involve the use of enzymes such as lysozyme and proteinase K (Lockhart et al., 1989), cetytrime-thylammonium bromice (CTAB) (Ghosh et al., 2009; Moyo et al., 2008; Khanuja et al., 1999; Novaes et al., 2009; Singh et al., 2000) or sodium dodecyl sulfate (SDS) (Kaufman et al., 1999; Dellaporta et al., 1983) treatment and extraction with organic solvents, detergent-induced lysis in conjunction with proteinase K and lysozyme (Perera et al., 1994) or lysis using guanidinium isothiocyanate (GITC)-containing solutions (Boom et al., 1990; Noordhoek et al., 1995; Chakravorty & Tyagi, 2001), among which, guanidinium thiocyanate has been shown to be a powerful agent in the purification of DNA because of its potential to lyse cells and its potential to inactivate nuclease (Boom et al., 1990; Chomczynski et al., 1987; Zeillinger et al., 1993). However, high amounts of gummy polysaccharides, polyphenols and other various secondary metabolites such as alkaloids, flavonoids, terpenes and tannins in the desert plants usually hamper the DNA isolation procedures and reactions such as DNA restriction, amplification and cloning (Moyo et al., 2008; Khanuja et al., 1999; Pang et al., 2011; Zhang K., 2011; Ji & Li, 2011). The main problems encountered in the isolation and purification of high

molecular weight DNA from plant species include degradation of DNA due to endonucleases and high levels of contaminants (polyphenols or polysaccharides) that co-precipitate with DNA. Endonucleases released from the vacuoles during the cell lysis process, which are co-isolated with highly viscous polysaccharide, lead to the degradation of DNA and remarkably reduce the yield of extracted DNA (Khanuja et al., 1999. Polyphenols released from the vacuoles during the cell lysis process are oxidized by cellular oxidases and undergo irreversible interactions with nucleic acids causing browning of the DNA (Varma et al., 2007; Moyo et al., 2008; Khanuja et al., 1999; Porebski et al., 1997). The presence of gelling polysaccharides prevents complete dissolution of nucleic acids and imparts a viscous constituency to the DNA making it stick to the wells during gel electrophoresis (Barnell et al., 1998; Diadema et al., 2003; Varma et al., 2007). Furthermore, inhibitor compounds like residual polyphenols, polysaccharides and other secondary metabolites inhibit enzymatic reactions such as restriction endonuclease cleavage (Raina and Chandlee, 1996) or Taq DNA polymerase amplifications (Shioda and Murakami-Murofushi, 1987; Tigst and Adams, 1992; Pandey et al., 1996) or ligase links (Moyo et al., 2008; Khanuja et al., 1999, Weishing et al., 1995). Thus, though several successful genomic DNA isolation protocols for high polyphenol and polysaccharide containing plant species have been developed, none of these are universally applicable to all plants (Varma et al., 2007), because qualitative and quantitative differences in the levels of polysaccharides, phenols and secondary metabolites in various plant tissues significantly alter the efficiency of nucleic acid extraction and purification procedures. Therefore researchers often modify a protocol or blend two or more different procedures to obtain DNA of the desired quality (Varma et al., 2007).

Reaumuria soongorica (Pall.) Maxim, an extreme xeric semi-shrub of Tamaricaceae, is a constructive and dominant species of desert shrub vegetation (Liu et al. 1982; Wang et al., 2011; Bai et al., 2008). It is distributed widely on a large area of sand wasteland (Fig. 1a) and saline land (Fig. 1b) in arid and semiarid regions of central Asia from the western Erdos, Alaskans, Hexi Corridor, Qaidam Basin to Tarim Basin and Jungar Basin (from the east to the west) and forms the vast and distinctive landscape of the salt desert (Liu and Liu, 1996). The distribution of *R. soongorica* in desert in northwestern China is shown in (Fig. 2) (Ma et al., 2007). It can inhabit on the alluvial plains of piedmont, hilly lands, eroded monadnocks, piedmont gravel mass, gravel alluvial fan and the Gobi. It is distributed on large span, wide range, and complex habitat where there are different climatic conditions among regions, especially with significantly different water conditions, such as the average annual rainfall in Lanzhou with 327.7 mm, Shapotou with 188.2 mm, and Ejina with 35.1 mm as it possesses the characteristics of drought resistance, salt tolerance, barrenness tolerance, and dune fixation. It is such a good candidate of desert plants that it is very significant for us to study its biological diversity and the mechanism of adverse environments resistance. However, the leaves of *R. soongorica* are evolved into the form of pellets suitable for arid environment, which are very hard in texture and contain high level of polysaccharides, polyphenols and secondary metabolites that co-precipitate with DNA, making DNA isolation difficult.

A good isolation protocol should be simple, rapid and efficient, yielding appreciable levels of high quality DNA suitable for molecular analysis. Krizman et al. (2006) were of the opinion that, among other factors, the amount of plant sample extracted could be critical in keeping an extraction procedure robust. In the present study, our objective was to create an improved DNA extraction procedure amenable for the isolation of high quality DNA in the

desert plant *R. soongorica*. Four methods for extracting DNA were tested in this study and they included the TianGen Plant Genomic DNA Kit, the modified TianGen Plant Genomic DNA Kit, the modified CTAB-A method and the modified CTAB-B method herein promoted by us. The results showed that the modified CTAB-B method was a relatively quick and inexpensive method and it was the best method for extraction DNA from leaf materials containing large quantities of secondary metabolites. Furthermore, it was further tested that the modified CTAB-B method for isolating DNA from leaves of *R. soongorica* yields DNA in a quantity and quality suitable for PCR amplification, DNA marker analysis and restriction digestion.

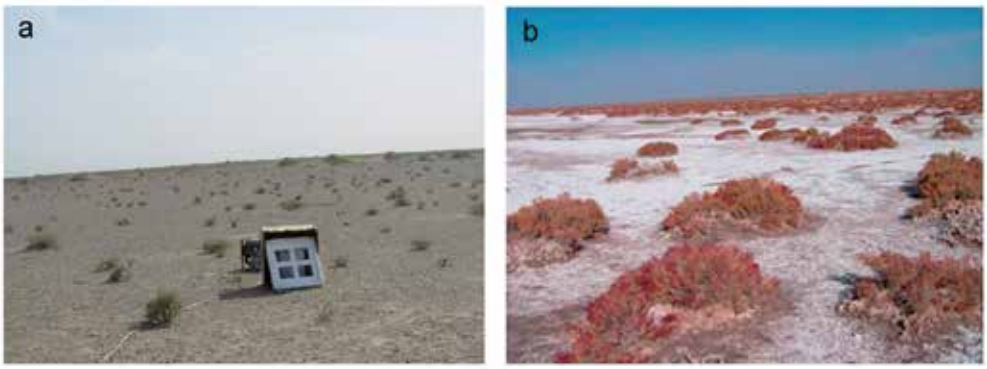


Fig. 1. The natural habitats of *Reaumuria soongorica*. a Populations of *R. soongorica* in sand wasteland or Go; b Populations of *R. soongorica* in saline land with a white visible salt on the ground

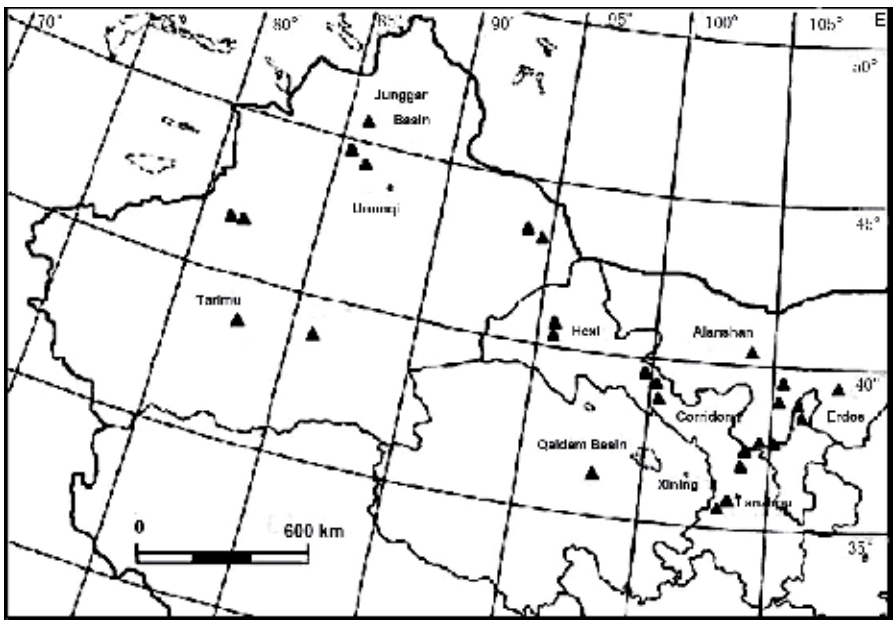


Fig. 2. Distribution map of the study plant *Reaumuria soongorica* in northwest China (the triangle symbol indicates the major distribution area)

2. Materials and methods

2.1 Plant materials

Tender *R. soongorica* leaves were collected from Ejina in Mogo, China and snap-frozen in liquid nitrogen. The frozen leaves were transported in liquid nitrogen and stored at -80ºC upon reaching the laboratory.

2.2 Equipments and solution preparation

Mortars, pestles, glassware and plasticware was autoclaved prior to use. The CTAB extraction buffer was composed of 2.0% CTAB (High Purity grade, Amresco), 100 mM Tris-HCl (pH 8.0)) (Ultra pure grade, Amresco), 2 M NaCl (Biotechnology grade, Amresco), 25 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0) (High purity Grade Amresco) in 0.1% ultra pure water. The components in the extraction buffer were mixed and autoclaved. The 5% PVPP (Sigma P-6755) was added when the material was grounded and the 5% beta-mercaptoethanol (Biotechnology grade, Amresco) was added before DNA extraction. The final solution was warmed in a water bath to 65°C for use in DNA extraction. TE buffer was prepared with 10 mM Tris-HCl (pH 8,0) and 1.0 mM EDTA (pH 8.0). A phenol (pH 5.0)/chloroform /isoamyl alcohol mixture (25/24/1) (Biotechnology grade, Amresco) and a chloroform /isoamyl alcohol mixture (24/1) were prepared before use, and all other solutions including 3 M sodium acetate (NaAc) (pH 5.2) (Biotechnology grade, Amresco), 1 M NaCl, pre-cooled 75% ethanol were prepared with ultra pure water and autoclaved. DNase-Free RNase (Ultra pure grade) was purchased from Amresco Corporation.

2.3 Grinding

The frozen fresh samples were transferred into a mortar with liquid nitrogen and the ceramic pestle was pre-chilled for grinding. By providing liquid nitrogen as a cooling jacket, the samples were sprinkled with PVPP and ground vigorously to fine powder using the ceramic pestle. This powder was used for the following extraction protocols of the CTAB methods. And the commercial DNA isolation kit was ground in liquid nitrogen free of adding to PVPP.

2.4 DNA extraction

For our modified CTAB method, the steps of this protocol was carried out as follows:

1. The ground powder sample (100 mg) was transferred to 2 ml micro-centrifuge tubes filled with 700 µl of pre-warmed CTAB extraction buffer containing 35 µl β-mercaptoethanol, following by incubation at 65°C for 30 min in a warm water bath. The mixture was regularly mixed three to four times by gently inversion during the incubation..
2. 200 µl 3 mol/L sodium acetate (NaAc) (pH 5.2) was added to the incubated mixture and mixed gently by inversion and incubated on ice for 30 min.
3. An equal volume of chloroform/isoamyl alcohol (24/1) was added to the homogenate and mixed thoroughly for 2 min, following by centrifugation at 12,000×g for 10 min at

room temperature. The upper aqueous phase was carefully collected from each sample without disturbing the interface. This step was repeated twice.

4. 2.5 volumes of absolute ethanol was added to the recovered supernatant and precipitated 30 min at -20ºC. A precipitate formed at this stage and the mixture was centrifuged at 12,000×g for 10 min at 4°C. DNA pellet was recovered by decanting the supernatant.
5. The crude nucleic pellet was dissolved in 1 ml of 1 M NaCl instead of dissolving it in Tris-EDTA (TE) buffer. The entire solution was transferred to a 2 ml microcentrifuge tube and treated with RNase at 37°C for 1 h. RNase contamination was removed by adding an equal volume of phenol (pH 5.0)/chloroform /isoamyl alcohol (25/24/1) and the aqueous phase was collected in a fresh microcentrifuge tube after centrifugation at 12,000×g for 5 min at room temperature.
6. An equal volume of chloroform/isoamyl alcohol (24/1) was added and mixed thoroughly. The samples were centrifuged at 14,000×*g* for 5 min at room temperature and the top aqueous phase was transferred to a fresh tube.
7. A double volume of absolute ethanol and 0.1 volumes of 3 M (pH 5.2) sodium acetate were added into the collected aqueous phase and were mixed gently by inversion. The samples were incubated at -80ºC for 30 min, followed by centrifugation at 12,000×g.
8. The DNA pellet was washed with 75% ethanol, absolute ethanol, air-dried and finally the purified DNA pellet was dissolved in 100 µl of TE buffer. and stored at -20°C.

Initial tests for DNA isolation from the leaves of *R. soongorica* were carried out with the modified CTAB-A method and the Plant Genomic DNA Kit (TIANGEN Biotech Co., Ltd., Beijing). The modified CTAB-A method was modified based on the classical Doyle and Doyle (1987) method. The steps of the modified CTAB-A are similar to those of the CTAB-B method before step 4 (the precipitate of crude nucleic pellet). The main difference is that the crude nucleic pellet in CTAB-A method was solved in TE and extracted by chloroform/isoamyl alcohol (24/1) again instead of being treated with DNase-free RNase. Briefly, the crude nuclei pellet was dissolved in 500 µl of TE buffer, followed by the steps 3 and 4 of the CTAB-B method repeatedly. The protocols for the commercial DNA isolation kit was performed according to the manufacturer' procedures on their website:http://www.tiangen.com/newEbiz1/EbizPortalFG/portal/html/ProductInfoExhibit.html?ProductInfoExhibit_ProductID=c373e923ec4bc4d68f7efc2e13bcb309&ProductInfoExhibit_isRefreshParent=false. The protocol of the modified TianGen Plant Genomic DNA Kit was based on those of the TianGen Plant Genomic DNA Kit with some slight modifications. The modifications were listed as follows: (1) The plant materials were ground free of liquid nitrogen, but were added to the cooled sterile mortar and ground with Gp1 buffer poured into the mortar. (2) The ground tissue was transferred to 2 ml micro-centrifuge tubes prepared a warm (65°C) Gp1extraction buffer, and then the 5% beta-mercaptoethanol and 10 µl DNase-free RNase were added to the mixture immediately and mixed gently by inversion. The other steps are carried out by the instructions of the kit. For each method, three independent experiments were done, and three samples were prepared in each independent experiment.

2.5 Testing the quality of the genomic DNA

Three microliters of each genomic DNA sample is examined by electrophoresis and remnant DNA sample is stored at –70°C. Mixing 1 µl of 5× DNA loading buffer (TIANGEN Biotech

(Beijing) Co. Ltd.) with 3 µl of genomic DNA at room temperature for 1 min. Then the sample was loaded on 0.8% agarose formaldehyde denaturing gels stained with ethidium bromide (EtBr) (Biotechnology grade, Amresco), and run on gels in the 1× formaldehyde electrophoresis buffer at 5-7 V/cm.

2.6 Assessment of the purity and the yield of the genomic DNA

Two microliters of each genomic DNA sample was diluted into 200 µl of sterilized ultra pure water (pH 7.0). The absorbance of each diluted genomic DNA sample was evaluated at 260 and 280 nm using a ND-2000C (Thermo, America). The yield of genomic DNA was calculated according to the formula: DNA yield = 50 × OD_{260} × dilution factor × volume of sample in milliliters/material weight (g). Measured the values at the wavelengths of 260, 280 nm and 230 nm and calculated the ratios of A260/A280 and the A260/A230.

2.7 ISSR amplification

The DNAs isolation from different *R. soongorica* populations by our promoted CTAB-B protocol were used as template for inter simple repeat sequence primers (ISSR) amplification (Gajera et al., 2010). ISSR amplification reactions were performed in 20 µl reaction volume containing 1 µl gDNA template, 0.25 mmol/L of each dNTPs, 2.5 mmol/L $MgCl_2$, 1×PCR buffer (10mmol/L Tris-HCl pH8.3, 5mmol/L KCl), 1 U Taq DNA polymerase and 0.5 mmol/L of UBC-807 primer (AGA GAG AGA GAG AGA GT). The amplification reaction were carried out on a thermocycler (Biometra) and programmed for an initial pre-denaturing at 94°C for 5 min, followed by 35 cycles of 1 min at 94°C (denaturation), 1 min at 48°C (annealing temperature), and 1.5 min at 72°C (extension) followed by a final extension step at 72°C for 10 min. Amplification products (5 µl) were electrophoresed in 1.5% agarose in 1× TBE buffer and stained with ethidium bromide.

3. Results and discussion

3.1 DNA isolation methodology

Commercial DNA isolation kits are widely used for their single-step methods and the relatively short amount of time required (usually about 1-2 h). These kits have also proven effective for isolating DNA from common plants such as rice, barley and Arabidopsis. We first attempted to isolate DNA from the leaves of *R. soongorica* using three commercial DNA isolation kit: the Plant Genomic DNA Kit (TIANGEN Biotech Co., Ltd., Beijing) which is designed specifically to extract DNA from plant tissues rich in secondary metabolites. In our hands, this kit was not able to isolate any DNA from the leaves of *R. soongorica* (data not shown). Then, we carried out the improved kit method according to the suggestion of this company' technical assistance employee, which yielded a small amount of DNA, but it was seriously contaminated (Fig. 3, lanes 1,2). The failure of the kit may be explained by the DNA likely formed a sticky, a glue-like gel in complex with these secondary metabolites and this could not be properly separated into two phases by centrifugation.

During the course of the RNA isolation, none of the kits were able to isolate any RNA from the leaves of *R. soongorica* (Wang et al. 2011), so we did not try more other commercial DNA kits to isolate DNA, but attempted to use a improved CTAB method to extract DNA from

the leaves of *R. soongorica*. When we carried out the modified CTAB-A method based on the classical Doyle and Doyle (1987) method, which consistently resulted in significant RNA contamination of the DNA samples (Fig. 3, lanes 3,4). To remove RNA contaminants, additional purification steps must be performed, which not only reduce DNA yield but also increase the time required for DNA extraction.

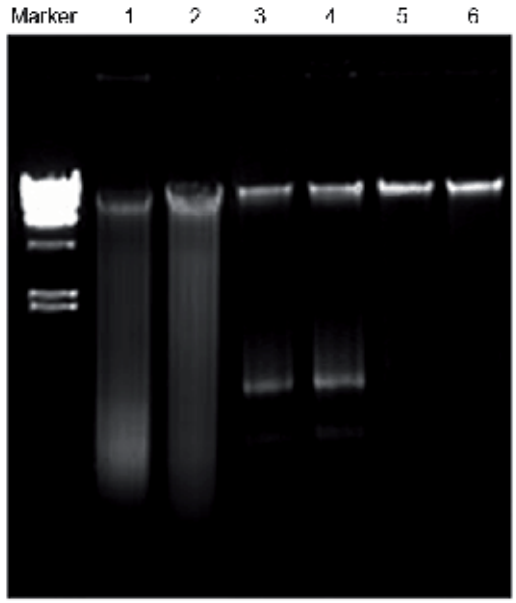


Fig. 3. Electrophoretic analysis of *R. soongorica* genomic DNA isolated using various extraction methods. The genomic DNA (3 μl) of each sample was loaded into the different well, and then was run on a 0.8% agarose gel stained with ethidium bromide in 1×TAE buffer. Marker indicates the Molecular weight marker – 1 kb DNA ladder. Lanes 1 and 2, the modified TianGen Plant Genomic DNA Kit; lanes 3 and 4, the modified CTAB-A method; lanes 5 and 6, our promoted CTAB-B method. As the TianGen Plant Genomic DNA Kit was able to isolate any DNA from the leaves, the photos were not shown in Fig. 3.

The present study was motivated by the need for better methods of extracting sufficient quantities of high-quality DNA from plant tissue rich in secondary metabolites for use in molecular marker assays. The promoted CTAB-B protocol described here efficiently eliminates most of the interfering molecules, including polyphenols, polysaccharides, and proteins, and it yields translucent and water-soluble DNA pellets without RNA contamination. The main protocols made in this method were grounding PVPP together with the plant material, an increase in the volume of high salt extraction buffer, adding 3 mol/L sodium acetate (NaAc) in extraction buffer, dissolving the crude pellet in 1 M NaCl followed by RNase treatment, the purification of acid phenol extraction (phenol: chloroform: isoamylal alcohol (PCI)=25:24:1) and the use of pre-cooled ethanol and sodium acetate in precipitation and all these modifications helped to remove the interference of secondary metabolites in the DNA isolation. PVPP was sprinkled directly onto the frozen fresh leaf tissue in the mortar and vigorously ground with the leaf tissue in the presence of liquid nitrogen, which can avoid the oxidation of released polyphenols into

quinines, which in turn bind to nucleic acids and hinder the isolation of high quality DNA. Increase in the volume of extraction buffer can completely break down the cell walls and make more nucleotide acid released resulted in increasing the yield of DNA isolation. Krizman et al. (2006) postulated that the plant tissue amount per volume of extraction buffer has an effect on DNA quality and yield. Since the extraction buffer is responsible for the lysis of membranes and liberation of DNA from cellular organelles (Weising et al., 2005), the smaller the quantity of plant tissue per unit volume, the more optimal the lysis process. Striking the correct balance between plant tissue amount and extraction buffer volume would reduce the probability of co-precipitation of contaminants with the DNA pellet as the saturation concentration during precipitation is less likely reached or exceeded (Krizman et al., 2006). During the extraction, 3 mol/L sodium acetate (NaAc) added combined with chloroform/isoamyl alcohol extraction can reduce markedly the co-precipitation of polysaccharides with the nucleic acids and remove most proteins, polysaccharides, polyphenols and other impurities for the first time. The crude nucleic pellet was dissolved in 1 ml of 1 M NaCl instead of dissolving it in Tris-EDTA (TE), which ensured further reduction of viscosity of the mucilaginous substances (Chen and Chen, 2004; Ghosh et al., 2009). DNase-free RNase was added to crude DNA samples dissolved in1 ml of 1 M NaCl to completely clear residual RNA. After RNase treatment, the DNA solution requires purification with an acid-phenol: chloroform : isoamyl alcohol (25: 24: 1) extraction because small amounts of protein in DNA pellets and salts in the RNase reaction buffer and stop solution both influence downstream molecular procedures such as restriction endonuclease digestion, ISSR-PCR amplification and full the genomic sequencing. Thus, we used acid-phenol to remove residual protein and the remaining salts after the RNase treatment. Finally, it is necessary to precipitate DNA simultaneously with sodium acetate (pH 5.2) and absolute ethanol which can completely remove the residual polysaccharides from the DNA sample resulted in increasing the yield of DNA isolated.

3.2 Assessment of the quality and quantity of the total DNA

The success of an DNA isolation protocol may be judged by the quality and quantity of DNA recovered. The quality of DNA was assessed by gel electrphoresis, spectrophotometry, restriction endonuclease digestion and PCR amplification. The mean yield of DNA extracted by our promoted method was approximately 60.29 ± 20.16 μg/100mg of fresh leaves, which was higher than that of the modified CTAB-A method (35.72 ± 15.41 μg/100mg) and the modified TianGen Plant Genomic DNA Kit (20.54 ± 8.43 μg/100mg) (Table 1). The DNA isolated by our promoted method also exhibited good purity. DNA absorbs UV light maximally at 260 nm, whereas protein absorbs at 280 nm and other contaminants including carbohydrates, phenol, and aromatic compounds generally absorb around 230 nm. Therefore, the A260/A280 and the A260/A230 ratios are often used as indicators of DNA sample purity. Generally, ratio values of A260/A280 in the range of 1.8 – 2.0 indicate high-purity DNA; the ratio values of A260/A280 less than 1.8 indicate protein contamination in DNA samples; the ratio values of A260/A280 more than 2.0 indicate much RNA or many DNA fragments in DNA samples. With our method, the A260/A280 and A260/A230 ratios were 1.86 ± 0.16 and 1.92 ± 0.13, respectively, indicating that the DNA was free of protein and polysaccharides/polyphenol contamination (Table 1). In addition, there were no other bands visible in the bands (Fig. 1, lanes 5,6), indicating that the DNA was free of genomic RNA contamination. In contrast, the DNA from the CTAB-A method had poor purity as assessed by A260/A280 (2.12 ± 0.18) and the A260/A230 ratios (1.55 ± 0.36 μg/100mg), and

the yield was lower than that obtained with our promoted CTB-B method and the DNA was badly contaminated with RNA because there were two other visible bands on the lane (Fig. 1, lanes 3,4; Table 1). The TianGen Plant Genomic DNA Kit was not able to isolate any DNA from the leaves of *R. soongorica*. Furthermore, even the improved kit method only yielded a small amount of DNA (20.54 ± 8.43 μg/100mg), and it was seriously contaminated (Fig. 1, lanes 1,2). The above results show that the improved CTAB-B protocol described herein efficiently eliminates most of the interfering molecules (including polyphenols, polysaccharides, proteins and salts), and it also provides a higher yield of DNA pellets that are translucent, water-soluble and lack RNA contamination, indicating that it is superior to the CTAB-A method and the commercial kits.

Method	Purity		Yield
	$A_{260/280}$	$A_{260/230}$	(μg/100mg)
Modified TianGen DNA kit	1.31 ± 0.15	0.87 ± 0.32	20.54 ± 8.43
Modified CTAB-A	2.12 ± 0.18	1.55 ± 0.36	35.72 ± 15.41
Modified CTAB-B (promoted by us)	1.86 ± 0.16	1.92 ± 0.13	60.29 ± 20.16

Table 1. The genomic DNA purity and yield in *R. soongorica* leaves by different methods

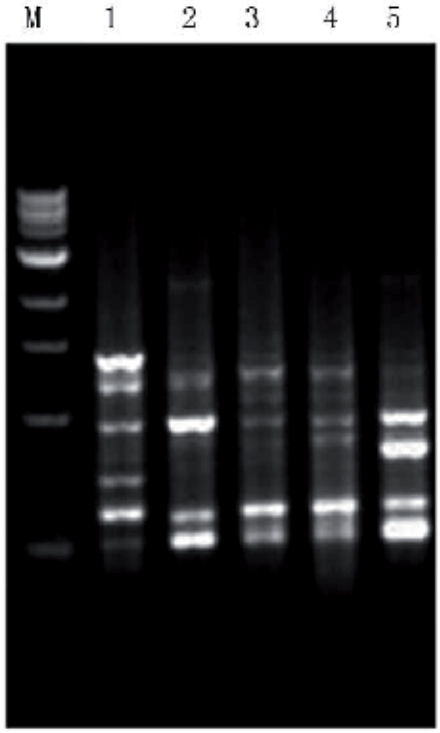


Fig. 4. ISSR-PCR profiles of the genomic DNAs isolation from different *R. soongorica* populations by our promoted CTAB-B protocol using the UBC-807 primer were analyzed on a 1.5% agarose gel stained with ethidium bromide in 1×TAE buffer. "M" represents the Molecular weight marker – DNA marker DL2000.

The suitability of extracted DNA for downstream molecular processes was further verified by molecular markers ISSR-PCR amplification. AS shown in Fig. 4, the genomic DNA of five different *R. soongorica* populations were highly amplifiable by ISSR-PCR as indicated by the amplification products resolved on 1.5% agarose gel. This further confirmed the purity of DNA, free of polysaccharide and polyphenol contamination, which would otherwise inhibit Taq DNA polymerase and restriction endonucleases (Ahmad et al., 2004). Plant molecular applications such as RAPD and AFLP necessitate the successful isolation of high quality DNA (Michiels et al., 2003; Ahmad et al., 2004), devoid of contaminants. Without high quality DNA such downstream molecular manipulations are not feasible (Varma et al., 2007). To confirm the applicability of our method, this DNA extraction method has also been found to be efficient in other desert plants, including *Tamarix ramosissima*, *Nitraria tangutorum* and *Caragana korshinskii* Kom. (data not shown).

4. Conclusion

Our results showed that the modified CTAB-B method promoted here was of high quality, purity and yield and was suitable for downstream molecular assays. Based on a CTAB method, the protocol has been improved as follows: the more volume of extraction buffer was used to completely break down the cell walls; the samples were ground with PVPP to effectively inhibit the oxidation of phenolics; during the extraction, 3 mol/L sodium acetate (NaAc) was added to reduce markedly the co-precipitation of polysaccharides with the nucleic acids; contaminating RNA was removed with RNase I; acid phenol extraction (phenol: chloroform: isoamylal alcohol (PCI)=25:24:1) was used to effectively removes the residual proteins and inhibitors in the RNase reagent. Thus, despite the high levels of secondary metabolites in the leaves of *R soongorica*, the high quality DNA is isolated from the nuclei without interference. Moreover, the new protocol is also suitable for isolating genomic DNAs from other desert plant species and tissues that are rich in secondary metabolites.

5. Acknowledgments

This study was supported by the National Key Technologies R&D Program: The Integration and Experiment Demonstration of Ecological Restoration Technology in Loess Hills-Sand Areas (2011BAC07B05) and by the Ministry of Forestry Commonweal Special Project: Monitoring and Assessment Technologies for Hydrological Regulation Function of Desert Ecosystem (201004010-05).

6. References

Ahmad S.M., Ganaie M.M., Qazi P.H., Verma V., Basir S.F., Qazi G.N. 2004. Rapid DNA isolation protocol for angiospermic plants. Bulgarian Journal of Plant Physiology 30, 25–33.

Bai J., Xu D.H., Kang H.M., Chen K., Wang G. 2008. Photoprotective function of photorespiration in *Reaumuria soongorica* during different levels of drought stress in natural high irradiance. Photosynthetica 46, 232–237.

Barnell P., Blanchard A.N., Bryant J.A., Smirnoff N., Weir A.F. 1998. Isolation of DNA from the highly mucilaginous succulent plant *Sedum telephium*. Plant Molecular Biology Reporter 16, 133–138.

Boom R., Sol C.J., Salimans M.M., Jansen C.L., Wertheim-van Dillen P.M., van der Noordaa J. 1990. Rapid and simple method for purification of nucleic acids. Journal of Clinical Microbiology, 28, 495–503.

Chakravorty S. & Tyagi J.S. 2001. Novel use of guanidinium isothiocyanate in the isolation of Mycobacterium tuberculosis DNA from clinical material. FEMS Microbiology Letters 205, 113–117.

Chen R.H., Chen W.Y., 2004. Rheological properties of the water-soluble mucilage of a green laver, Monostroma nitidium. Journal of Applied Phycology 13, 481–488.

Chomczynski P. & Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162, 156–159.

Dellaporta S.L., Wood J., Hicks J.B. 1983. A plant DNA mini-preparation: Version 2. Plant Molecular Biology Reports 1, 19–22.

Diadema K., Baumel A., Lebris M., Affre L. 2003. Genomic DNA isolation and amplification from callus culture in succulent plants, *Carpobrotus* species (Aizoaceae). Plant Molecular Biology Reporter 21, 173a–173e.

Doyle J.J. & Doyle J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19, 11–15.

Gajera B.B., Kumar N., Singh A.S., Punvar B.S., Ravikiran R., Subhash N., Jadeja G.C. 2010. Assessment of genetic diversity in castor (*Ricinus communis* L.) using RAPD and ISSR markers. Industrial Crops and Products 32, 491–498.

Ghosh R., Paul S., Ghosh S.K., Roy A. 2009. An improved method of DNA isolation suitable for PCR-based detection of begomoviruses from jute and other mucilaginous plants. Journal of Virological Methods 159, 34–39.

Ji R.M.T. & Li J.Q. 2011. The comparative study of different methods for total DNA extraction in Tibet Planteau *Potentilla anserine*. Northern Horticulture 6, 152–154.

Kaufman B., Richards S., Dierig D. A. 1999. DNA isolation method for high polysaccharide *Lesquerella* species. Industrial Crops and Products 9, 111–114.

Khanuja S.P.S., Shasany A.K., Darokar M.P., Kumar S. 1999. Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. Plant Molecular Biology Reporter 17, 1–7.

Križman M., Jakše J., Baričevič D., Javornik B., Prošek M. 2006. Robust CTAB-activated charcoal protocol for plant DNA extraction. Acta Agriculturae Slovenica 87, 427–433.

Liu J.Q., Qiu M.X., Pu J.C., Lu Z.M. 1982. The typical extreme xerophyte *Reaumuria soongorica* in the desert of China. *Acta Botanic Sinica, 24*, 485−488.

Liu S. & Liu X.Y. 1996. The estimating model of upper plant weight on *Reaumuria soongorica* semishrub. Arid Zone Research 13, 36–41.

Lockhart P.J., Larkum A.W.D., Beanland T.J., Howe C.J. 1989. The isolation of high molecular mass DNA from the prochlorophyte *Prochloron didemni*. FEBS Letters 248, 127–130.

Ma J.Y., Chen F.H., Zhang H.W., Xia D.S. 2007. Spatial distribution characteristics of stable carbon isotope compositions in desert plant Reaumuria soongorica. Frontiers of Earth Science in China 1, 150–156.

Michiels A., Van den Ende W., Tucker M., Van Riet L., Van Laere A. 2003. Extraction of high-quality genomic DNA from latex-containing plants. Analytical Biochemistry 315, 85–89.

Moyo M., Amoo S.O., Bairu M.W., Finnie J.F., Van Staden J. 2008. Optimising DNA isolation for medicinal plants. South African Journal of Botany 74, 771–775.

Noordhoek G.T., Kaan J.A., Mulder S., Wilke H., Kolk A.H.J. 1995. Routine application of the polymerase chain reaction for detection of Mycobacterium tuberculosis in clinical samples. Journal of Clinical Pathology 48, 810–814.

Novaes R.M.L., Rodrigues J.G., Lovato M.B. 2009. An efficient protocol for tissue sampling and DNA isolation from the stem bark of Leguminosae trees. Genetics and Molecular Research 8, 86–89.

Pandey R.N., Adams R.P., Flournoy L.E. 1996. Inhibition of random amplified polymorphic DNAs (RAPDs) by plant polysaccharides. Plant Molecular Biology Reporter 14, 17–22.

Pang Y.W., Duan L.Z., Song D.Y., Ren Z.M. 2011. A modified method of genomic DNA extraction and establishment of AFLP systems in *Rhus chinensis*. Guihaia 31, 36–38.

Perera J., Chandrasekharan N.V. and Karunanayake E.H. 1994. PCR based detection of Mycobacterium tuberculosis : effect of sample preparation. Southeast Asian Journal Tropical Medical & Public Health 25, 693–697.

Porebski S., Bailey L.G. and Baum B.R. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter 15, 8–15.

Raina K. & Chandlee J.M. 1996. Recovery of genomic DNA from fungus (*Sclerotinia homoeocarpa*) with high polysaccharide content. Biology Techniques 21, 1030–1032.

Shioda M., Murakami-Murofushi K. 1987. Selective inhibition of DNA polymerase by a polysaccharide purified from slime of *Physarum polycephalum*. Biochemical and Biophysical Research Communications 146, 61–66.

Singh K.K., Muralidhar M., Kumar A., Chattopadhyaya T.K., Kapila K., Singh M.K., Sharma S.K., Jain N.K., Tyagi J.S. 2000. Comparison of in house polymerase chain reaction with conventional techniques for the detection of *Mycobacterium tuberculosis* DNA in granulomatous lymphadenopathy. Journal of Clinical Pathology 53, 355–361.

Tigst D. & Adams R.P. 1992. The effect of plant polysaccharides and buffer additives on PCR. Biology Techniques 12, 332–334.

Varma A., Padh H., Shrivastava N. 2007. Plant genomic DNA isolation: an art or a science. Biotechnology Journal 2, 386–392.

Wang X.H., Xiao H.L., Chen G.X., Zhao X., Huang C.H., Chen C.Y., Wang F. 2011. Isolation of high-quality RNA from Reaumuria soongorica, a desert plant rich in secondary metabolites. Molecular Biotechnology 48, 165–172.

Weishing K., Nybom H., Wolff K., Meyer W. 1995. DNA isolation and purification. In: DNA fingerprinting in plants and fungi, pp 44–59. CRC Press, Boca Raton, Florida.

Weising K., Nybom H., Wolff K., Kahl G., 2005. DNA Fingerprinting in Plants: Principles, Methods and Applications, 2nd ed. CRC Press, Boca Raton.

Zeillinger R., Schneeberger C., Speiser P., Kury F. 1993. A simple method for isolation of DNA from blood clots suited for use in PCR. Biotechniques, 14, 202–203.

Zhang K. 2011. Comparison of DNA extraction methods from *Herba cistanche*. Journal of Anhui Agri Sci 39, 3957–3959.

2

Olive Tree Genetic Resources Characterization Through Molecular Markers

Sónia Gomes, Paula Martins-Lopes and Henrique Guedes-Pinto
Institute for Biotechnology and Bioengineering, Centre of Genomics and Biotechnology - University of Trás-os-Montes and Alto Douro (IBB/CGB-UTAD), Vila Real Portugal

1. Introduction

The *Olea europaea* L. is considered one of the most widely grown fruit crop in the countries of the Mediterranean basin. The olive products, such as olive oil, table olives, and olive pastes are the basic constitutes of the Mediterranean diet due to their benefits for human health, besides other applications such as in cosmetics field.

The olive is one of the most ancient cultivated fruit trees. Olive cultivation has a very long history which started from the Third Millennium BC (Loukas & Krimbas, 1983) in the Eastern region of the Mediterranean sea and spread later around the basin following land and maritime routes to Italy, Spain, North Africa and France. Nowadays there are about 805 million of olive trees, 98% of which are grown in the Mediterranean countries (Tsitsipis et al., 2009). The foremost consuming countries are also the main olive oil producers. According to Food and Agriculture Organization of the United Nations, Mediterranean countries produce more than 90% of world olives, and the biggest olive producers are Spain, Italy, Greece, Turkey, Tunisia, Morocco, Syria, and Portugal (FAO, 2008). Other consuming countries are the United States, Canada, Australia and Japan (Hatzopoulos et al., 2002; Pinelli et al., 2003). Though the olive oil consumption has been mainly constricted to Mediterranean countries, actually it has been extended to other areas due to its health beneficial proprieties (Bracci et al., 2011). Over the centuries, olive trees were propagated mainly vegetatively and were selected based on olive quantitative and qualitative traits. However, this procedure did not exclude the problematic of natural crossing between the newly introduced cultivars and the local germplasm and somatic mutation events, and genetic variability among the olive tree collections has been reported by several authors (Angiolillo et al., 1999; Bautista et al., 2003; Belaj et al., 2002, 2003, 2004, 2006; Cordeiro et al., 2008; Gemas et al., 2000; Gomes et al., 2008, 2009; Martins-Lopes et al., 2007, 2009; Sefc et al., 2000). In addition, the olive tree is allogamous, easily generating crosses between cultivars which give rise to high genetic variability between and within cultivars (Mekuria et al., 1999; Ouazzani et al., 1996; Zohary, 1994).

More than 2600 cultivars have been described for *Olea europaea* L. using morphologic analyzes (Rugini & Lavee, 1992), although many of them might be synonyms, homonyms, ecotypes or the result of crosses between neighbouring olive cultivars (Barranco et al., 2000). Bartolini et al. (1998) reported that there are 79 olive collections located in 24 countries

which contain about 1200 cultivars with more than 3000 different names. The high number of olive cultivars causes a huge problem in the germplasm collections management and traceability and authenticity of olive oils produced, once there is an uncertainty about its olive cultivar correct denomination (Cipriani et al., 2002).

Until recent years, cultivars' identification was based only on morphological and agronomic traits. However, recognition of olive cultivars based on phenotypic characters revealed to be problematic, especially in early stages of tree development. Traditionally diversity within and between olive tree cultivars was determined by assessing differences in olive tree, namely leaf shape and color, and olive fruits morphology. These measures have the advantage of being readily available, do not require sophisticated equipment and are the most direct measure of phenotype, thus they are accessible for immediate use, an important attribute. However, these morphological and phenological markers have the disadvantage of the small number of polymorphism detected and of being environmentally dependent (Mohan et al., 1997; Tanksley & Orton, 1983). Besides that, some of the phenological characteristics are only accessible for a limited (e.g., olive fruits) or when the olive tree achieves a mature stage, which may delay the correct identification. Due to the high genetic diversity level observed in olive germplasm and the presence of homonyms and synonyms cases, efficient and rapid discriminatory methods are urgently.

In recent years, molecular markers have been applied in olive germplasm to identify cultivars and to determine the relationships between cultivars. Molecular markers are, according to Kahl (2004), any specific DNA segment whose base sequence is polymorphic in different organisms. Such markers can be visualised by hybridization-based techniques such as restriction fragment length polymorphism (RFLP) or by polymerase chain reaction (PCR)-based methods.

Molecular markers present numerous advantages over conventional phenotype based alternatives. The choice and selection of an adequate marker system depends upon the type of study to be undertaken and whether it will fulfil at least a few of the mentioned criteria: (a) highly polymorphic between two organisms, inherited codominantly, (b) evenly distributed throughout the genome and easily visualized, (c) occurs frequently in the genomes, (d) stable over generations, (e) simple, quick and inexpensive, (f) small amounts of DNA samples required, and (g) no prior information about the sample's genome (Agarwal et al., 2008; Hatzopoulos et al., 2002).

Because of their high polymorphism level and discerning power, molecular markers have been used as a powerful tool for olive gene pools' characterization. Molecular markers have played a crucial role to distinguish, characterize, and to elucidate olive germplasm origin and diversity. Different molecular markers have been applied for olive genetic diversity assessment, such as the dominant random amplified polymorphic DNA (RAPD) (Belaj et al., 2003; Cordeiro et al., 2008; Gemas et al., 2004; Gomes et al., 2009; Martins-Lopes et al., 2007, 2009; Trujillo et al., 1995) and inter simple sequence repeat (ISSR) markers (Essadki et al., 2006; Gemas et al., 2004; Gomes et al., 2009; Martins-Lopes et al., 2007, 2009). The codominant microsatellite (SSR) (Belaj et al., 2003; Bracci et al., 2009; Gomes et al., 2009; Sabino et al., 2006; Sarri et al., 2006; Sefc et al., 2000), and amplified fragment length polymorphism (AFLP) (Ercisli et al., 2009; Grati-Kamoun et al., 2006; Montemurro et al., 2005) have been used for olive germplasm characterization.

However, the disadvantageous associated with some type of markers, like the less sensibility, and reproducibility of RAPD or the complexity of the AFLP assay, makes it necessary to convert interesting markers (bands) into sequence-characterized amplified regions (SCAR) or sequence-tagged site (STS) markers (Olson et al., 1989; Paran & Michelmore, 1993). During the olive genome exploration different molecular markers have emerged. The single nucleotide polymorphisms (SNP) has been used to discriminate 49 olive cultivars, selected among the most widely cultivated, for olive oil production, in the Mediterranean area (Consolandi et al., 2007). The presence of retrotransposon-like elements in the olive genome was reported during SCAR development for olive cultivar identification (Hernández et al., 2001b). It is generally accepted that retrotransposons have played an important role in olive genetic instability and genome evolution. The use of retrotransposon sequences to generate molecular markers (e.g., REMAP: retrotransposon microsatellite amplification polymorphism) has been used in olive tree (Natali et al., 2007).

The increasing openness of genetic markers in olive tree allows the detailed studies and evaluation of genetic diversity. Within this context, a review of the state of the art of molecular marker techniques applied for olive cultivars characterization and their applicability in olive germplasm conservation will be presented. This will give a prospect of what has been attained and what still needs to be done in order to better understand this crop that has lived for centuries and still remains to be discovered and understood.

2. Olive tree origins

Mythologically olive tree was a gift of Athens goddess to the Greeks. However, olive tree geographical origin still remains unclear. According to botanists, the olive tree and oleaster correspond to *Olea europaea* subsp. *europaea* L. var. *europaea* and var. *sylvestris*, respectively. Oleaster is the wild form, while the olive is the cultivated form (Breton et al., 2006). The olive tree is self-incompatible. Out-crossing is mediated by the wind that transports pollen over long distances, with cytoplasmic male-sterile cultivars being pollinated efficiently by surrounding cultivars or even by oleasters (Besnard et al., 2000). It is assumed that cultivars have originated from the wild Mediterranean olive (oleasters), and have been disseminated all around the Mediterranean countries following human displacement. It is also presumed that crosses between wild and cultivated forms could have led to new cultivars around Mediterranean countries (Besnard et al., 2001).

In order to understand olive domestication, random amplified polymorphic DNA (RAPD) profiles of 121 olive cultivars were compared to those of 20 natural oleaster populations from Eastern and Western parts of the Mediterranean Basin. The differences observed between groups of cultivars were clear (Besnard et al., 2001). Cultivars from Israel, Turkey, Syria, Greece and Sicily were close to the Eastern oleasters group; on the other hand, clones from Italy, France, Corsica, Spain and the Maghreb were closer to the Western group. Multiple origins for Mediterranean olive (*Olea europaea* L. ssp. *europaea*) based upon mitochondrial DNA variations have been reported (Besnard & Bervillé, 2000). The phylogeographic study revealed the presence of three mitotypes (ME1, MOM and MCK) in both cultivated olive and oleaster; while a fourth mitotype, ME2, was unique to a few cultivars from East to West. This information led to the conclusion that a great majority of the cultivars were originated by maternal descent from the Eastern populations once they carry the mitotypes ME1 or ME2. The cultivars with the Western mitotypes, MOM or MCK,

generally kept a nuclear RAPD profile close to the profile of Western natural populations. Consequently, they could result from exclusively local material (as for Corsica), while ME1 and ME2 are characteristic of the East Mediterranean populations. The presence of these different mitotypes reflects the complexity of olive domestication: the Western Mediterranean is probably a zone where olive trees from the East, once introduced, have been hybridized and back-crossed with the indigenous olives (Besnard et al., 2001).

Bronzini de Caraffa et al. (2002) have performed a study of nuclear and mitochondrial DNAs of cultivated and wild olives, from two Corsican and Sardinian Mediterranean islands, using both random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) markers. The results have indicated that the combination of mitotype and RAPD markers can be used as a powerful tool for differentiating two groups in the wild forms: the Western true oleasters and the feral forms. A recently study has investigated the genetic diversity between Eastern and Western oleasters and between cultivars (Breton et al., 2006). The oleaster genetic diversity, obtained by chloroplast and SSR markers, is divided into seven *reconstructed panmictic oleaster populations* regions (RPOP) in both Eastern and Western populations that could overlay glacial refuges. The authors argue that the gene flow has occurred in oleasters mediated by cultivars spread by human migration or through trade. However, a complex origin for this species, higher than expected initially, was reported (Breton et al., 2006).

3. Non PCR-based markers

3.1 Restriction Fragment Length Polymorphism (RFLP)

The restriction fragment length polymorphism (RFLP) markers are based on the analysis of patterns derived from cleaved DNA sequences with known specific restriction enzymes and hybridization with specific probes (Mohan et al., 1997). Based on this method genetic variation and relationship between 89 very old olive trees and 101 oleasters, cultivated around the Mediterranean basin, have been evaluated by cytoplasmic DNA markers (Amane et al., 1999). A similar approach was used to study chloroplast DNA variation in wild and cultivated Morocco olives (Amane et al., 2000). The analysis revealed the presence of four distinct chlorotypes. Nowadays, restriction fragment length polymorphism (RFLP) markers are not very widely used due to several constrains of the method: (a) time consuming, (b) radioactive and/or toxic reagents, (c) large quantity of high quality genomic DNA, and (d) prior sequence information for probe generation; increasing overall the complexity of the methodology (Agarwal et al., 2008). With the development of PCR based methodologies, this marker has been limited for diversity studies, once PCR methods are more expedite. However, the use of RFLP combined with other molecular techniques has been used for olive tree diversity studies. Besnard et al. (2002) combined the RFLP technique with PCR to analyze the chloroplast DNA diversity in the olive complex.

4. PCR-based markers

Since the PCR has been introduced by Mullis et al. (1986) the molecular studies have profoundly changed the way in which they are conducted. The subsequent development of methods for DNA fingerprinting has introduced the possibility to univocally identify cultivars or clones from a specific area. The different PCR based methods used for olive

diversity evaluation and germplasm characterization will be described in this review, considering always the purpose of the work performed in olive around Mediterranean countries.

4.1 Random Amplified Polymorphic DNA (RAPD)

A new DNA polymorphism assay was first described in 1990 by Williams et al. (1990) and Welsh & McClelland (1990). The random amplified polymorphic DNA (RAPD) marker is based on the amplification by PCR of random DNA segments, using single primers of arbitrary nucleotide sequence. The amplified DNA fragments, referred to as RAPD markers, were shown to be highly useful in the construction of genetic maps. With RAPD method the resulted polymorphisms are detected by electrophoresis as different DNA fragments. The different DNA fragments are generated once the primers used usually anneal with multiple sites in different regions of the genome, producing multiple amplified products that often contain repetitive DNA sequences (Paran & Michelmore, 1993).

The first study using RAPD markers, to evaluate olive germplasm polymorphism was reported by Fabbri et al. (1995). All RAPD data suggest a high degree of genetic diversity in the olive germplasm (Belaj et al., 2006). Several reports detected also a high degree of genetic variability within cultivars of different countries: Iran (Shahriari et al., 2008), Spain (Belaj et al., 2002), and Portugal (Cordeiro et al., 2008; Gemas et al., 2004; Martins-Lopes et al., 2007). Most of the olive cultivars in these studies were clustered according to their fruit's end-use and ecological adaptation. Belaj et al. (2004) found that a combination of three highly polymorphic RAPD primers (OPK16, OPA19 and OPX09) was optimal to discriminate among 103 cultivars. Inter- and intravarietal variation of three olive cultivars, 'Galega Vulgar', 'Cordovil de Serpa' and 'Verdeal Alentejana', were also observed with RAPD markers (Gemas et al., 2000). The clonal diversity has been accessed using RAPD markers in combination with inter simple sequence repeat (ISSR) and simple sequence repeat (SSR) in two important Portuguese olive cultivars, 'Verdeal-Transmontana' (Gomes et al., 2008) and 'Cobrançosa' (Martins-Lopes et al., 2009). The highest proportion of polymorphic products, observed in 'Verdeal-Transmontana' clones was generated using primer OPO10 (88%), and the mean level of polymorphism was 28%. In the 'Cobrançosa' the authors reported a considerable polymorphism among the DNA fingerprints of the clones. The RAPD primers amplified 150 reproducible fragments, of which 75% were polymorphic. The high level of polymorphism reported demonstrates that the Portuguese 'Verdeal-Transmontana' and 'Cobrançosa' cultivars were genetically heterogeneous, confirming that olive is a highly variable species. Recently, genetic similarities and distances among Turkish wild olive trees were studied in order to improve genetic resources and knowledge of cultivars evolutionary background (Sesli & Yegenoglu (2009).

However, RAPD methodologies have its criticisms due to the low data reproducibility between laboratories, although it may be quite reliable at the same laboratory. The fact of low cost, low time usage, low DNA amount even if not of good quality and no previous DNA sequence knowledge made this molecular marker technique one of the first to be used to access genetic variability. A variation of the RAPD technique, such as arbitrarily primed PCR (AP-PCR) (Welsh & McClelland, 1990), that involves the increase of annealing temperature during the PCR cycles, has been used, in Turkey, to characterize and select six important olive clones used for olive oil production Kockar & Ilıkcı (2003). Claros et al.

(2000) used the same methodology for olive geographic location and confirmed the hypothesis of autochthonic origin of most olive-tree cultivars. Nowadays, obtaining specific markers, such as sequence characterized amplified region (SCAR) and sequence-tagged-site (STS), from RAPD markers could be a way to overcome the lack of reproducibility, proper of RAPD markers.

4.2 Amplified Fragment Length Polymorphism (AFLP)

The principle of amplified fragment length polymorphism (AFLP) (Vos et al., 1995) technique is basically simple and its procedure consists of three main steps: (a) template preparation, (b) fragment amplification, and (c) gel analysis. The fingerprinting patterns are obtained by detection of genomic restriction fragments by PCR amplification. This technique has been widely employed. Because of its effectiveness, reliability and efficiency in genetic diversity studies (Ercisli et al., 2009), the AFLP technique has been widely used in olive Spanish cultivars considering intra-varietal diversity (Sanz-Cortés et al., 2003), and to assess genetic inter-relationships among cultivated cultivars in the Eastern Mediterranean Basin (Owen et al., 2005). The results showed significant genetic distance between Greek and Turkish cultivars, and a clear separation of most of the Spanish and Italian clones, suggesting that an East-West divergence of olive cultivars occurred. Using the AFLP markers Angiolillo et al. (1999) have shown that wild olives from the Western Mediterranean and cultivated cultivars did not cluster together, and were relatively distant. However, a few oleasters clustered with the cultivars suggesting a common origin.

The first linkage map of the olive genome was constructed using a combination of molecular markers (e.g., RAPD, AFLP, RFLP and SSR) (De la Rosa et al., 2003). Maps can be used to select important traits and to study genes that control expression of polygenic traits. Molecular marker linkage maps are widely recognized as essential tools for genetic research and breeding in many species.

4.3 Microsatellites (SSR)

The simple sequence repeat (SSR) (Tautz et al., 1986; Litt & Luty, 1989) consists of short (1-6 base pair long) stretches of DNA tandem repeated several times, occurring in the genomes of many higher organisms (Rafalski & Tingey, 1993; Wu & Tanksley, 1993). The simple sequence repeat or microsatellites, as one of the most popular marker system, are widely used in plant genetic research for diversity studies, namely in olive tree and to test the breeding success as they are transferable, highly polymorphic, ideal for genetic map development, linkage analysis, marker-assisted selection and fingerprinting studies (Bracci et al., 2009; Cipriani et al., 2002; De la Rosa el al., 2004; Gomes et al., 2009; Karp et al., 1996; Muzzalupo et al., 2009; Rallo et al., 2002; Sefc et al., 2000). When compared with RAPD or AFLP markers, the SSR have the advantage of their codominant nature, as two alleles may be identified at each locus. The main constrain of SSR markers is the development requires previous DNA sequencing for primer designing.

The microsatellites loci have been isolated from olive tree (Carriero et al., 2002; Cipriani et al., 2002; De la Rosa et al., 2002; Rallo et al., 2000; Sefc et al., 2000) and are used either alone or in combination with other molecular markers to characterize olive cultivars (Belaj et al., 2004; Gomes et al., 2009; Khadari et al., 2003; Wu & Sedgley, 2004). This methodology has

been used to analyze the genetic variability of the somatic embryogenesis induction process in *Olea europaea* L. and *Olea europaea* var. *maderensis*. The authors reported the maintenance of the genomic integrities between species suggesting the absence of somaclonal variation (Lopes et al., 2009). New insights about genetic diversity and gene flow between the wild (oleaster) and the cultivated form, using SSR marker was reported (Breton et al., 2006). A database containing a consensus list of SSR profiles for true-to-type olive genotyping has been constructed. This platform will allow results' comparison among laboratories, in order to establish a common olive database (Baldoni et al., 2009).

During many years the agarose gel electrophoresis has been used as the common detection method for SSR analysis. The agarose gel is efficient when the alleles are long enough, that is, more than 200-300 base pair and the differences among alleles are also significant to be visualized (i.e., more than 10-20 base pair). The high resolution polyacrylamide gels have been used when small differences between alleles, less than 1-10 base pair, must be identified. Nowadays, the separation of SSR markers using sequencing apparatus revealed to be very suitable, since the detection of alleles is performed automatically. The major advantages of automated detection are: (a) faster in obtaining results, (b) automated data analysis, (c) multiplex analysis, (d) high reproducibility, and (e) exclusion of silver-staining procedure. However, between different apparatus there may be found a shift among allele size, which has to be undertaken when comparing results among laboratories.

4.4 Inter Simple Sequence Repeats (ISSR)

In order to resolve some of the inconveniences associated with RAPD (low reproducibility), the high AFLP cost, and the need to know the flanking sequences in order to developed primers for SSR polymorphism, ISSR were developed (Terzopoulos et al., 2005; Zietkiewicz et al., 1994). ISSR markers are based on the amplification of regions (200-2000 base pair) between inversely oriented closely spaced microsatellites. The ISSR show the specificity of microsatellite markers, but need no sequence information for primer synthesis. The ISSR alone or in combination with other marker systems, have been widely used to analyze clonal variation and genetic variability in olive cultivars (Gemas et al., 2004; Gomes et al., 2008; Martins-Lopes et al., 2007, 2009; Terzopoulos et al., 2005).

Previous studies have concluded that ISSR markers are efficient in assessing phylogenetic relationships in the *O. europaea* complex (Gemas et al., 2004; Hess et al., 2000) and for olive fruits and leaves identification (Pasqualone et al., 2001). The simultaneous use of ISSR with other markers such as RAPD has made possible the discrimination between 30 Portuguese and 8 foreign olive cultivars (Martins-Lopes et al., 2007).

4.5 Sequence Characterized Amplified Region (SCAR)

Since PCR-based molecular markers have been developed, several PCR-based markers modifications have emerged. Due to the certification process of orchards and regions, crucial for protected denomination of origin (PDO), there is an urgent need for early and efficient methods able to discriminate and indentify olive cultivars. The development of cultivar-specific DNA markers can also be useful in olive industry in order to avoid olive oil adulteration that affects the oil quality (Marieschi et al., 2011; Pafundo et al., 2007).

The sequence characterized amplified region (SCAR) have been widely developed for plant breeding studies in several species such as wheat (Hernández et al., 1999), grapevine (Vidal et al., 2000), tomato (Zhang & Stommel, 2001), and pear (Lee et al., 2004; Marieschi et al., 2011). In olive, this type of marker has also been applied for olive germplasm evaluation and mapping (Bautista et al., 2003; Busconi et al., 2006; Hernández et al., 2001a), and for analysis of complex agro-food matrixes (olive oil traceability) (Pafundo et al., 2007).

The development of sequence characterized amplified region (SCAR) involves cloning of the amplified product, and then sequencing the two ends of the cloned product that appeared to be specific. The SCAR has the advantage of being inherited in a codominant fashion in contrast to RAPD which are inherited in a dominant manner (Mohan et al., 1997). Bautista et al. (2003) used this technology to develop specific markers useful for olive cultivar identification and mapping. They demonstrated that the use of SCAR markers is enough to provide a simple, cheap, and reliable procedure to identify geographically related olive cultivars. The development of SCAR markers by directly sequencing olive RAPD bands was reported by Hernández et al. (2001a) and they demonstrated that the generated markers were useful for the marker assisted selection of the high flesh/stone ratio. This type of marker has also been applied for olive germplasm evaluation and mapping (Bautista et al., 2003; Busconi et al., 2006). Wu et al. (2004) combined RAPD, SCAR and SSR markers to construct a linkage map from a cross-pollinated F_1 population of 'Frantoio' × 'Kalamata' olive cultivars.

4.6 Single Nucleotide Polymorphisms (SNP)

The single nucleotide polymorphisms are a marker system that can differentiate individuals based on variations detected at the level of a single nucleotide base in the genome. Such variations are present in large abundance in the genomes of higher organisms including plants (Agarwal et al., 2008). The SNP-based markers have been used in many plant species.

In olive, due to olive unknown genome, this technique has not been widely applied. Reale et al. (2006) used SNP markers to genotype 65 olive samples obtained from Europe and Australia, and observed that 77% of the cultivars were clearly discriminate. However, the authors developed SNP markers from olive gene sequences available in the GenBank database and from arbitrary sampling using the sequence-related amplification polymorphism (SRAP) method.

5. Conclusions

Nowadays, the olive industry requires certified olive cultivars with elite agronomic characteristics and adapted to modern intensive mechanized orchards (Hatzopoulos et al., 2002). Very few cultivars are grown commercially in more than one region or country, while most of them have a local diffusion. The cases of cultivars homonyms and synonyms associated with high genetic diversity makes the olive tree germplasm very difficult to characterize. The PCR-based markers opened the possibility to develop, over the last two decades, new molecular techniques for cultivar identification and further certification purposes in order to certify the propagated material. It is essential to study the genetic base of olive germplasm in order to characterize and compare with other genetic, phenotypic and agronomic data. Different molecular markers have been used in genetic diversity studies which give us information about the relationships between cultivars and the olive domestication process.

The choice on which molecular technique would be the most suitable for olive genetic resource characterization depend on a number of factors as the level of variability of the species, and the resources available (Belaj et al., 2006). Technological advancement has contributed to the development, in every aspect, of molecular genetic markers, making them technically simpler, efficient, cost-effective, and faster than the classic methods.

However, molecular approaches (nuclear and cytoplasmic) should not be considered alone or as substitutes of morphological characterization but as complementary tools, more complete and effective, for olive genetic resources studies. The several molecular markers used for germplasm variability studies may play a major role in olive tree breeding programs when using marker assisted selection for biotic and abiotic stress tolerance, olive fruits and oil quality traits.

However, there are still aspects of cultivars synonymous that still needed to be addressed in order to develop a complete database, in order to have an overview of the genetic variability available. As soon as the recent olive genome sequence is released new strategies may be taken in order olive germplasm management, breeding strategies and certification issues.

6. References

Agarwal, M.; Shrivastava, N. & Padh H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*, Vol.27, pp. 617–631.

Amane, M.; Lumaret, R.; Hany, V.; Ouazzani, N.; Debain, C.; Vivier, G. & Deguilloux, M.F. (1999). Chloroplast-DNA variation in cultivated and wild olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, Vol.99, No.1-2, pp. 133-139.

Amane, M.; Ouazzani, N.; Lumaret, R. & Debain, C. (2000). Chloroplast DNA variation in the wild and cultivated olives (*Olea europaea* L.) of Morocco. *Euphytica*, Vol.116, pp. 59-64.

Angiolillo, A.; Mencuccini, M. & Baldoni, L. (1999). Olive genetic diversity assessed using amplified fragment length polymorphisms. *Theoretical and Applied Genetics*, Vol.98, pp. 411–421.

Baldoni, L.; Nicolò, G.C.; Mariotti, R.; Ricciolini, C.; Arcioni, S.; Vendramin, G.V.; Buonamici, A.; Porceddu, A.; Sarri, V.; Ojeda, M.A.; Trujillo, I.; Rallo, L.; Belaj, A.; Perri, E.; Salimonti, A.; Muzzalupo, I.; Casagrande, A.; Lain, O.; Messina, R. & Testolin, R. (2009). A consensus list of microsatellite markers for olive genotyping. *Molecular Breeding*, Vol.24, pp. 213–231.

Barranco, D.; Cimato, A.; Fiorino, P.; Rallo, L.; Touzani, A.; Castaneda, C. & Trujillo, I. (2000). *World catalogue of olive varieties*. International Olive Oil Council, Madrid, Espanha.

Bartolini, G.; Prevost, G.; Messeri, C. & Carignani, G. (1998). Olive germplasm: cultivars and world-wide collections. FAO. Rome.

Bautista, R.; Crespillo, R. & Canovas, F.M. (2003). Identification of olive-tree cultivars with SCAR markers. *Euphytica*, Vol.129, pp. 33-41.

Belaj, A.; Trujillo, L. & Rallo, L. (2002). RAPD's analysis supports the autochthone origin of olive cultivars. *Acta Horticulturae*, Vol.586, pp. 83-86.

Belaj, A.; Caballero, J.M. & Barranco, D. (2003). Genetic characterization and identification of new accessions from Syria in an olive germplasm bank by means of RAPD markers. *Euphytica*, Vol.134, pp. 261-268.

Belaj, A.; Satovic, Z.; Rallo, L. & Trujillo I. (2004). Optimal use of RAPD markers for varietal identification in olive (*Olea europaea* L.) germplasm collections. *Journal of the American Society for Horticultural Science,* Vol.129, No.2, pp. 266-270.

Belaj, A.; Ojeda, M.A.; Muñoz, C.; Rodríguez, E.; Díaz, A.; Rallo, P.; De la Rosa, R.; Barranco, D.; Rallo, L. & Trujillo, I. (2006). The use of molecular markers to characterize olive (*Olea europaea* L.) germplasm and its wild relatives. OliveBioteq. pp.391-396.

Besnard, G. & Bervillé, A. (2000). Multiple origins of the Mediterranean olive (*Olea europaea* L. ssp. *europaea*) based upon mitochondrial DNA polymorphisms. *Compte Rendu Académie des Sciences, Sciences de la Vie*, Vol.323, pp. 173–181.

Besnard, G.; Khadari, B.; Villemur, P. & Bervillé, A. (2000). A cytoplasmic male sterility in olive cultivars (*Olea europaea* L.): phenotypic, genetic and molecular approaches. *Theoretical and Applied Genetics*, Vol.100, pp.1018–1024.

Besnard, G.; Baradat, P.; Breton, C.; Khadari, B. & Bervillé, A. (2001). Olive domestication from structure of oleasters and cultivars using RAPDs and mitochondrial RFLP. *Genetics Selection Evolution*, Vol.33, S251–S268.

Besnard, G.; Khadari, B.; Baradat, P. & Bervillé, A. (2002). Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasmic genetic differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics*, Vol.105, pp. 139-144.

Bracci, T.; Sebastiani, L.; Busconi, M.; Fogher, C.; Belaj, A. & Trujillo, I. (2009). SSR markers reveal the uniqueness of olive cultivars from the Italian region of Liguria. *Scientia Horticulturae*, Vol.122, pp. 209-215.

Bracci, T.; Busconi, M.; Sebastiani, L. & Fogher, C. (2011). Molecular studies in olive (*Olea europaea* L.): overview on DNA markers applications and recent advances in genome analysis. *Plant Cell Report*, Vol.30, pp. 449–462.

Breton, C.; Tersac, M. & Berville, A. (2006). Genetic diversity and gene flow between the wild olive (*oleaster*, *Olea europaea* L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *Journal Biogeography*, Vol.33, pp. 1916-1928.

Bronzini de Caraffa, V.; Maury, J.; Gambotti, C.; Breton, C.; Bervillé, A. & Giannettini, J. (2002). Mitochondrial DNA variation and RAPD mark oleasters, olive and feral olive from Western and Eastern Mediterranean. *Theoretical and Applied Genetics*, Vol.104, pp. 1209-1216.

Busconi, M.; Sebastiani, L. & Fogher, C. (2006). Development of SCAR markers for germplasm characterisation in Olive Tree (*Olea europaea* L.). *Molecular Breeding*, Vol.17, No.1, pp. 59-68.

Carriero, F.; Fontanazza, G.; Cellini, F. & Giorio, G. (2002). Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, Vol.104, pp. 301-307.

Cipriani, G.; Marrazzo, M.T.; Marconi, R.; Cimato, A. & Testolin, R. (2002). Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theoretical and Applied Genetics*, Vol.104, pp. 223-228.

Claros, M.G.; Crespillo, R.; Aguilar, M.L. & Cánovas, F.M. (2000). DNA fingerprinting and classification of geographically related genotypes of olive-tree (*Olea europaea* L.). *Euphytica*, Vol.116, pp. 131–142.

Consolandi, C.; Palmieri, L.; Doveri, S.; Maestri, E., Marmiroli, N.; Reale, S.; Lee, D.; Baldoni, L; Tosti, N.; Severgnini, M.; De Bellis, G. & Castiglioni, B. (2007). Olive variety

identification by ligation detection reaction in a universal array format. *Journal of* Biotechnology, Vol.129, pp. 565–574.

Cordeiro, A.I.; Sanchez-Sevilla, J.F.; Alvarez-Tinaut, M.C. & Gomez-Jimenez, M.C. (2008). Genetic diversity assessment in Portugal accessions of *Olea europaea* by RAPD markers. *Biologia Plantarum*, Vol.52, No.4, pp. 642-647.

De la Rosa, R.; Martin, A.; Rallo, L.; Angiolillo, A.; Guerrero, C. & Baldoni, L. (2002). RAPD and AFLP analysis for Olive mapping. *Acta Horticulturae*, Vol.586, pp. 79-82.

De la Rosa, R.; Angiolillo, A.; Guerrero, C.; Pellegrini, M.; Rallo, L.; Besnard, G.; Berville, A.; Martin, A. & Baldoni, L. (2003). A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. *Theoretical and Applied Genetics*, Vol.106, pp. 1273-1282.

De la Rosa, R.; James, C.M. & Tobutt, K.R. (2004). Using microsatellites for paternity testing in olive progenies. *HortScience*, Vol.39, pp. 351–354.

Ercisli, S.; Barut, E. & Ipek, A. (2009). Molecular characterization of olive cultivars using amplified fragment length polymorphism markers. *Genetics and Molecular Research*, Vol.8, pp. 414-419.

Essadki, M.; Ouazzani, O.; Lumaret, R. & Moumni, M. (2006). ISSR variation in olive-tree cultivars from Morocco and other western countries of the Mediterranean Basin. *Genetic Resource and Crop Evolution,* Vol.53, No.3, pp. 475–482.

Fabbri, A.; Hormaza, J.I. & Polito, V.S. (1995). Random amplified polymorphic DNA analysis of olive (*Olea europaea* L.) cultivars. *Journal of the American Society for Horticultural Science*, Vol.120, No.3, pp. 538–542.

FAO (2008). Agricultural Statistics of the Food and Agriculture Organization of the United Nations, Rome. www.FAO.org. Accessed March 18 2010.

Gemas, V.J.V.; Rijo-Johansen, M.J., Tenreiro, R. & Fevereiro, P. (2000). Inter- and intra-varietal analysis of tree *Olea europaea* L. cultivars using RAPD technique. *Journal of Horticultural Science and Biotechnology*, Vol.75, No.3, pp. 312-319.

Gemas, V.J.V.; Almadanim, M.C.; Tenreiro, R.; Martins, A. & Fevereiro, P. (2004). Genetic diversity in the Olive tree (*Olea europaea* L. subsp. *europaea*) cultivated in Portugal revealed by RAPD and ISSR markers. *Genetic Resource and Crop Evolution*, Vol.5, pp. 501-511.

Gomes, S.; Martins-Lopes, P.; Lima-Brito, J.; Meirinhos, J.; Lopes, J.; Martins, A. & Guedes-Pinto, H. (2008). Evidence of Clonal Variation in Olive 'Verdeal-Transmontana' cultivar using RAPD, ISSR and SSR Markers. *Journal of Horticultural Science and Biotechnology*, Vol.83, No.4, pp. 395-400.

Gomes, S.; Martins-Lopes, P.; Lopes, L. & Guedes-Pinto, H. (2009). Assessing genetic diversity in *Olea europaea* L. using ISSR and SSR markers. *Plant Molecular Biology Reports*, Vol.123, pp. 82-89.

Grati-Kamoun, N.; Mahmoud, F.; Rebaï, A.; Gargouri, A.; Panaud, O. & Saar, A. (2006). Genetic diversity of Tunisian olive tree (*Olea europaea L.*) cultivars assessed by AFLP markers. *Genetic Resource and Crop Evolution*, Vol.53, pp. 265-275.

Hatzopoulos, P.; Banilas, G.; Giannoulia, K.; Gazis, F.; Nikoloudakis, N.; Milioni, D. & Haralampidis, K. (2002). Breeding molecular markers and molecular biology of the olive tree. *European Journal of Lipid Science and Technology,* Vol.104, pp. 574-586.

Hernández, P.; Martín, A. & Dorado, G. (1999). Development of SCARs by direct sequencing of RAPD products: practical tool for the introgression and marker-assisted selection of wheat. *Molecular Breeding*, Vol.5, pp. 245-253.

Hess, J.; Kadereit, J.W. & Vargas, P. (2000). The colonization history of *Olea europaea* L. in *Macaronesia* based on internal transcribed spacer 1 (ITS-1) sequences, randomly amplified polymorphis DNAs (RAPD), and intersimple sequence repeats (ISSR). *Molecular Ecology*, Vol.9, pp. 857–868.

Hernández, P.; de la Rosa, R.; Rallo, L. & Dorado, G. (2001a). Development of SCAR markers in olive (*Olea europaea*) by direct sequencing of RAPD products: applications in olive germplasm evaluation and mapping. *Theoretical and Applied Genetics*, Vol.103, pp. 788-791.

Hernández, P.; de la Rosa, R.; Rallo, L.; Martin, A. & Dorado, G. (2001b). First evidence of a retrotransposon-like element in olive (*Olea europaea*): Implications in plant variety identification by SCAR marker development. *Theoretical and Applied Genetics*, Vol.102, pp. 1082-1087.

Kahl, G. (2004). The dictionary of gene technologies. Genomics, transcriptomics, proteomics. 3rd Edition. Volume 1 and 2 Wiley – VCH Verlag, Gmb H &Co. KGaA, Weinheim.

Karp, A.; Seberg, O. & Buiatti, M. (1996). Molecular techniques in the assessment of botanical diversity. *Annals of Botany*, Vol.78, pp. 143-149.

Khadari, B.; Breton, C.; Moutier, N.; Roger, J.P., Besnard, G.; Bervillé, A. & Dosba, F. (2003). The use of molecular markers for germplasm management in a French olive collection. *Theoretical and Applied Genetics*, Vol.106, pp. 521-529.

Kockar, F. & Ilıkcı, R. (2003). Investigations of genetic variation between olive (*Olea europaea* L.) cultivars using arbitrarily primed polymerase chain reaction (AP-PCR). *PCR Applied to Genetics of Olives*, Vol.58, pp. 837- 842.

Lee, G.P.; Lee, C.H. & Kim, C.S. (2004). Molecular markers derived from RAPD, SCAR, and the conserved 18S rDNA sequences for classification and identification in *Pyrus pyrifolia* and *P. communis*. *Theoretical and Applied Genetics*, Vol.108, pp. 1487–1491.

Litt, M. & Luty, J.A. (1989). A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetic*, Vol.44, pp. 397-401.

Lopes, T.; Capelo, A.; Brito, G.; Loureiro, J. & Santos, C. (2009). Genetic variability analyses of the somatic embryogenesis induction process in *Olea* spp. using nuclear microsatellites. *Trees-Structure and Function,* Vol.23, pp. 29-36.

Loukas, M. & Krimbas, C.B. (1983). History of Olive cultivars based on their genetic distances. *Journal of Horticultural Science,* Vol.58, pp. 121–127.

Marieschi, M.; Torelli, A.; Bianchi, A. & Bruni, R. (2011). Development of a SCAR marker for the identification of *Olea europaea* L.: A newly detected adulterant in commercial Mediterranean oregano. *Food Chemistry*, Vol.126, No.2, pp. 705-709.

Martins-Lopes, P.; Lima-Brito, J.; Gomes, S.; Meirinhos, J.; Santos, L. & Guedes-Pinto, H. (2007). RAPD and ISSR molecular markers in *Olea europaea* L.: Genetic variability and cultivar identification. *Genetic Resource and Crop Evolution*, Vol.54, pp. 117-128.

Martins-Lopes, P.; Gomes, S.; Lima-Brito, J.; Lopes, J. & Guedes-Pinto, H. (2009). Assessment of clonal genetic variability in *Olea europaea* L. 'Cobrançosa' by molecular markers. *Scientia Horticulturae*, Vol.123, pp. 82-89.

Mekuria, G.; Collins, G. & Sedgley, M. (1999). Genetic variability between different accessions of some common commercial olive cultivars. *Journal of Horticultural Science and Biotechnology*, Vol.74, pp. 309–314.

Mohan, M.; Nair, S.; Bhagwat, A.; Krishna, T.G.; Yano, M.; Bhatia, C.R. & Sasaki, T. (1997). Genome mapping, molecular markers and markers-assisted selection in crop plants. *Molecular Breeding*, Vol.3, pp. 87-103.

Mullis, K.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology*, 51 Pt 1, pp. 263–373.

Montemurro, C.; Simeone, R.; Pasqualone, A.; Ferrara, E. & Blanco, A. (2005). Genetic relationships and cultivar identification among 112 olive accessions using AFLP and SSR markers. *Journal of Horticultural Science and Biotechnology*, Vol.80, pp. 105-110.

Muzzalupo, I.; Stefanizzi, F. & Perri, E. (2009). Evaluation of olives cultivated in southern Italy by simple sequence repeat markers. *HortScience*, Vol.44, pp. 582-588.

Natali, L.; Giordani, T.; Buti, M. & Cavallini, A. (2007). Isolation of Ty1-copia putative LTR sequences and their use as a tool to analyse genetic diversity in *Olea europaea* *Molecular Breeding*, Vol.19, pp. 255-265.

Olson, M.; Hood, L.; Cantor, C.H. & Botstein, D. (1989). A common language for physical mapping of the human genome. *Science*, Vol.24, pp.1434-1435.

Owen, C.A., Bita, E.C.; Banilas, G.; Hajjar, S.E.; Sellianakis, V. & Aksoy, U. (2005). AFLP reveals structural details of genetic diversity within cultivated olive germplasm from the Eastern Mediterranean. *Theoretical and Applied Genetics*, Vol.110, pp. 1169–1176.

Ouazzani, N.; Lumaret, R. & Villemur, P. (1996). Genetic variation in the olive tree (*Olea europaea* L.) cultivated in Morocco. *Euphytica*, Vol.91, pp. 9-20.

Pafundo, S.; Agrimonti, C.; Maestri, E. & Marmiroli, N. (2007). Applicability of SCAR markers to food genomics: olive oil traceability. *Journal of Agricultural and Food Chemistry*, Vol.55, pp. 6052–6059.

Paran, I. & Michelmore, W.R. (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics*, Vol.85, pp. 985-993.

Pasqualone, A.; Caponio, F. & Blanco, A. (2001). Inter-simple sequence repeat DNA markers for identification of drupes from different *Olea europaea* L. cultivars. *European Food Research Technology*, Vol.213, pp. 240-243.

Pinelli, P.; Galardi, C.; Mulinacci, N.; Vincieri, F. F.; Cimato, A.; Romani, A. (2003). Minor polar compound and fatty acid analyses in monocultivar virgin olive oils from Tuscany. *Food Chemistry*, Vol.80, pp. 331–336.

Rafalski, J.A. & Tingey, S.V. (1993). Genetic diagnostics in plant breeding: RAPD, microsatellites and machines. *Trends in Genetics*, Vol.9, pp. 275-279.

Rallo, P.; Dorado, G. & Martin, A. (2000). Development of Simple Sequence Repeats (SSRs) in Olive tree (*Olea europaea* L.). *Theoretical and Applied Genetics*, Vol.101, pp. 984–989.

Rallo, P.; Dorado, G. & Martin, A. (2002). Application of Microsatellite Markers in olive Breeding. *Acta Horticulturae*, Vol.586, pp. 69-71.

Reale, S.; Doveri, S.; Diaz, A.; Angiolillo, A.; Lucentini, L.; Pilla, F.; Martín, A.; Donini, P. & Lee, D. (2006). SNP-based markers for discriminating olive (*Olea europaea* L.) cultivars. *Genome*, Vol.49, No.9, pp. 1193-1205.

Rugini, E. & Lavee, S. (1992). Olive. In: F.A. Hammerschlag and R.E. Litz (eds.). Biotechnology of perennial fruit crops. CAB Intl., Wallingford, U.K. pp. 371-382.

Sabino, G.F.; Busconi, M.; da Câmara, A. & Fogher, C. (2006). Development and characterization of microsatellite loci from *Olea europaea*. *Molecular Ecology Notes*, Vol.6, pp. 1275–1277.

Sanz-Cortés, F.; Parfitt, D.E.; Romero, C.; Struss, D.; Llacer, G. & Badenes, M.L. (2003). Intraspecific olive diversity assessed with AFLP. *Plant Breeding*, Vol.122, pp. 173–177.

Sarri, V.; Baldoni, L.; Porceddu, A.; Cultrera, N.G.M.; Contento, A.; Frediani, M.; Belaj, A.; Trujillo, I. & Cionini, P.G. (2006). Microsatellite markers are powerful tools for

discriminating among olive cultivars and assigning them to geographically defined populations. *Genome*, Vol.49, pp.1606–1615.

Sefc, K.M.; Lopes, M.S.; Mendonça, D.; Rodrigues dos Santos, M.; Câmara Machado, M.L. & Câmara Machado, A. (2000). Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. *Molecular Ecology*, Vol.9, pp. 1171–1173.

Sesli, M. & Yegenoglu, E.D. (2009). Genetic analysis on wild olives by using RAPD markers. *African Journal of Agriculture Research,* Vol.8, pp. 707-712.

Shahriari, M.; Omrani, A.; Falahati-Anbaran, A.; Ghareyazei, B. & Nankali, A. (2008). Identification of Iranian olive cultivars by using RAPD and microsatellite markers. *Acta Horticulturae*, Vol.791, pp. 109-115. j

Tanksley, S.D. & Orton, T.J. (1983). Isozymes in plant genetics and breeding. Elsevier, Amsterdam.

Tautz, D.; Trick, M. & Dover, G.A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature*, Vol.322, pp. 652-656.

Terzopoulos, P.J.; Kolano, B.; Bebeli, P.J.; Kaltsikes, P.J. & Metzidakis, I. (2005). Identification of *Olea europaea* L. cultivars using inter-simple sequence repeat markers. *Scientia Horticulturae*, Vol.105, pp. 45–51.

Trujillo, I.; Rallo, L. & Arus, P. (1995). Identifying olive cultivars by isozyme analysis. *Journal of the American Society for Horticultural Science*, Vol.120, pp. 318–324.

Tsitsipis, J.A.; Varikou, K.; Kalaitzaki, A.; Alexandrakis, V.; Margaritopoulos, J. & Skouras, P. (2009). Chemical control of olive pests: Blessing or curse?. 4th European Meeting of the IOBC/WPRS Working Group Integrated Protection of Olive Crops'.. Abstract book pp. 31.

Vidal, J.R.; Delavault, P.; Coarer, M. & Defontaine, A. (2000). Design of grapevine (*Vitis vinifera* L.) cultivar-specific SCAR primers for PCR fingerprinting. *Theoretical and Applied Genetics*, Vol.101, pp.1194–1201.

Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; van de Lee, T.; Hornes, M.; Friters, A.; Pot, J.; Paleman, J.; Kuiper, M. & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research,* Vol.21, pp. 4407-4414.

Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, Vol.18, pp. 7213-7218.

Williams, J.G.K.; Kubelik, A.R.; Livak, K.J.; Rafalski, J.A. & Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, Vol.18, pp. 6531-6535.

Wu, K.S. & Tanksley, S.D. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular General Genetics*, Vol. 241, pp. 225-235.

Wu, S.B.; Collins, G. & Sedgley, M. (2004). A molecular linkage map of olive (*Olea europaea* L.) based on RAPD, microsatellite, and SCAR markers. *Genome*, Vol.47, pp. 26-35.

Zhang, Y. & Stommel, J.R. (2001). Development of SCAR and CAPS markers linked to the Beta gene in tomato. *Crop Science*, Vol.41, pp.1602-1608.

Zietkiewicz, E.; Rafalski, A. & Labuda, D. (1994). Genome fingerprinting and simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, Vol.20, pp. 176-183.

Zohary, D. (1994). The wild genetic resources of the cultivated olive. *Acta Horticulturae*, Vol.356, pp. 62-65.

3

Association Mapping in Plant Genomes

Braulio J. Soto-Cerda[1,2] and Sylvie Cloutier[1]
[1]*Cereal Research Center, Agriculture and Agri-Food Canada*
[2]*Agriaquaculture Nutritional Genomic Center, CGNA (R10C1001)*
Genomics and Bioinformatics Unit
[1]*Canada*
[2]*Chile*

1. Introduction

One of the many goals of plant geneticists and breeders pertains to the explanation of phenotypic variation as it relates to changes in DNA sequence (Myles et al., 2009). The development of molecular markers for the detection and exploitation of DNA polymorphisms in plant systems is one of the most significant developments in the field of molecular biology and biotechnology. Linkage mapping has been a key tool for identifying the genetic basis of quantitative traits in plants. However, for linkage studies, suitable crosses, sometimes limited by low polymorphism or small population size, are required. In addition, only two alleles per locus and few recombination events are considered to estimate the genetic distances between marker loci and to identify the causative genomic regions for quantitative trait loci (QTL), thereby limiting the mapping resolution. To circumvent these limitations, linkage disequilibrium (LD) mapping or association mapping (AM) has been used extensively to dissect human diseases (Slatkin, 2008). This approach has received increased attention during the last few years. AM has the potential to identify a single polymorphism within a gene that is responsible for phenotypic differences. AM involves searching for genotype-phenotype correlations among unrelated individuals. Its high resolution is accounted for by the historical recombination accumulated in natural populations and collections of landraces, breeding materials and varieties. By exploiting broader genetic diversity, AM offers three main advantages over linkage mapping: mapping resolution, allele number and time saving in establishing a marker-trait association and its application in a breeding program (Flint-Garcia et al., 2003). Although AM presents clear advantages over linkage mapping, they are often applied in conjunction, especially to validate the associations identified by AM, thus reducing spurious associations.

The inherent nature of AM brings its own limits such as the fact that biological and evolutionary factors affect LD distribution and mapping resolution. The strength of AM relies on the analysis of common variants, which explain at most 5%-10% of the heritable component of human diseases (Asimit & Zeggini, 2010). The role of rare variants in explaining the remaining heritable variation is becoming more important. New statistical models for AM are being developed to better consider rare variants because early methods allocated most of the statistical power to higher frequency alleles.

Since most of the traits important for environmental fitness and agricultural value are quantitative in nature (Yu & Buckler, 2006), there is tremendous interest in using AM to examine them. In this chapter, we introduce the concept of linkage disequilibrium, which plays a central role in association analysis. For this reason, it is critical to understand LD measurement, its variation across the genome and how it is affected by population structure and relatedness. Recent AM studies in plants, advantages and disadvantages of AM, and its integration with other mapping methods are also reviewed and discussed. An overview of the software currently available for AM and their main characteristics is presented. Future perspectives of AM in plants, application in other emerging research areas, potential usefulness for new cultivar development and for the conservation of adaptive genetic variation are outlined.

2. Linkage disequilibrium and association mapping concepts

The terms LD and AM have often been used interchangeably in the literature. However, they present subtle differences. According to Gupta et al. (2005), AM refers to the significant association of a marker locus with a phenotype trait while LD refers to the non random association between two markers or two genes/QTLs (Figure 1). Thus, AM is actually an application of LD. In other words, two markers in LD represent a non random association between alleles, but do not necessarily correlate/associate with a particular phenotype, whereas association implies a statistical significance and refers to the covariance of a marker and a phenotype of interest. Although it lies outside the scope of this section, we would like to also clarify the difference between linkage and LD because they too are commonly confused. Linkage refers to the correlated inheritance of loci through the physical connection on a chromosome, whereas LD refers to the correlation between alleles in a population (Flint-Garcia et al., 2003). Although tight linkage between alleles on the same chromosome generally translate into high LD, significant LD may also exist between distant loci, and even between loci located on different chromosomes. The latter, reviewed in depth below, is the result of other forces such as selection, mutation, mating system, population structure, etc.

Both QTL and AM approaches are therefore based on LD between molecular markers and functional loci. In QTL mapping, LD is generated by the mating design while in AM, LD is a reflection of the germplasm collection under study (Stich & Melchinger, 2010). In a mapping population, LD is influenced only by recombination in the absence of segregation distortion. In AM, LD may also be influenced by other forces such as those mentioned above as well as by recombination.

The concept of LD was first described by Jennings in 1917, and its quantification (D) was developed by Lewontin in 1964 (Abdurakhmonov & Abdukarimov, 2008). The simplified explanation of the commonly used LD measure, D or D' (standardized version of D), is the difference between the observed gametic frequencies of haplotypes and the expected gametic frequencies of haplotype under linkage equilibrium.

$$D = P_{AB} - P_A P_B \quad (1)$$

Where P_{AB} is the frequency of gametes carrying allele A and B at two loci; P_A and P_B are the product of the frequencies of the allele A and B, respectively. In the absence of other forces, recombination through random mating breaks down the LD with $D_t = D_0 (1 - r)^t$, where D_t is the remaining LD between two loci after t generations of random mating from the original D_0 (Zhu et al., 2008).

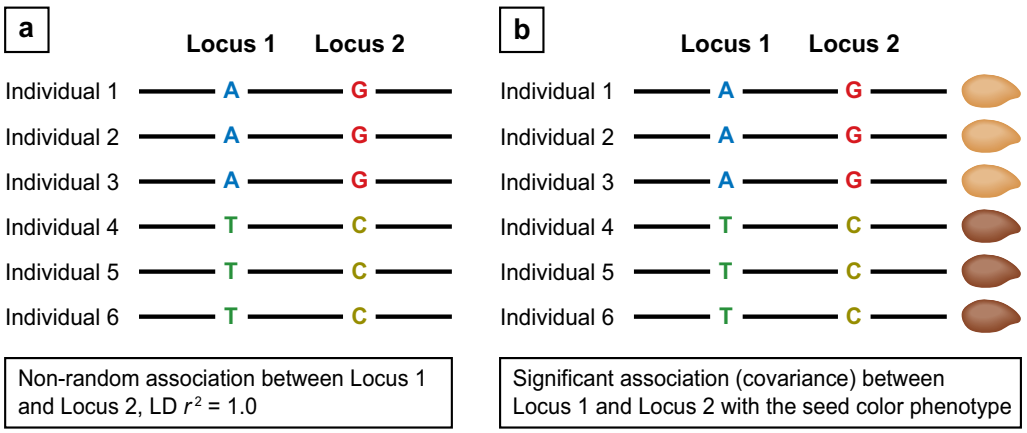


Fig. 1. Principles of linkage disequilibrium and association mapping. a. Linkage disequilibrium. Locus 1 and Locus 2 present an unusual pattern of association between alleles A-G and T-C, which deviate from Hardy-Weinberg expectations, but without any statistical correlation with a phenotype. b. Association mapping. Locus 1 and Locus 2 are in LD. Significant covariance with the seed colour phenotype is considered evidence of association.

Several statistics have been proposed for LD, and these measurements generally differ in how they are affected by marginal allele frequencies and sample sizes. Here, we introduce the two most utilized statistics for LD. Both D' (Lewontin, 1964) and r^2, the square of the correlation coefficient between two loci (Hill & Robertson, 1968), reflect different aspects of LD and perform differently under various conditions. D' only reflects the recombinational history and is therefore a more accurate statistic for estimating recombination differences, whereas r^2 summarizes both recombinational and mutational history (Flint-Garcia et al., 2003). For two biallelic loci, D' and r^2 have the following formula:

$$D' = |D| \,/\, D_{max} \tag{2}$$

$$D_{max} = \min (P_A P_b, P_a P_B) \text{ if } D > 0;$$

$$D_{max} = \min (P_A P_B, P_a P_b) \text{ if } D < 0$$

$$r^2 = D^2 \,/\, P_A P_a P_B P_b \tag{3}$$

D is limited because its range is determined by allele frequencies. D' was developed to partially normalize D with respect to the maximum value possible for the allele frequencies and give it a range between 0 and 1 (Zhu et al., 2008). The r^2 statistic has an expectation of $1/(1+4Nc)$, where N is the effective population size and c is the recombination rate, and it also varies between 0 and 1 (Hill & Robertson, 1968).

Choosing the appropriate LD statistics depends on the objective of the study. Most studies on LD in animal populations used D' to measure population-wide LD of microsatellite data (Du et al., 2007). However, D' is inflated by small sample size and low allele frequencies; therefore, intermediate values of D' are unsafe for comparative analyses of different studies and should be verified with r^2 before being used for quantification of the

extent of LD (Oraguzie et al., 2007). Although r^2 is still considered to be allele frequency dependent, the bias due to allele frequency is considerably smaller than in D' (Ardlie et al., 2002). Currently, most LD mapping studies in plants use r^2 for LD quantification because it also provides information about the correlation between markers and QTL of interest (Flint-Garcia et al., 2003; Gupta et al., 2005). Typically, r^2 values of 0.1 or 0.2 are often considered the minimum thresholds for significant association between pairs of loci and to describe the maximum genetic or physical distance at which LD is significant (Zhu et al., 2008).

3. Visualization and statistical significance of LD

Since D' and r^2 are pairwise measurements between polymorphic loci, it is difficult to obtain summary statistics of LD across a region (Gupta et al., 2005). There are two common ways to visualize the extent of LD and the genomic regions or haplotype blocks found to be in significant LD. LD scatter plots are used to estimate the rate at which LD declines with genetic or physical distance (Figure 2a). An average genome-wide decay of LD can be estimated by plotting LD values, from a data set covering an entire genome, against distance. Alternatively, the extent of LD can be estimated for a particular region carrying a gene/QTL of interest previously identified by linkage mapping. These scatter plots are useful to determine the average effective distance threshold above which significant LD (commonly 0.5 for D' and 0.1 for r^2) is expected based on the curve of a nonlinear logarithmic trend drawn through the data points of the scatter plot (Breseghello & Sorrells, 2006). Disequilibrium matrices or LD heat maps are also very useful for visualizing the linear arrangement of LD between polymorphic sites within a short physical distance such as a gene, along an entire chromosome or across the whole genome (Figure 2b) (Flint-Garcia et al., 2003). LD heat maps are colour-coded triangular plots where the diagonal represents ordered loci and the different intensity coloured pixels depict significant pairwise LD level expressed as D' or r^2. Blocks of high intensity pixels afford an easy visualization of loci in significant LD. In this figure, the larger the blue blocks of haplotypes along the diagonal of the triangular plot, the higher the level and extent of LD between adjacent loci in the blocks, meaning that there has been either limited or no recombination since the LD block formation (Abdurakhmonov & Abdukarimov, 2008). These graphical representations enable us to determine the optimum number of markers to detect significant marker-trait associations and the resolution at which a QTL can be mapped. Because LD estimation based on D' or r^2 can be sensitive to marker density, highly saturated and representative linkage groups are ideal for LD calculations.

The statistical significance of LD is typically determined using a χ^2 test of a 2 × 2 contingency table. A p-value threshold of 0.05 is often used to declare lack of independence of alleles at two loci, thus suggesting association (Gupta et al., 2005). From a 2 × 2 contingency table, the probability (P) of independence of alleles at two loci is generally calculated through a Fisher's exact test (Fisher, 1935; as cited in Gupta et al., 2005). Statistically significant LD can also be calculated using a multifactorial permutation analysis to compare sites with more than two alleles such as microsatellite markers. These statistical methods are implemented in software such as PowerMarker (Liu & Muse, 2005) and TASSEL (Trait Analysis by aSSociation Evolution and Linkage) (Bradbury et al., 2007).

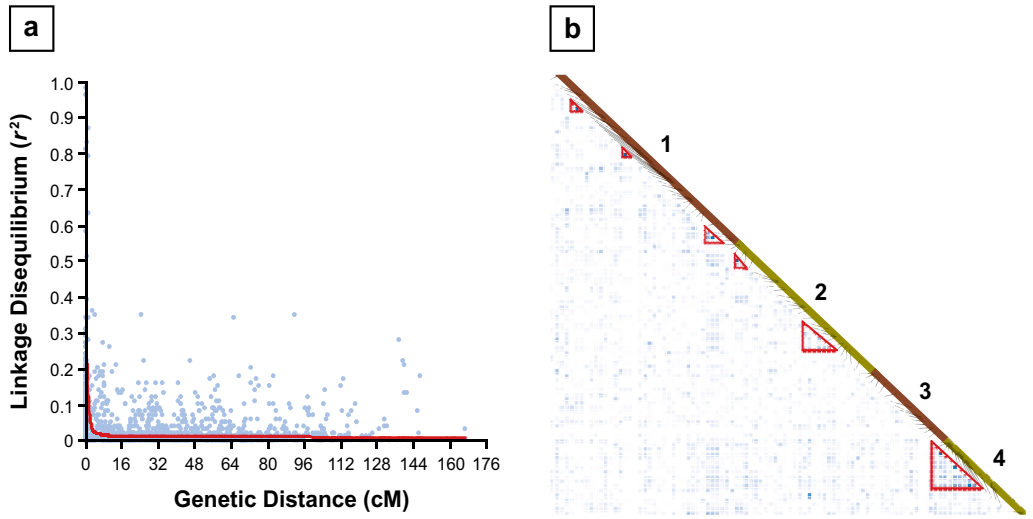


Fig. 2. Visualization of linkage disequilibrium in flax (*Linum usitatissimum* L.). a. Scatter plot of LD decay (r^2) against genetic distance (cM), representing a measure of an average genome-wide LD. b. Heatmap of LD variation between pairwise polymorphic loci of four linkage groups. Blocks in significant LD are highlighted by red triangles. LD distribution is heterogeneous within and between linkage groups.

4. LD variation as an effect of biological factors

4.1 Recombination

Several biological factors influence LD strength and its distribution across genomes. Many regions of the human genome display rates of recombination that differ significantly from the genome average recombination rate of 1 cM/Mb (Arnheim et al., 2003). These regions have been called "hotspots" and "coldspots" for high and low recombination rates, respectively. LD is strongly influenced by localized recombination rate and is correlated with other associated factors such as GC content and gene density (Dawson et al., 2002). In principle, local sequence features can affect LD directly and indirectly. For example, GC-rich sequences may be associated with higher rates of recombination and/or mutation, two phenomena that could directly lower surrounding levels of LD. Furthermore, in some protein-coding sequences, changes created by recombination or mutation may affect the fitness of an individual, and these sequences could be indirectly associated with unique patterns of LD as a consequence of natural selection (Smith et al., 2005).

Because LD is broken down by recombination, and recombination is not distributed homogeneously across the genome, blocks of LD are expected. Also, differences in LD between micro chromosomes and macro chromosomes have been reported (Stapley et al., 2010) as well as intra-chromosomal variation, where centromeric regions showed higher levels of LD. Teo et al. (2009) conducted a comprehensive analysis of genomic regions with different patterns of LD to unravel the consequences of this patterning for AM in human populations. Plant genomes have revealed similar general conclusions with regards to LD distribution. Inter-chromosomal LD variation has been reported in barley (*Hordeum vulgare*), maize (*Zea mays*), tomato (*Solanum lycopersicum*) and bread wheat (*Triticum aestivum*)

(Malysheva-Otto et al., 2006; Robbins et al., 2011; Yan et al., 2009; Zhang et al. 2010), where it varied between less than 1 cM to more than 30 cM ($r^2 > 0.1$). As a consequence, investigation of LD variation at the genome and chromosome scale to accurately estimate marker density for each chromosome is required to provide insights to the most cost-effective AM approach.

4.2 Mating system

The mating system has profound effects on LD (Myles et al., 2009). Selfing reduces opportunities for effective recombination because individuals are more likely to be homozygous than in outcrossing species (Flint-Garcia et al., 2003). In self-pollinated species such as rice (*Oryza sativa*), Arabidopsis (*Arabidopsis thaliana*) and wheat (*Triticum aestivum*) (Garris et al., 2005; Nordborg, 2000; Zhang et al. 2010), LD extends much further as compared to outcrossing species such as maize (*Zea mays*), grapevine (*Vitis vinifera*) and rye (*Secale cereale*) (Li et al., 2011b; Myles et al., 2009; Tenaillon et al., 2001). As a result, genetic polymorphisms tend to remain correlated, and LD is expected to be maintained over long genetic or physical distances (Gaut & Long, 2003). However, because LD declines more rapidly in outcrossing plant species than self-pollinated plants, a higher resolution is expected, enabling more accurate fine mapping and potentially facilitating the cloning of candidate genes. A detailed review of LD decay in self-pollinated and outcrossing species can be found in Flint-Garcia et al. (2003) and Abdurakhmonov & Abdukarimov (2008).

4.3 Germplasm

The germplasm plays a key role in LD variation because the extent of LD is influenced by the level of genetic diversity captured by the population under consideration. In general, the larger the genetic variation, the faster the LD decay, a direct consequence of the broader historical recombination. The population sample effect is evident in maize (*Zea mays*) where LD decays within 1 kb in landraces, approximately doubles (~ 2kb) in diverse inbred lines and can extend up to several hundred kb in commercial elite inbred lines (Jung et al., 2004). Tenaillon et al. (2001) investigated sequence diversity at 21 loci on chromosome 1 in a diverse group of maize germplasm, including exotic landraces and US accessions. An average LD decay was determined to occur within 400 bp ($r^2 = 0.2$), but extended up to 1000 bp in a group of US inbred lines. In Michigan local Arabidopsis populations, LD decay varied within 50 kb up to 50-100 cM. The latter was explained as a genetic bottleneck or founder effect, which reduced dramatically the genetic variation (Nordborg et al., 2002). In cotton (*Gossypium hirsutum*), the genome-wide average LD ($r^2 \leq 0.1$) declined to 10 cM in landraces, but was up to 30 cM in varieties (Abdurakhmonov et al., 2008). Myles et al. (2011) studied LD variation in over 1000 samples of domesticated grape (*Vitis vinifera*) and its wild relatives, reporting a rapid LD decay, even greater than in maize, as result of a weak domestication bottleneck followed by thousands of years of widespread vegetative propagation.

Estimates of genome-wide average LD decay may not reflect LD patterns between different populations of the same species. Each of these populations should be explored independently for the extent of LD in order to conduct successful association mapping studies (Abdurakhmonov & Abdukarimov, 2008). Taking into account these three important biological factors, an obvious question is whether an increased or decreased level of LD is

favourable in AM? Populations with either rapid or slow LD decay can be useful in AM, depending on the purposes of the study. Thus, populations with narrow genetic diversity and long extent of LD are amenable to coarse mapping with fewer markers requiring fine mapping in more genetically diverse populations, assuming that the causal genetic factors are sufficiently similar across different germplasm groups.

5. LD variation as effect of evolutionary factors

5.1 Selection

Initial interest in LD arose from questions surrounding the *modus operandi* of natural selection. Simply stated, if alleles at two loci are in LD and they both affect positively reproductive fitness, the response to selection at one locus might be accelerated by selection affecting the other (Slatkin, 2008). Thus, positive selection will increase LD between and in the vicinity of the selected loci, a phenomenon known as genetic hitchhiking (Maynard Smith & Haigh, 1974; as cited in Slatkin, 2008). Even if the second locus is selectively neutral, the selection applied over the first will increase LD between them. The LD level between the two loci will remain constant over time depending on the genetic distance, the recombination rate and the effective population size (N). In contrast, if both loci are maintained by balancing selection, then LD can persist indefinitely (Lewontin, 1964). Nonetheless, LD should be higher in loci affected by positive selection because a strong positive selection limits genetic diversity as opposed to a balancing selection which tends to maintain or increase polymorphism. In general, disease resistance genes in plants (*R*-genes) are affected by balancing selection with low intragenic LD and rapid decay (Yin et al., 2004), which could facilitate fine mapping of disease resistance genes providing high marker saturation. Artificial selection also has dramatic effects on LD. Mosaics of large LD blocks are observed, especially in regions carrying agronomic-related genes. Domestication bottlenecks followed by strong selection for specific environments and end-use traits have modified the genome architecture in many crops reducing genetic diversity and creating population structure, which may be the main factor affecting the power of AM.

5.2 Population structure

Selection affects the genome and LD in locus-specific manner. In contrast, population structure affects LD throughout the genome. Consequently, genome-wide patterns of LD can help to understand the history of changes in populations (Slatkin, 2008). However, the power of AM can be strongly reduced as a consequence of population structure (Balding, 2006). Population structure occurs from the unequal distribution of alleles among subpopulations of different ancestries. When these subgroups are sampled to construct a panel of lines for AM, the intentional or unintentional mixing of individuals with different allele frequencies creates LD. Significant LD between unlinked loci results in false-positive associations between a marker and a trait. The effect is obvious in the following case. Suppose that one subpopulation is fixed for A and B alleles at two loci whereas another is fixed for a and b. Any mixture of individuals from the two subpopulations would contain only AB and ab haplotypes, implying that they are in perfect LD, when in fact there is no LD in either subpopulation (Slatkin, 2008). By definition, polymorphisms at two or more loci must exist to estimate the level of LD. In the above example, both loci are monomorphic in their respective subpopulations. However, when individuals are mixed, in the newly created artificial single population, false

polymorphisms and consequently significant but spurious LD is observed. Thornsberry et al., (2001) reported significant associations between polymorphisms at the maize *Dwarf8* gene and variation in flowering time, but they also stated that up to 80% of the false positive associations resulted from population structure. The occurrence of spurious associations is markedly higher in adaptation-related genes because they show positive correlations with the environmental variables under which they have evolved, and, as a result, the genomic regions carrying these genes could present stronger population differentiation. Several statistical models take into account the potential effect of population structure. Commonly used algorithms are those of Pritchard & Rosenberg (1999) implemented in the software STRUCTURE (Hubisz et al., 2009; Pritchard et al., 2000). Other methods are based on Principal Component Analysis (PCA) (Price et al., 2006), and Principal Coordinate Analysis and Modal Clustering (PCoA-MC) (Reeves & Richards, 2009).

5.3 Genetic drift, population bottleneck and gene flow

The effect of genetic drift in a small population results in the consistent loss of rare allelic combinations which increases LD level (Flint-Garcia et al., 2003). Genetic drift can create LD between closely linked loci. The effect is similar to taking a small sample from a large population. Even if two loci are in linkage equilibrium, sampling only few individuals can create LD (Slatkin, 2008).

LD can also be created in populations that have experienced a reduction in size (called a bottleneck) with accompanying extreme genetic drift (Dunning et al., 2000; as cited in Flint-Garcia et al., 2003). After a bottleneck, some haplotypes will be lost; generally resulting in increased LD. Subsequent bottlenecks will further contribute to augment LD by increasing the effect of genetic drift. Colonizing species undergo repeated bottlenecks, and many models of the history of hominids assume the occurrence of a bottleneck when modern humans first left Africa (Noonan et al., 2006). Several studies of humans have argued that long distance LD in humans is the result of this early bottleneck in human history (Schmegner et al., 2005). In plants, comparisons with wild ancestors indicate that, in maize, approximately 80% of the allele richness has been lost as a consequence of domestication bottlenecks (Wright & Gaut, 2005) while this number is 40-50% in sunflower (Liu & Burke, 2006) and 10-20% in rice (Zhu et al., 2007). Gene flow introduces new individuals or gametes with different ancestries and allele frequencies among populations. If selection maintains differences in allele frequencies at two or more loci among subpopulations, LD in each subpopulation will persist (Slatkin, 1975; as cited in Slatkin, 2008), but generally when random mating and recombination take place, LD caused by gene flow eventually breaks down.

Factors such as genetic drift, population bottlenecks and gene flow can contribute to generating artificial LD and negatively impact the ability to use LD in AM for the precise localization of QTL. In general, any biological or evolutionary forces that contribute to an increase of LD beyond that expected by chance in an "ideal" population will result in false-positive associations (Gaut & Long, 2003).

6. Approaches for AM

Many methodologies have been developed and are widely used for AM in humans (Schulze & McMahon, 2002), and several are perfectly applicable without change or with case-to-case

modifications for a wide range of organisms, including plants. The methods to study marker-trait association using LD may differ for discrete and quantitative traits (Nielsen & Zaykin, 2001). Here, we will examine several approaches: Multiparent Advanced Generation Intercross (MAGIC), Case-control (CC), Transmission Disequilibrium Test (TDT) and other approaches that incorporate corrections for population structure such as genomic control (GC) and structured association (SA).

6.1 Multiparent Advanced Generation Intercross (MAGIC)

MAGIC is an extension of the advanced intercross method in which an intermated mapping population is created from multiple founder lines. A Recombinant Inbred Line (RIL) population is created from multiple founder lines, in which the genome of the founders are first mixed by several rounds of mating, and subsequently inbred to generate a stable panel of inbred lines. The larger number of parental accessions increases the allelic and phenotypic diversity over traditional RILs, potentially increasing the number of QTL that segregate in the population. The successive rounds of recombination cause LD to decay, thereby increasing the precision of QTL location (Mackay & Powell, 2007). In both crops and animals, the MAGIC design has the ability to capture the majority of the variation available in the gene pool. Although it might take several years before these populations are suitable for fine mapping, they are relatively inexpensive to develop and their value as mapping resources increases with each generation (Mackay & Powell, 2007). In plants, MAGIC can be used to combine coarse mapping with low marker densities on lines derived from an early generation, with fine mapping using lines derived from a more advanced generation and a higher marker density. Regardless of the generation used, LD decay remains the critical factor determining the mapping resolution.

6.2 Case-control (CC)

The classical methodology and design of AM is the "case and control" (CC) approach. If a mutation increases disease susceptibility, then we can expect it to be more frequent among affected individuals (cases) than among unaffected individuals (controls). The essential idea behind CC-based AM is that markers close to the disease mutation may also have allele frequency differences between cases and controls if there is LD between the marker locus and the "susceptibility" mutations (Schulze & McMahon, 2002). For accurate mapping, this design requires an equal number of unrelated and unstructured *case-control* samples. The Pearson χ^2 test, Fisher's exact test or Yates continuity correction can be used to compare allele frequencies and detect association between a phenotype and a marker (Abdurakhmonov & Abdukarimov, 2008). The CC tests are sensitive to overall population LD between a marker and a locus affecting the trait. As previously discussed, LD can exist between unlinked loci, meaning that strong marker-trait association is not necessarily evidence for physical proximity between a marker and the gene affecting the phenotype. As a consequence, the CC approach is highly sensitive to population structure (Schulze & McMahon, 2002). To efficiently eliminate the confounding effects caused by population structure, Spielman et al. (1993) developed the Transmission Disequilibrium Test (TDT).

6.3 Transmission Disequilibrium Test (TDT)

The ability to map QTL in collections of breeding lines, landraces or samples from natural populations has merit. In these populations, LD often decays more rapidly than in

controlled crosses, enabling fine mapping. The challenge is to distinguish the effects of population subdivision from LD caused by linkage (syntenic LD). A robust method to test for this partitioning is the TDT (Spielman et al., 1993) that permits the detection of linkage in the presence of disequilibrium. Neither linkage alone nor disequilibrium alone (non syntenic LD) will generate a positive result in a TDT. As a consequence, the TDT is a robust method to control false positives (Mackay & Powell, 2007). In brief, TDT compares the transmission versus the non transmission of alleles to the offspring using a χ^2 test, assuming a linkage between a marker and a trait. The TDT design requires genotyping of markers from three individuals: one heterozygous parent, one homozygous parent and one affected offspring. In the absence of linkage between QTL and marker, the expected ratio of transmission to non transmission is 1:1 (Nielsen & Zaykin, 2001). In the presence of linkage, it is distorted to an extent that depends on the strength of LD between the marker and the QTL. In addition, the power of the association will depend on the effectiveness of selection of extreme progeny in driving segregation away from expectation (Mackay & Powell, 2007).

The initial TDT approach did not address the cases of multiallelic markers, multiple markers, missing parental information, large pedigrees and complex quantitative traits (Schulze & McMahon, 2002). A variety of extensions of the TDT approach have been developed and applied to resolve multiallelic marker issues (i.e., GTDT, ETDT, MCTm); reviewed by Schulze & McMahon (2002).

In crops, parental and progeny lines are often separated by several generations of gametogenesis rather than one, as is often the case of human studies. For this reason, the TDT, while still valid, may be less robust because the breeding process may result in increased segregation distortion (Mackay & Powell, 2007).

6.4 Other approaches

Population structure arising from recent migration, population admixture and artificial selection will generate non syntenic LD. Assuming that such population structure has a similar effect on all loci, a random set of markers can be used to statistically assess the extent with which population structure is responsible for non syntenic LD (Stich & Melchinger, 2010). This is the basis of genomic control (GC). For example, for a case-control analysis of candidate genes, the GC approach computes χ^2 test statistics for independence for both null (random) and candidate loci. An average χ^2 of null loci greater than 1.0 indicates the presence of significant structure. By using the magnitude of the χ^2 test observed at the null loci, a multiplier is derived to adjust the critical value for significance tests for candidate loci (Mackay & Powell, 2007). By contrast, structure association (SA) analysis developed by Pritchard et al. (2000), first uses a set of random markers to estimate population structure (*Q*-matrix), and then incorporates this estimate into a general linear model (GLM) analysis which enables correction for false associations. Yu et al. (2006) developed a new methodology, the mixed linear model (MLM), which incorporates both population structure and familial relatedness or so-called "kinship" (*K*-matrix). To perform MLM: (1) a *Q*-matrix is generated using for example, STRUCTURE; (2) the pairwise relatedness coefficients between individuals of a germplasm collection (*K*-matrix) is estimated using for example, SpaGeDi software (Hardy & Vekemans, 2002); and (3) both *Q*- and *K*-matrices are used in AM to control spurious associations. Studies conducted in human, Arabidopsis and bread

wheat (Raman et al., 2010; Yu et al., 2006; Zhao et al., 2007a) have demonstrated the effectiveness of the MLM approach over the GLM.

Another type of mixed model used in AM incorporates PCA instead of the *Q*-matrix. Promising results as a fast and effective way to identify population structure have been reported (Price et al., 2006). The PCA-based MLM model is computationally effective as compared to the *Q*-matrix estimated from STRUCTURE. Also, STRUCTURE has been found to overestimate the "true" number of subpopulations under particular scenarios (Evanno et al., 2005).

7. AM studies in plants

Some of the first LD mapping studies in plants were done in maize (*Zea mays*) (Bar-Hen et al., 1995), rice (*Oryza sativa*) (Virk et al., 1996) and oat (*Avena sativa*) (Beer et al., 1997). Bar-Hen et al. (1995) and Virk et al. (1996) predicted the association of quantitative traits using RAPD and isozymes markers, respectively. Beer et al. (1997) associated 13 QTL with RFLP loci using 64 oat varieties and landraces. In these studies, a low number of genome-wide distributed markers were assessed without considering the population structure. The first empirical candidate gene association taking into account background molecular markers to correct for population structure was performed in maize looking at the D8 locus and its association with flowering time (Pritchard, 2001). In Arabidopsis, most of the AM studies focused on providing proof of concept, identification of QTL involved in adaptation and detection of additional alleles to supplement other mutagenesis approaches (Ersoz et al., 2007). Aranzana et al. (2005) performed the first attempt at a genome wide association study (GWAS) in Arabidopsis, reporting previously known flowering time and three known pathogen-resistance genes. GWAS refers to the use of many markers that span an entire genome to identify functional common variants in LD with at least one of the genotyped markers. Numerous research papers focusing on LD and AM have since been published on more than a dozen plant species. These studies have been reviewed by Gupta et al. (2005) and more recently by Zhu et al. (2008).

In the last five years, plant AM studies have expanded because of advances in sequencing technologies which enable more efficient and cost-effective development of a large number of molecular markers such as Single Nucleotide Polymorphisms (SNPs). In Arabidopsis, new studies have been carried out aiming to dissect downy mildew resistance genes and climate-sensitive QTL, with special efforts focused on the understanding of adaptative variation (Li et al., 2010; Nemri et al., 2010). The first applied a CG approach, and the second a GWAS based on no fewer than 213,497 SNPs. In maize, recent studies dissected the quantitative genetic nature of the northern leaf blight (NLB) resistance, southern leaf blight (SLB) resistance and leaf architecture, scanning the genome using ~ 1.6 million SNPs (Kump et al., 2011; Poland et al., 2011; Tian et al., 2011). Poland et al. (2011) identified several loci with small additive effects carrying candidate genes related to plant defense, including receptor-like kinase genes. Kump et al. (2011), from the same research group, identified 32 QTL with predominantly small additive effects related to SLB resistance. Similarly, Tian et al. (2011) demonstrated that the genetic architecture of leaf traits is dominated by small effects and that the *liguleless* genes have contributed to more upright leaves. Currently, whole genome scanning has moved beyond Arabidopsis and maize to other species such as rice and barley. Huang et al. (2010) uncovered the genetic basis of 14 rice agronomic traits

based on ~ 3.6 million SNPs. The loci identified through GWAS explained ~ 36% of the phenotypic variance, on average. In barley, GWAS of 15 morphological traits identified one putative anthocyanin pathway gene, *HvbHLH1*, carrying a deletion resulting in a premature stop codon and which was diagnostic for the absence of anthocyanin in the germplasm studied (Cockram et al., 2010). Efforts towards understanding adaptation-related genes have been undertaken in wheat. Raman et al. (2010) applied GWAS in order to identify genetic factors associated with aluminium resistance, one of the most restrictive abiotic stresses on acid soils worldwide. The study confirmed previously identified loci and identified putative novel ones. Subsequently, Rousset et al. (2011) studied the genetic nature of flowering time in wheat to investigate the effect of candidate genes on flowering time. The *Vrn-3* gene explained a high percentage of the phenotypic variation of earliness followed to a lesser extent by *Vrn-1*, *Hd-1* and *Gigantea* (*GI*). In *Brassica napus*, several seed oil related loci were identified, with a few corresponding to previously reported genomic regions associated with oil variation (Zou et al., 2010). In tetraploid alfalfa (*Medicago sativa*), 15 SSR markers showed strong association with yield in different environments (Li et al., 2011a). In sugar beet (*Beta vulgaris*), genetic variation of six agronomic traits was dissected using GWAS, identifying several QTL with major effects and others with epistatic effects (Würschum et al., 2011). Thus, LD mapping, considered a few years ago as an emerging tool in plant genomics, has recently been shown to be a powerful method to dissect complex traits in crops. Table 1 summarizes these and other recently published AM studies in plants. Earlier publications are summarized elsewhere (Gupta et al., 2005; Zhu et al., 2008).

8. Benefits and limitations of AM

The potential high resolution in localizing a QTL controlling a trait of interest is the primary advantage of AM as compared to linkage mapping (Figure 3). AM has the potential to identify more and superior alleles and to provide detailed marker data in a large number of lines which could be of immediate application in breeding (Yu & Buckler, 2006). Furthermore, AM uses breeding populations including diverse and important materials in which the most relevant genes should be segregating. Complex interactions (epistasis) between alleles at several loci and genes of small effects can be identified, pinpointing the superior individuals in a breeding population (Tian et al., 2011). Sample size and structure do not need to be as large as for linkage studies to obtain similar power of detection. Finally, AM has the potential not only to identify and map QTL but also to identify causal polymorphisms within a gene that are responsible for the difference between two phenotypes (Palaisa et al., 2003).

AM suffers from some limitations such as when the trait under consideration is strongly associated with population structure. Most traits under local adaptation or in balancing selection in different populations may be thus affected (Stich & Melchinger, 2010). When statistical methods to correct for population structure are applied, the differences between subpopulations are disregarded when searching for marker-trait associations. Therefore, all polymorphisms responsible for the phenotypic differences between subpopulations remain undetected, thus underpowering AM. LD mapping often requires a large number of markers for genotyping in GWAS. The number of markers depends in large part on the genome size and the expected LD decay; linkage mapping generally requires fewer markers to detect significant QTL. A high density of markers can only be achieved through the

Species	Germplasm	Trait	Marker system	Reference
Arabidopsis	Diverse accessions	Flowering time/pathogen resistance	Sequences	Aranzana et al. (2005)
	Diverse accessions	Multiple traits	SSRs/SNPs	Ersoz et al. (2007)
	Natural accessions	Flowering time	SNPs	Brachi et al. (2010)
	Diverse accessions	Climate-sensitive QTL	SNPs	Li et al. (2010)
	Landraces	Downy mildew	SNPs	Nemri et al. (2010)
Maize	Inbred lines	Aluminum tolerance	SNPs	Krill et al. (2010)
	Inbred lines	Drought tolerance	SNPs	Lu et al. (2010)
	Inbred lines	Northern leaf blight	SNPs	Poland et al. (2011)
	Inbred lines	Southern leaf blight	SNPs	Kump et al. (2011)
	Inbred lines	Leaf architecture	SNPs	Tian et al. (2011)
Teosinte	Landraces	Domestication-related genes	SNPs	Weber et al. (2009)
Wheat	Cultivars	Kernel size, milling quality	SSRs	Breseghello & Sorrells, (2006)
	Diverse accessions	Aluminum resistance	DArT	Raman et al. (2010)
	Breeding lines	Stem rust resistance	DArT	Yu et al. (2011)
	Diverse accessions	Flowering time	SNPs	Rousset et al. (2011)
Barley	Inbred lines	Growth habit	SNPs	Rostoks et al. (2006)
	Cultivars	Anthocyanin pigmentation	SNPs	Cockram et al. (2010)
	Breeding lines	Winterhardiness	SNPs	Von Zitzewitz et al. (2011)
Oat	Diverse cultivars	Agronomic and kernel quality traits	AFLPs	Achleitner et al. (2008)
Rice	Landraces	Heading date, plant height and panicle length	SSRs	Wen et al. (2009)
	Landraces	Multiple agronomic traits	SNPs	Huang et al. (2010)
Canola	Diverse accessions	Leaf traits, flowering time and phytate content	AFLPs	Zhao et al. (2007b)
	Diverse accessions	Oil content	SSRs	Zou et al. (2010)
Soybean	Breeding lines	Iron deficiency chlorosis	SSRs	Wang et al. (2008)
Cotton	Diverse cultivars	Fiber quality	SSRs	Abdurakhmonov et al. (2009)
Peanut	Diverse accessions	Seed quality traits	SSRs-SNPs	Wang et al. (2011)
Sugar beet	Inbred lines	Sugar content and yield	SSRs	Stich et al. (2008)
	Inbred lines	Multiple traits	SNPs	Würschum et al. (2011)
Alfalfa	Cultivars	Biomass yield and stem composition	SSRs	Li et al. (2011a)

SNPs: Single Nucleotide Polymorphisms; SSRs: Simple Sequence Repeats; DArT: Diversity Arrays Technology; AFLPs: Amplified Fragment Length Polymorphisms.

Table 1. Association mapping studies in plants.

development of an integrated genotyping by sequencing (GBS) platform. Thus, the analysis of cost-benefit must be conducted in the light of the real impacts that such investments will have in the future market appreciation of that plant species. Alternative approaches such as linkage mapping and CG could be feasible for other studied traits. The power of AM to detect an association is influenced by allele frequency distribution at the functional polymorphism level. The results of empirical studies suggest that a high percentage of alleles are rare (Myles et al., 2009). Rare alleles cannot be evaluated adequately because, by definition, they are present in too few individuals and consequently lack resolution power. As a consequence, an important piece of heritability remains undetected. For such rare alleles, linkage mapping may be used because correlation between population structure and phenotypes can be broken, and allele frequencies can be inflated to enhance the power of mapping (Stich & Melchinger, 2010). In this regard, several studies have combined linkage mapping and LD mapping, a methodology known as "nested association mapping", which reduces spurious associations caused by population structure, particularly for traits strongly affected by local geographic patterns (Brachi et al., 2010; Poland et al., 2011). With the growing interest in finding the missing heritability not accounted for by common alleles (Asimit & Zeggini, 2010), several new association analysis methods for rare variants are being proposed, with some important advances in complex trait dissection (Li & Leal, 2008).

9. Computer programs for AM

A variety of software packages are available for AM, and it can be inferred from the previous sections that LD studies are computationally demanding. Thus, newer and more powerful programs are constantly under development. TASSEL is a commonly used software for LD mapping in plants, frequently updated with newly developed methods. Current examples include the GLM and the multiple regression models combined with the estimates for false discovery rate. TASSEL can also be used for calculation and graphical display of LD statistics, analysis of population structure using PCA, and tree plots of genetic distance. Although TASSEL can handle both SSR and SNP markers, the latest version only accepts SNPs. For SSR analysis, users must continue with TASSEL v. 2.1. Alternatively, GenStat offers traditional statistical analyses as well as linkage and AM analyses for SSRs.

GenStat performs structure analysis based on PCA, LD decay and single trait association analysis using PCA-based MLM. Version 14 was recently released and can be downloaded for non profit purposes from http://www.vsni.co.uk/2011/asides/genstat-14-released/. Gupta et al. (2005) and Excoffier & Heckel (2006) comprehensively reviewed the most common software for population genetics and LD mapping analyses but the majority of them can only handle a few thousand marker loci. Progress in sequencing technologies has solved the past issue of genotyping large populations with high marker densities and software development has also moved quickly. Nowadays, the main issue is the time required for processing large data sets and the availability of powerful statistical models to adjust for multiple testing. JMP Genomics v.5 is a Windows based program that offers several solutions for handling large SNP data sets (http://www.jmp.com/software/genomics). Among its main characteristics, JMP Genomics is capable of handling data sets as large as 1.5 million SNPs for 15,000 samples on a 32-bit desktop work station using CG or GWA. It also corrects for relatedness and population structure using association tests, and calculates identical by descent (IBD), identical by state (IBS) and allele-sharing individual relationship matrices. Interactive triangular plots

and zooming features permit visualization of LD blocks. Association between SNPs and multiple traits can be tested separately or jointly, while adjusting for covariates. JMP Genomics 5 also simplifies the analysis of rare and common variants, and includes features for high quality graphs and figures.

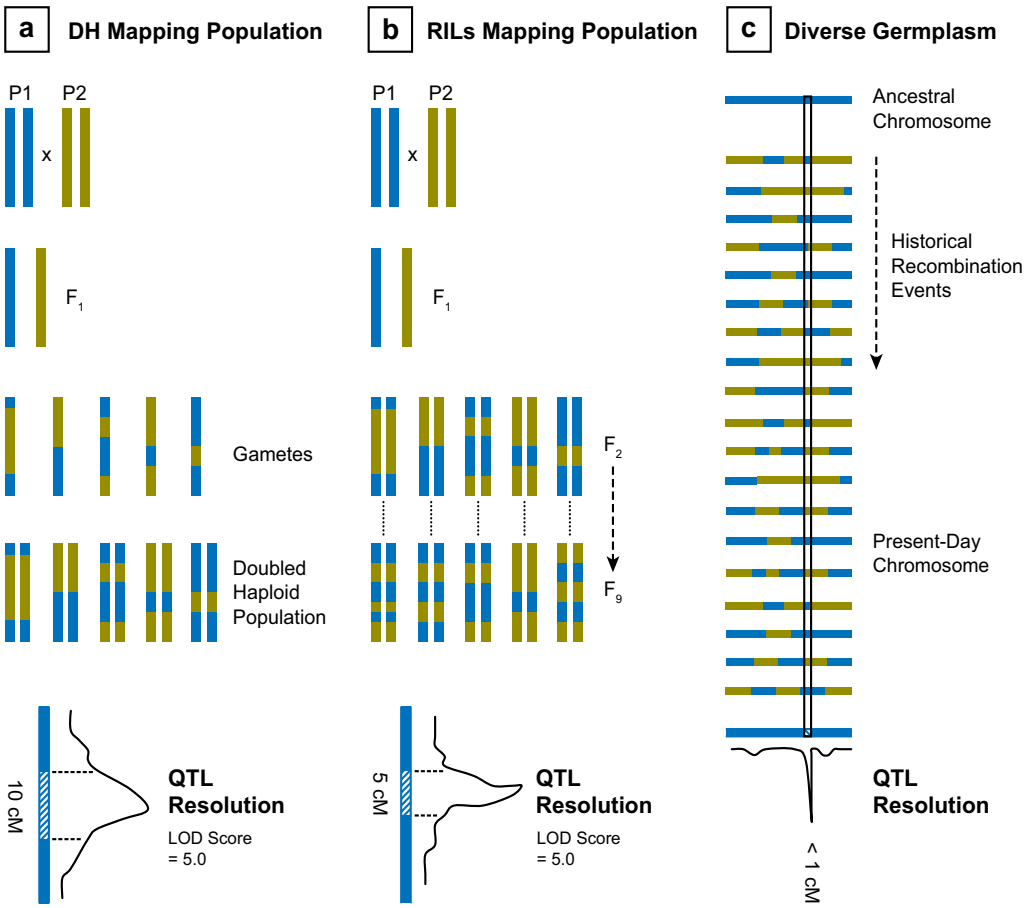


Fig. 3. Comparison of mapping resolution between linkage mapping and AM. a. A Doubled Haploid (DH) mapping population. b. A Recombinant Inbred Line (RIL) mapping population. c. A collection of diverse germplasm. **a** and **b** present low QTL resolution as a consequence of few meiosis events accumulated; **c** presents a high QTL resolution because a larger number of recombination events have accumulated during the population history.

Similar applications can be found in GenAMap software, which incorporates visualization strategies for structured AM (http://cogito-b.ml.cmu.edu/genamap/). It has a processing capacity of 1 million SNPs in approximately 1 hour. The analysis is performed on a remote cluster complete with complex parallelization schemes to optimize run-time efficiency. GenAMap gives an overview of the association results through a heatmap view where SNPs are plotted against a network of candidate genes, shows interactions between genes, integrates the association strengths of the genes to SNPs in the genome, and creates a tree view of structured genes to explore and identify functional relevant branches of the tree that

are associated with a genomic region. Although GenAMap was primarily developed for human diseases, it can be applied to plant AM as well.

PLINK software v. 1.07 (Purcell et al., 2007; http://pngu.mgh.harvard.edu/purcell/plink/) is an open source C/C++ GWAS tool set. With PLINK, large data sets comprising hundreds of thousands of SNPs and individuals can be readily manipulated and analyzed. PLINK offers five main characteristics. Data management is a simple interface for reordering, recording and filtering genotypic information. Summary statistics to determine the randomness of genotyping failure highlights the test of missingness on a simple haplotypic case-control test. Population stratification is measured on the basis of a genome average proportion of alleles sharing identical by state (IBS) between any two individuals. PLINK offers tools to cluster individuals into homogeneous subsets to identify potential outlier individuals causing genotyping or pedigree errors, and to incorporate this stratification in GWAS. Association analyses include CC, stratified analysis, TDT, QTDT, sib TDT and correction for multiple tests. Table 2 summarizes these and other software based on their analytical focus.

10. Future perspectives of AM

Large scale GWAS have already been carried out in plants and many more are in progress. The technological problem of efficiently genotyping 1 million or more SNPs has been solved, and the cost of genotyping continues to decline (Slatkin, 2008). With this increased resolution of LD patterns, the study of crop history will shift in focus from understanding the average history of populations to understanding the history of different genomic regions in depth. GWAS will not be limited to the identification of QTL but will also provide in depth understanding of the genomic changes that have shaped crop plants as a consequence of domestication and selection. Such information will translate into improved design of breeding populations and germplasm collections capturing adaptative variation.

Design and implementation of genotyping assays is no longer time-consuming or expensive. To fully exploit and benefit from the large amount of achievable genotyping data, care must be given to proper and powerful experimental design (Myles et al., 2009). Because LD mapping often involves a relatively large number of diverse accessions, phenotypic data collection with adequate replications across multiple years and locations can be challenging. Efficient field design, appropriate statistical methods and consideration for QTL × environmental interactions should be explored to increase the mapping power, particularly if field conditions are not homogeneous. Reducing errors associated with phenotypic measurements remains a priority.

One of the limitations of LD mapping is that it provides little insight into the mechanistic basis of LD detected, so that genomic localization and cloning of genes based on LD may not be always straightforward. This limitation occurs because strong LD is sometimes the result of a recent occurrence of LD (recent mutations) rather than a close physical linkage between two loci exhibiting LD (Gupta et al., 2005). As a consequence, we anticipate increased usage of nested AM, because it has the power to simultaneously capture information about the linkage of the markers and the degree of LD historically created. Linkage mapping and LD mapping are complementary and their successful combination has been demonstrated in plant systems (Brachi et al., 2010; Poland et al., 2011). As mentioned earlier, AM is one of

Software	Focus	Description	Website
STRUCTURE 2.3	Population structure	Compute a MCMC Bayesian analysis to estimate the proportion of the genome of an individual originating from the different inferred populations	http://pritch.bsd.uchicago.edu/software.html
BAPS 5.0	Population structure	Compute Bayesian analysis to estimate the proportion of the genome of an individual and assign individuals to genetic clusters by either considering them as immigrants or as descendents from immigrants	http://web.abo.fi/fak/mnf/mate/jc/software/baps.html
mStruct	Population structure	Detection of population structure in the presence of admixing and mutations from multi-locus genotype data. It is an admixture model which incorporates a mutation process on the observed genetic markers	http://www.cs.cmu.edu/~suyash/mstruct.html
LDheatmap	LD	R environment software for LD estimation (r^2) displayed as heatmap plots using SNPs	http://www.jstatsoft.org/v16/c03
LDhat 2.1	Recombination rates and LD	R environment software for LD estimation and identification of hotspots using a Bayesian reversible jump MCMC scheme for SNPs	http://www.stats.ox.ac.uk/mcvean/ldhat.html
MIDAS	LD	Compute LD heatmaps for D' and r^2 providing both inter loci and inter allelic LD variation	http://www.genes.org.uk/software/midas/
Arlequin 3.5	Genetic analysis and LD	Hierarchical analysis of genetic structure (AMOVA), LD for D' and r^2. Version 3.5 incorporates a R function to parse XML output files to produce publication quality graphics	http://cmpg.unibe.ch/software/arlequin35/

Software	Focus	Description	Website
Haploview 4.2	Haplotype analysis and LD	LD and haplotype block analysis, haplotype population frequency estimation, single SNP and haplotype association tests, permutation testing for association significance	http://www.broad.mit.edu/mpg/haploview/
GGT 2.0	Genetic analysis, LD and AM	Compute genetic distance based on Jaccard similarity, dendrograms are displayed using a Neighbour-Joining algorithm. Displays LD heatmaps and LD scatter plots for D' and r^2 and performs simple AM analysis	http://www.plantbreeding.wur.nl/UK/software_ggt.html
SVS 7	Stratification, LD and AM	Estimate stratification, LD, haplotypes blocks and multiple AM approaches for up to 1.8 million SNPs and 10,000 samples	http://www.goldenhelix.com
TASSEL	Stratification, LD and AM	SSR markers, GLM and MLM methods	http://www.maizegenetics.net
GenStat	Stratification, LD and AM	SSR markers, GLM and MLM-PCA methods	http://www.vsni.co.uk/
JMP genomics	Stratification, LD and structured AM	SNPs, CG and GWAS, analysis of common and rare variants	http://www.jmp.com/software/genomics
GenAMap	Stratification, LD and structured AM	SNPs, tree of functional branches, multiple visualization tools	http://cogito-b.ml.cmu.edu/genamap
PLINK	Stratification, LD and structured AM	SNPs, multiple AM approaches, IBD and IBS analyses	http://pngu.mgh.harvard.edu/purcell/plink/

Table 2. List of software used in LD and AM.

many applications of LD. With the increasing availability of molecular markers, it is now feasible to scan a genome to identify signatures of selection (both positive and balancing selection). This approach, known as population genomics, simultaneously studies thousands

of marker loci distributed across the genome to better understand the roles that evolutionary processes have played in the current pattern of genetic variation across populations (Luikart et al., 2003). Among different approaches reviewed by Oleksyk et al. (2010), LD can be used to Identify loci that have been targets of selection. Strong positive selection quickly increases the frequency of an advantageous allele, resulting in linked loci remaining in unusually strong LD with that allele in the phenomenon known as genetic hitchhiking. Since conditions vary from one locality to another and differ considerably between ecosystems (Oleksyk et al., 2010), it is expected that genomic differentiation occurred between populations. Patterns of contrasting LD between populations can assist in identifying adaptative genetic diversity for emerging global problems such as drought tolerance, UV radiation, heavy metal related genes, and ultimately, food security. Since climate change is likely to affect a wide range of species and habitats, LD could assist in the development of specific "adapted germplasm collections" suitable for cultivar development and conservation rather than collections capturing mostly neutral variation. Studies on the adaptation of natural populations to local ecosystems based on LD variations have already been reported (Li & Merilä, 2011).

Although GWAS have been successful in finding new causative alleles, usually tests for common variants are underpowered for detecting variants of lower frequency leaving a high proportion of undetected heritability. In human genetics, there is a growing interest in the role of rare variants in multifactorial disease etiology and there is an increasing body of evidence pointing to the role of rare variants in complex traits (Bansal et al., 2010). The frequency of any single rare or low-frequency variant is less than 5%, but collectively the number of rare variants is substantial. According to the multiple rare variant (MRV) hypothesis, there are many large effect rare variants in the population and cases of common inherited diseases have been the result of additive effects of a few of these moderate to high penetrance MRVs (Bodmer & Bonillna, 2008, as cited in Asimit & Zeggini, 2010). In the search for causal variants of type 1 diabetes (T1D), Nejentsev et al. (2009) identified four disease-associated rare variants in the *IFIH1* gene, which are protective of T1D. Involvement of rare variants in hypertension has also been shown (Ji et al., 2008). Despite their importance, rare variants have not been studied as extensively as common variants because of cost limitations in next generation sequencing technologies and the lack of an appropriate analytical toolbox to enable powerful rare variant association analysis (Asimit & Zeggini, 2010). With this in mind, several strategies for association studies involving rare variants have been proposed. The simplest approach is to test them individually using standard contingency table and regression methods such as those implemented in the genetic software PLINK (Purcell et al., 2007). This method, called "single-locus test" is highly problematic, given, for example, the poor power that such statistical tests have to detect small differences between diagnostic or phenotypic groups (Gorlov et al., 2008, as cited in Bansal et al., 2010). Other methods that overcome the power issues associated with testing rare variants individually include the collapsing strategy, methods based on summary statistics, multiple regression and data mining which are comprehensively reviewed by Bansal et al. (2010). Approaches involving direct sequencing have been tested by Li & Leal (2009). Since epigenetic factors are also likely to contribute to common complex traits, epigenome-wide association studies (EWASs) have been proposed to uncover another missing piece of heritability unexplained by common variants (Rakyan et al., 2011), specifically involving the study of variation in DNA methylation across the genome.

Future scenarios in plant AM will likely include a combination of studies involving common and rare variants to explain most of the phenotypic variation observed for agronomic and adaptative traits. The reader will have noticed the influence of human genetics in much of the discussion of LD mapping. Plant geneticists continue to follow human genetics research in order to improve QTL studies. However, plants offer advantages that cannot be afforded in humans such as population design and size, which promise to make plant GWAS a powerful tool. Overall, we anticipate witnessing advances in plant AM as a result of new approaches in human association studies in combination with the benefits of plant genetics that enable us to uncover and understand levels of plants genome complexity not seen before.

11. Acknowledgments

We are grateful to Andrzej Walichnowski for help with manuscript editing and Mike Shillinglaw for figures.

12. References

Abdurakhmonov, I. & Abdukarimov, A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics*, Vol. 2008, 1-18

Abdurakhmonov, I., Kohel, R., Yu, J., Pepper, A., Abdullaev, A., Kushanov, F., Salakhutdinov, I., Buriev, Z., Saha, S., Scheffler, B., Jenkins, J. & Abdukarimov. A. (2008). Molecular diversity and association mapping of fiber quality traits in exotic *G. Hirsutum* L. germplasm. *Genomics*, Vol. 92, No. 6, 478-487

Abdurakhmonov, I., Saha, S., Jenkins, J., Buriev, Z., Shermatov, S., Scheffler, B., Pepper, A., Yu, J., Kohel, R. & Abdukarimov A. (2009). Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica*, Vol. 136, No. 3, 401-417

Achleitner, A., Tinker, N., Zechner, E. & Buerstmayr, H. (2008). Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with quantitative traits. *Theoretical and Applied Genetics*, Vol. 117, No. 7, 1041-1053

Aranzana, M., Kim, S., Zhao, K., Bakker, E., Horton, M., Jacob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P. & Nordborg M. (2005). Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, Vol. 1, No. 5, e60

Ardlie, K., Kruglyak, L. & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Review Genetics*, Vol. 3, No. 4, 299-309

Arnheim, N., Calabrese, P. & Nordborg, M. (2003). Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. *The American Journal of Human Genetics*, Vol. 73, No. 1, 5-16

Asimit, J. & Zeggini, E. (2010). Rare variant association analysis methods for complex traits. *Annual Review of Genetics,* Vol. 44, 293-308

Balding, D. (2006). A tutorial on statistical methods for population association studies. *Nature Review Genetics*, Vol. 7, No. 10, 781-791

Bansal, V., Libiger, O., Torkamani, A. & Schork, N. (2010). Statistical analysis strategies for association studies involving rare variants. *Nature Review Genetics*, Vol. 11, No. 11, 773-785

Bar-Hen, A., Charcosset, A., Bourgoin, M. & Guiard, J. (1995). Relationship between genetic markers and morphological traits in a maize inbred line collection. *Euphytica*, Vol. 84, No. 2, 145-154

Beer, S., Siripoonwiwat, W., O'donoughue, L., Souza, E., Mathews, D. & Sorrels, M. (1997). Associations between molecular markers and quantitative traits in oat germplasm pool: can we infer linkages? *Journal of Agricultural Genomics*, Vol. 3

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J. & Roux, F. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics*, Vol. 6, No. 5, e1000940

Bradbury, P., Zhang, Z., Kroon, D., Casstevens, T., Ramdoss, Y. & Buckler, E. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, Vol. 23, No. 19, 2633-2635

Breseghello, F. & Sorrells, M. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, Vol. 172, No. 2, 1165-1177

Cockram, J., White, J., Zuluaga, D., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P., Stein, N., Schulte, D., Steuernagel, B., Marshall, D., Thomas, W., Ramsay, L., Mackay, I., Balding, D., AGOUEB Consortium., Waugh, R. & O'Sullivan, D. (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 50, 21611-21616

Dawson, E., Abecasis, G., Bumpstead, S., Chen, Y., Hunt, S., Beare, D., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganske, R., Lõhmussaar, E., Zernant, J., Tõnisson, N., Remm, M., Mägi, R., Puurand, T., Vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D., Cardon, L. & Dunham, I. (2002). A first-generation linkage disequilibrium map of human chromosome 22. *Nature*, Vol. 418, No. 6897, 544-548

Du, F., Clutter, A. & Lohuis, M. (2007). Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences*, Vol. 3, No. 3, 166-178

Ersoz, E., Yu, J. & Buckler, E. (2007). Application of linkage disequilibrium and association mapping in crop plants, In: *Genomics-Assisted Crop Improvement*, R. Varshney, & R. Tuberosa, (Ed.), 97-119, Springer, ISBN 987-1-4020-6294-0, Dordrecht, the Netherlands

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, Vol. 14, No. 8, 2611-2620

Excoffier, L. & Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Review Genetics*, Vol. 7, No. 10, 745-758

Flint-Garcia, S., Thornsberry, J. & Bukler, E. (2003). Structure of linkage disequilibrium in plants. *Annual Review Plant Biology*, Vol. 54, 357-374

Garris, A., Tai, T., Coburn, J., Kresovich, S. & McCouch, S. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics*, Vol. 169, No. 3, 1631-1638

Gaut, B. & Long, A. (2003). The lowdown of linkage disequilibrium. *The Plant Cell*, Vol. 15, No. 7, 1502-1506

Gupta, P., Rustgi, S. & Kulwal, P. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology*, Vol. 57, No. 4, 461-485

Hardy, O. & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, Vol. 2, No. 4, 618-620

Hill, W. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, Vol. 38, No. 6, 226-231

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E., Qian, Q., Zhang, Q., Li, J. & Han, B. (2010). Genome-wide studies of 14 agronomic traits in rice landraces. *Nature Genetics*, Vol. 42, No. 11, 961-967

Hubisz, M., Falush, D., Stephens, M. & Pritchard, J. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, Vol. 9, No. 5, 1322-1332

Jennings, H. (1917). The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. *Genetics*, Vol. 2, No. 2, 97-154

Ji, W., Foo, J., O'Roak, B., Zhao, H., Larson, M., Simon, D., Newton-Cheh, C., State, M., Levy, D. & Lifton, R. (2008). Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nature Genetics*, Vol. 40, No. 5, 592-599

Jung, M., Ching, A., Bhattramakki, D., Dolan, M., Tingey, S., Morgante, M. & Rafalski, A. (2004). Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh*1 locus in elite maize germplasm. *Theroretical and Applied Genetics*, Vol. 109, No. 4, 681-689

Krill, A., Kirst, M., Kochian, L., Buckler, E. & Hoekenga, O. (2010). Association and linkage analysis of aluminum tolerance genes in maize. *PLoS One*, Vol. 5, No. 4, e9958

Kump, K., Bradbury, P., Wisser, R., Buckler, E., Belcher, A., Oropeza-Rosas, M., Zwonitzer, J., Kresovich, S., McMullen, M., Ware, D., Balint-Kurti, P. & Holland, J. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics*, Vol. 43, No. 2, 163-168

Lewontin, C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, Vol. 49, No. 1, 49-67

Li, B. & Leal, S. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *The American Journal of Human Genetics*, Vol. 83, No. 3, 311-321

Li, M. & Merilä, J. (2011). Population differences in levels of linkage disequilibrium in the wild. *Molecular Ecology*, Vol. 20, No. 14, 2916-2928

Li, X., Wei, Y., Moore, K., Michaud, R., Viands, D., Hansen, J., Acharya, A. & Brummer, E. (2011a). Association mapping of biomass yield and stem composition in a tetraploid alfalfa breeding population. *The Plant Genome*, Vol. 4, No. 1, 24-35

Li, Y., Haseneyer, G., Schön, C., Ankerst, D., Korzun, V., Wilde, P. & Bauer, E. (2011b). High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biology*, Vol. 11, No. 6, 1-14

Li, Y., Huang, Y., Bergelson, J., Nordborg, M. & Borevitz, J. (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 49, 21119-21204

Liu, A. & Burke, J. (2006). Patterns of nucleotide diversity in wild and cultivated sunflowers. *Genetics*, Vol. 173, No. 1, 321-330

Liu, K. & Muse, S. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics,* Vol. 21, No. 9, 2128-2129

Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., Farkhari, M., Ribaut, J., Cao, M., Rong, T. & Xu, Y. (2010). Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 45, 19585-19590

Luikart, G., England, P., Tallmon, D., Jordan, S. & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Review Genetics*, Vol. 4, No. 12, 981-994

Mackay, I. & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, Vol. 12, No. 2, 57-63

Malysheva-Otto, L., Ganal, M. & Röder, M. (2006). Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genetics*, Vol. 7, No. 6, 1-14

Myles, S., Peiffer, J., Brown, P., Ersoz, E., Zhang, Z., Costich, D. & Buckler, E. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell*, Vol. 21, No. 8, 2194-2202

Myles, S., Boyko, A., Owens, C., Brown, P., Grassi, F., Aradhya, M., Prins, B., Reynolds, A., Chia, J., Ware, D., Bustamante, C. & Buckler, E. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 9, 3530-3535

Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. (2009). Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*, Vol. 324, No. 5925, 387-389

Nemri, A., Atwell, S., Tarone, A., Huang, Y., Zhao, K., Studholme, D., Nordborg, M. & Jones, J. (2010). Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 22, 10302-10307

Nielsen, D. & Zaykin, D. (2001). Association mapping: where we've been, where we're going. *Expert Review of Molecular Diagnostics*, Vol. 1, No. 3, 334-342

Noonan, J., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J. & Rubin, E. (2006). Sequencing and analysis of Neanderthal genomic DNA. *Science*, Vol. 314, No. 5802, 1113-1118

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics*, Vol. 154, No. 2, 923-929

Nordborg, M., Borevitz, J., Bergelsom, J., Berry, C., Chory, J., Hagenbland, J., Kreitman, M., Maloof, J., Noyes, T., Oefner, P., Stahl, E. & Weigel, D. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, Vol. 30, No. 2, 190-193

Oleksyk, T., Smith, M. & O'Brien, S. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society of London Biological Sciences*, Vol. 365, No. 1537, 185-205

Oraguzie, N., Wilcox, P., Rikkerink, H. & de Silva, H. (2007). Linkage disequilibrium, In: *Association mapping in plants*, N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner & H.N. de Silve, (Ed.), 11-39, Springer, ISBN-13 9780387358444, New York, USA

Palaisa, K., Morgante, M., Tingey, S. & Rafalski, A. (2003). Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *The Plant Cell*, Vol. 15, No. 8, 1795-1806

Poland, J., Bradbury, P., Buckler, E. & Nelson, R. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 17, 6893-6898

Pritchard, J. & Rosenberg, N. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *The American Journal of Human Genetics*, Vol. 65, No. 1, 220-228

Pritchard, J., Stephens, M., Rosenberg, N. & Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, Vol. 67, No. 1, 170-181

Pritchard, J. (2001). Deconstructing maize population structure. *Nature Genetics*, Vol. 28, No. 3, 203-204

Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, Vol. 38, No. 8, 904-909

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., Sklar, P., de Bakker, P., Daly, M. & Sham, P. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, Vol. 81, No. 3, 559-575

Rakyan, V., Down, T., Balding, D. & Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nature Review Genetics*, Vol. 12, No. 8, 529-541

Raman, H., Stodart, B., Ryan, P., Delhaize, E., Emebiri, L., Raman, R., Coombes, N. & Milgate, A. (2010). Genome-wide association analysis of common wheat (*Triticum aestivum* L.) germplasm identifies multiple loci for aluminium resistance. *Genome*, Vol. 53, No. 11, 957-966

Reeves, P. & Richards, C. (2009). Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. *PLoS One*, Vol. 4, No. 1, e4269

Robbins, M., Sim, S., Yang, W., Deynze, A., van der Knaap, E., Joobeur, T. & Francis, D. (2011). Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. *Journal of Experimental Botany*, Vol. 62, No. 6, 1831-1845

Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, P., Roose, M., Svensson, J., Stein, N., Varshney, R., Marshall, D., Graner, A., Close, T. & Waugh R. (2006). Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 103, No. 49, 18656-18661

Rousset, M., Bonnin, I., Remoué, C., Falque, M., Rhoné, B., Veyrieras, J., Madur, D., Murigneux, A., Balfourier, F., Le Gouis, J., Santoni, S. & Goldringer I. (2011). Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, Vol. 123, No. 6, 907-926

Schmegner, C., Hoegel, J., Vogel, W. & Assum, G. (2005). Genetic variability in a genomic region with long-range linkage disequilibrium reveals traces of a bottleneck in the history of the European population. *Human Genetics*, Vol. 118, No. 2, 276-286

Schulze, T. & McMahon, F. (2002). Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. *American Journal of Medical Genetics*, Vol. 114, No. 1, 1-11

Slatkin, M. (2008). Linkage disequilibrium: Understanding the evolutionary past and mapping the medical future. *Nature Review Genetics*, Vol. 9, No. 6, 477-485

Smith, A., Thomas, D., Munro, H. & Abecasis, G. (2005). Sequence features in regions of weak and strong linkage disequilibrium. *Genome Research*, Vol. 15, No. 11, 1519-1534

Spielman, R., McGinnis, R. & Ewens, W. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *The American Journal of Human Genetics,* Vol. 52, No. 3, 506-516

Stapley, J., Birkhead, T., Burke, T. & Slate, J. (2010). Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Research*, Vol. 20, No. 4, 496-502

Stich, B., Piepho, H., Schulz, B. & Melchinger, A. (2008). Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics,* Vol. 117, No. 6, 947-954

Stich, B. & Melchinger, A. (2010). An introduction to association mapping in plants. *CAB Reviews: Perspectives in agriculture, Veterinary Science, Nutrition and Natural Resources*, Vol. 5, No. 39, 1-9

Tenaillon, M., Sawkins, M., Long, A., Gaut, R., Doebley, J. & Gaut, B. (2001). Patterns of DNA sequence polymosphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 98, No. 16, 9161-9166

Teo, Y., Fry, A., Bhattacharya, K., Small, K., Kwiatkowski, D. & Clark, T. (2009). Genome-wide comparisons of variation in linkage disequilibrium. *Genome Research*, Vol. 19, No. 10, 1849-1860

Thornsberry, J., Goodman, M., Doebley, J., Kresovich, S., Nielsen, D. & Buckler, E. (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*, Vol. 28, No. 3, 286-289

Tian, F., Bradbury, P., Brown, P., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T., McMullen, M., Holland, J. & Buckler, E. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, Vol. 43, No. 2, 159-162

Virk, P., Ford-Lloyd, B., Jackson, M., Pooni, H., Clemeno, T. & Newbury, H. (1996). Predicting quantitative variation within rice germplasm using molecular markers. *Heredity*, Vol. 76. No. 3, 296-304

Von Zitzewitz, J., Cuesta-Marcos, A., Condon, F., Castro, A., Chao, S., Corey, A., Filichkin, T., Fisk, S., Gutierrez, L., Haggard, K., Karsai, I., Muehlbauer, G., Smith, K., Veisz, O. & Hayes, P. (2011). The genetics of winterhardiness in barley: perspectives from genome-wide association mapping. *The Plant Genome*, Vol. 4, No. 1, 76-91

Wang, J., McClean, P., Lee, R., Goos, R. & Helms, T. (2008). Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. *Theoretical and Applied Genetics*, Vol. 116, No. 6, 777-787

Wang, M., Sukumaran, S., Barkley, N., Chen, Z., Chen, C., Guo, B., Pittman, R., Stalker, H., Holbrook, C., Pederson, G. & Yu, J. (2011). Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. *Theoretical and Applied Genetics*, Vol. 123, No. 8, 1307-1317

Weber, A., Zhao, Q., McMullen, M. & Doebley, J. (2009). Using association mapping in teosinte to investigate the function of maize selection-candidate genes. *PLoS One*, Vol. 4, No. 12, e8227

Wen, W., Mei, H., Feng, F., Yu, S., Huang, Z., Wu, J., Chen, L., Xu, X. & Luo, L. (2009). Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, Vol. 119, No. 3, 459-470

Wright, S. & Gaut, B. (2005). Molecular population genetics and the search for adaptative evolution in plants. *Molecular Biology and Evolution*, Vol. 22, No. 3, 506-519

Würschum, T., Maurer, H., Kraft, T., Janssen, G., Nilsson, C. & Reif, J. (2011). Genome-wide association mapping of agronomic traits in sugar beet. *Theoretical and Applied Genetics*, Vol. 123, No. 7, 1121-1131

Yan, J., Shan, T., Warburton, M., Buckler, E., McMullen, M. & Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *Plos One*, Vol. 4, No. 12, e8451

Yin, T., DiFazio, S., Gunter, L., Jawdy, S., Boerjan, W. & Tuskan, G. (2004). Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. *New Phytologist*, Vol. 164, No. 1, 95-105

Yu, J. & Buckler, E. (2006). Genetic association mapping and genome organization of maize. *Current Opinions in Biotechnology*, Vol. 17, No. 2, 155-160

Yu, J., Pressoir, G., Briggs, W., Vroh Bi, I., Yamasaki, M., Doebley, J., McMullen, M., Gaut, B., Nielsen, D., Holland, J., Kresovich, S. & Buckler, E. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, Vol. 38, No. 2, 203-208

Yu, L., Lorenz, A., Rutkoski, J., Singh, R., Bhavani, S., Huerta-Espino, J. & Sorrells, M. (2011). Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. *Theoretical and Applied Genetics*, Vol. 123, No. 8, 1257-1268

Zhang, D., Bai, G., Zhu, C., Yu, J. & Carver, B. (2010). Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. *The Plant Genome*, Vol. 3, No. 2, 117-127

Zhao, K., Aranzana, M. & Kim, S. (2007a). An Arabidopsis example of association mapping in structured samples. *PLoS Genetics*, Vol. 3, No. 1, e4

Zhao, J., Paulo, M., Jamar, D., Lou, P., van Eeuwijk, F., Bonnema, G., Vreugdenhil, D. & Koornneef, M. (2007b). Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. *Genome*, Vol. 50, No. 10, 963-973

Zhu, Q., Zheng, X., Luo, J., Gaut, B. & Ge, S. (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Molecular Biology and Evolution*, Vol. 24, No. 3, 875-888

Zhu, C., Gore, M., Buckler, E. & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, Vol. 1, No. 1, 5-20

Zou, J., Jiang, C., Cao, Z., Li, R., Long, Y., Chen, S. & Meng, J. (2010). Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. *Genome*, Vol. 53, No. 11, 908-916

4

Fruit Germplasm Characterization: Genomics Approaches for the Valorisation of Genetic Diversity

Innocenzo Muzzalupo[1], Enzo Perri[1] and Adriana Chiappetta[2]
[1]*Agricultural Research Council - Olive growing and oil industry research centre (CRA-OLI), Rende (CS)*
[2]*University of Calabria (UNICAL), Dep. of Ecology Arcavacata di Rende (CS) Italy*

1. Introduction

The introduction of modern varieties of staple crop in agriculture seem to lead to an overall decrease in genetic diversity, although within the released varieties themselves the data are inconsistent and no overall narrowing of the genetic base can be discerned. The situation regarding genetic erosion in landraces and crop wild relatives is equally complex. While many recent studies have confirmed that diversity in farmers' fields and protected areas has eroded, this is not universally the case.

Many reports expressed continuing concern over the extent of genetic vulnerability and the need for a greater deployment of diversity. However, better techniques and indicators are needed to monitor genetic diversity, to establish baselines and monitor trends.

In recent years, there is evidence of growing public awareness with regard to the importance of genetic diversity, both to meet increasing demands for greater dietary diversity, as well as to meet future production challenges. The increased environmental variability that is expected to result from climate change implies that in the future, farmers and plant breeders will need to be able to access an even wider range of plant genetic resources for food and agriculture than today.

The existing *ex situ* collections of fruit trees germplasm may valuably provide either a source of genes potentially useful as raw material in plant breeding, or plants directly valid for a sustainable production. With respect to the latter item, we refer to those local varieties that, having evolved for a very long period in a location, and having developed adaptative traits, well integrated with the environmental, agronomic, cultural and traditional features of the site and more or less recently have been replaced with new varieties. The requirements of modern agriculture, such as sustainability call for the cultivation of a wider range of diverse material that could better respond to the different aspects involved. Specifically, if it is necessary to obtain new varieties with a broader genetic base, capable of producing under diverse conditions and to respond to different stresses – *i.e.* pests, drought, low fertility of the soil etc. On the other hand, in some cases, the re-introduction of old local varieties and

the safeguard of traditional farming systems and landscapes can be very profitable from an economic and socio-economic point of views. In general, the lack of information about plant genetic resources conserved have the effect of limiting the use that can be made of large existing collections, restricting the value and the usefulness of a collection even within the owning institute and among other potential users. Hence, assessing the traits of the germplasm conserved in a collection is an essential prerequisite to a proper and wide utilization of the plant material conserved and it is the first step toward a further definition of the roles that the varieties can play in sustainable production, through the direct use or in breeding programmes.

In the past few years, several studies based on comparative high throughput sequencing of plant transcriptomes have, indeed, allowed the identification of new gene functions, contaminant sequences from other organisms, alterations of gene expression in response to genotype, tissue or physiological changes, as well as large scale discovery of SNPs (Single Nucleotide Polymorphisms) in a number of model and non model species, such us maize, grapevine and eucalyptus (Costa *et al.*, 2010).

Among the cultivated plants, olive (*Olea europaea* L.) is the sixth most important oil crop in the world, presently spreading from the Mediterranean region of origin to new production areas, due to the beneficial nutritional properties of olive oil and to its high economic value. The Mediterranean basin is the traditional area of olive cultivation and has 95% of the olive orchards of the world. From the Mediterranean basin, olive cultivation is presently expanding into areas of Australia, South and North America (Argentina, Chile, United States) and South Africa.

It belongs to the family of Oleaceae, order of Lamiales, which includes about 10 families for a total of about 11,000 species. Members of this order are important sources of fragrances, essential oils and phenolics claiming for numerous health benefits, or providing valuable commercial products, such as wood or ornamentals. Information on the genome sequence and transcript profiles are completely lacking. Olive is a diploid species, predominantly allogamous. In spite of its economical importance and metabolic peculiarities, very few data are available on gene sequences controlling the main metabolic pathways.

In spite of its economical importance and metabolic peculiarities, very few data are available on gene sequences controlling the main metabolic pathways in olive. With regard to oil, a range of biochemical methods to study the traceability of olive oil has to be presented. In fact, the analysis of minor and major components present in olive oil represents a valuable tool for authentication purposes.

Food authenticity has become a focal point for producers, consumers and policy markers. The DNA based technology is gaining a great attention in the field of food authenticity. This technology makes use of molecular markers such as RAPD, AFLP and SSR more efficient and constitutes promising approach for variety characterization and oil traceability in olives.

In this contest, the acquisition of additional information on biochemical markers in olive represents a fundamental and indispensable step to preserve the main olive varieties and also to safeguard the minor genotypes, in order to avoid a loss of genetic diversity and offer an important genetic basis for future breeding programs.

2. Management of germplasm collection

Plant genetic resources are essential to a sustainable agriculture and food security. FAO estimates humans have used some 10,000 species for food throughout history. However, only about 120 cultivated species provide around 90% of food requirements and 4 species (Maize, Wheat, Rice and Potatoes) provide about 60% of human dietary energy for the world's population. Of the myriad of varieties of these crops developed by farmers over millennia, which form an important part of agricultural biodiversity, more than 75% have been lost in the past 100 years.

Some fear that corporate financial interests might prevent safeguarding of livelihoods, promotion of food security, biodiversity-rich farming under control of local communities.

The best way of conserving fruit germplasm collection is their utilization. However, today these resources are not only underutilized but also under conserved. The Global Plan of Action therefore supports activities improving *in situ* and *ex situ* conservation of plant collection. Regarding *ex situ* conservation, millions of accessions are already stored in hundreds of germplasm collections around the world for both conservation and utilization purposes. Find short descriptions about these germplasm databases and links to their websites by searching either by database type or by free text search.

2.1 The *in situ* management of germplasm collection

Awareness of the importance and value of crop wild relatives and of the need to conserve them *in situ* has increased. A global strategy for crop wild relatives preservation and use has been drafted, protocols for the *in situ* conservation of crop wild relatives are now available. The number and coverage of protected areas are expanding the last years and this has indirectly led to a greater protection of crop wild relatives.

Important progress has been made in the development of tools and techniques to assess and monitor plant genetic resources for food and agriculture within agricultural production systems. Countries now report a greater considerate of the amount and distribution of genetic diversity in the field, as well as the value of local seed systems in maintaining such diversity. More consideration is now being paid in several countries to increasing genetic diversity within production systems as a way to reduce risk, particularly in light of changes in climate, pests and diseases. The number of on-farm management projects is increased somewhat and new legal mechanisms have been put in place in several countries to enable farmers to market genetically diverse varieties. There is still a need for more effective policies, regulations governing the *in situ* and on-farm management of plant genetic resources for food and agriculture, both inside and outside protected areas, and closer collaboration and coordination are needed between the agriculture and environment sectors. Many aspects of *in situ* management still require further research and strengthened research capacity is required in such areas as the taxonomy of crop wild relatives and the use of molecular tools to conduct inventories and surveys.

2.2 The *ex situ* management of germplasm collection

The total number of varieties conserved *ex situ* international has reached 7.4 million. While new collecting accounted for at least 240,000 varieties, and possibly considerably more,

much of the overall increase is the result of exchange. It is estimated that less than 30% of the total number of varieties are distinct (FAO, 2010). There is still a need for greater rationalization among collections globally.

The existing *ex situ* collections of fruit tree germplasm may valuably provide either a source of genes potentially useful as raw material in plant breeding, or plants directly valid for a sustainable production. With respect to the latter item, we refer to those local varieties that, having evolved for a very long period in a location, and having developed adaptative traits well integrated with the environmental, agronomic, cultural and traditional features of the site and more or less recently have been replaced with new varieties. The needs of modern agriculture, such as sustainability call for the cultivation of a wider range of diverse material that could better respond to the different aspects involved. Specifically, if it is necessary to obtain new varieties with a broader genetic base, capable of producing under diverse conditions and to respond to different stresses – *i.e.* drought, pests, low fertility of the soil etc. –, on the other hand, in some cases, the reintroduction of old local varieties and the safeguard of traditional farming systems and landscapes, can be very profitable from an economic and socio-economic point of views. In general, the lack of information about plant genetic resources conserved have the effect of limiting the use that can be made of large existing collections, restricting the value and the usefulness of a collection even within the owning institute and among other potential users. Hence, assessing the traits of the germplasm conserved in a collection is an essential prerequisite to a proper and wide utilization of the plant material conserved and it is the first step toward a further definition of the roles that the varieties can play in sustainable production, through the direct use or in breeding programmes.

Germplasm collections established and maintained by genebanks provide for the present and future utilization of plant genetic resources. In the early stages of collection development, the focus was mainly on acquisition *per se*, and less on optimizing collection composition. Many germoplasm collections were started from working collections that had been used to support specific purposes, including breeding, crop improvement and taxonomic studies. In many cases, germplasm collections expanded their collections thereafter by including obsolete varieties, research lines or samples obtained from collecting missions to natural distribution areas of crops and their wild relatives.

There is still a need for greater rationalization among collections globally. Scientific understanding of the on-farm management of genetic diversity has increased. While this approach to the conservation and use of plant genetic resources for food and agriculture is becoming increasingly mainstreamed within national programmes, further efforts are needed in this regard. With the development of new molecular techniques, the amount of data available on genetic diversity has increased dramatically, leading to an improved understanding of issues such as domestication, genetic erosion and genetic vulnerability.

The largest total numbers of *ex situ* varieties are of wheat, rice, barley and maize accounting for 77% of the total cereal and pseudo-cereal holdings. Other large cereal holdings include sorghum (about 235,000 varieties) and pearl millet (more than 65,000 varieties; FAO, 2010). In some tropical countries, roots and tubers, including cassava, potato, yam, sweet potato and aroids, are more important as staple foods than cereals, but being more difficult to conserve, collection sizes tend to be smaller. Centro Internacional de la Papa (CIP, Spain)

holds the world's largest sweet potato collection (more than 6,400 varieties) as well as the third largest potato collection (representing about 8% of total world holdings of about 98,000 varieties) after those of the Institut National de la Recherche Agronomique (INRA, France) and N.I. Vavilov All-Russian Scientific Research Institute of Plant Industry (Russian Federation).

Other important collections of olive tree (*Olea europaea* L. subsp. *europaea* var. *europaea*) are found at several Mediterranean countries at Aegean Agricultural Research Institute of Turkey (AARI, Turkey), Consiglio per la Ricerca e la Sperimentazione in Agricoltura - Centro di Ricerca per l'Olivicoltura e l'Industria Olearia (Agricultural Research Council - Olive growing and OiL Industry research centre, CRA-OLI, Italy), Horticulture and Subtropical Crops Research Institute (HSCRI, Azerbaijan), Junta de Andalucía, Instituto Andaluz de Investigación Agroalimentaria y Pesquera, Centro de Investigación y Formación Agroalimentaria Córdoba (CIFACOR, Spain), National Plant Gene Bank of Iran was placed in the Seed and Plant Improvement Institute (NPGBI-SPII, Iran). The largest olive collection (accounting for 17% of the total olive trees with more than 500 varieties) is held by CRA-OLI in Italy, followed by the collections of the CIFACOR in Spain.

The systematic collection of Italian olive varieties for deposit into specific catalogue fields began in Italy in the 1980s. A similar international collection was begun in 1997 by CRA-OLI of Rende, Italy. Collection entailed the following steps: a survey of the territory, individuation, basic characterization, and introduction into the gene bank field. Material identified by other international scientific institutions (International Treaty on Plant Genetic Resources for Food and Agriculture - Plant Genetic Resources RGV-FAO Projects) was also included. To date, roughly 500 varieties have been introduced into the CRA-OLI collection, and this list has been published (web site http://apps3.fao.org/wiews/olive/oliv.jsp).

The olive tree is one of the oldest cultivated plants, and its fruit has been used for nourishment for more than 5,000 years in the Mediterranean regions where it originated. Over the last few centuries, cultivation of the olive tree has spread to North and South America, as well as Japan, South Africa, and Australia. Due to the tree's need for a warm but not excessively hot climate, it can be cultivated in both the northern and southern hemispheres between 30 and 45 degrees latitude, with the exception of some equatorial regions where olive trees are grown at high altitude. Nowadays, olives are produced in more than 40 countries spread across all six inhabited continents, and even in exotic places like Hawaii.

A useful olive germplasm collection also requires an organizational system devoid of homonymy, synonymy and mislabelling so that a reliable classification of all varieties can be achieved without unnecessary confusion. Recent research has focused on using morphology and biochemical and molecular markers to characterize and identify olive varieties. The identification of varieties and varieties using molecular markers is a crucial aim of modern horticulture, because such a technique would greatly facilitate breeding programmes and germplasm collection management.

3. Olive germplasm characterization

The genetic patrimony of the Mediterranean Basin's olive trees are very rich and is characterised by and abundance of varieties. Based on estimates by the FAO Plant

Production and Protection Division Olive Germplasm (FAO, 2010), the world's olive germplasm contains more than 2,629 different varieties, with many local varieties and ecotypes contributing to this richness.

The olive tree is a member of the Oleaceae family, which contains the genera *Fraxinus*, *Forsythia*, *Forestiera*, *Ligustrum*, and *Syringa*, in addition to the genus *Olea*. The genus *Olea* of the sub-family *Oleideae*, includes two sub-genera, *Olea* and *Paniculatae*. According to recent revisions of the *Olea europaea* taxonomy (Green, 2002), this species is divided into the following six sub-species based on morphology and geographical distribution:

1. subsp. *europaea*, divided into the two botanical varieties: the wild olive or oleaster (var. *sylvestris*) and the cultivated olive (var. *europaea*), distributed in the Mediterranean Basin;
2. subsp. *cerasiformis*, present in Madeira Island;
3. subsp. *cuspidata*, distributed from South Africa to southern Egypt and from Arabia to northern India and south-west China;
4. subsp. *guanchica*, present in the Canary Islands;
5. subsp. *laperrinei*, localized to the Sahara region;
6. subsp. *maroccana*, present in south-western Morocco.

Commercial olives are products of *Olea europaea* subsp. *europaea* var. *europaea*, as only this species produces edible fruit. The cultivated olive tree can reach heights ranging from just a few meters to 20 meters. The trunk is irregular, and the branches bear evergreen, elliptical and/or lanceolate leaves whose upper and lower surfaces are green and silvery, respectively. The olive tree (photo 1) is a long lived evergreen and some specimens have been reported to live for nearly 2,000 years. Its wood can resist decay, and when mechanical damage or environmental extremes kill the top of the tree, new growth arises from the root system.

Olive trees were multiplied by using different explants including ovule (spheroblast) and subsequently leafy stem cutting and grafting on seedlings or clonal stocks. Vegetative reproduction potential varies, which is dependent on genotype, e.g. easy to rooting and recalcitrant to root initiation (Hartmann and Kester 1968). Micropropagation of the olive variety was successful on OM medium (Rugini, 1984) and subsequently several other researchers slightly modified the culture medium by adding different growth substances or rooting conditions (Cozza *et al.*, 1997; Mencuccini, 2003). The micropropagated materials (photo 2) can be used to screen for resistance to biotic and abiotic stress and for genetic improvement activity (Rugini *et al.*, 2000; Sasanelli *et al.*, 2000; Bartolozzi *et al.*, 2001).

When propagated by either seed or cuttings, the root system generally is shallow, spreading to only 0.9-1.2 meters even in deep soils. The above ground portion of the olive tree is recognizable by its dense assembly of limbs, short internodes, and the compact nature of the foliage. Light does not readily penetrate into the interior of an olive tree unless the tree is pruned to create light channels. If left unkempt, olive trees develop multiple branches with cascading limbs. The branches are able to bear large quantities of fruit on their terminal twigs, which are pendulous, flexible, and sway with the slightest breeze.

Olive leaves (photo 3) are thick, leathery, and oppositely arranged. The silvery green leaves are oblong in shape, measuring 4–10 centimetres long and 1–3 centimetres wide. Leaves

have stomata on their lower surfaces only. Stomata are nestled in peltate trichomes that restrict water loss and make the olive tree relatively resistant to drought. Some multicellular hairs are present on the leaf surfaces. Each leaf grows over a two year period. Olive leaves usually abscise in the spring after they are 2 or 3 years old. As with other evergreens, however, leaves older than 3 years are often present. Flower bud inflorescences (photo 4) are borne on each leaf's axil. The small white, feathery flowers, with ten cleft caly x and corolla, two stamens and bifid stigma. The bud is usually formed during one season, at which point it can remain dormant for more than a year before beginning visible growth during the subsequent season. After the buds become viable inflorescences, flowers bloom a season later than expected. Each inflorescence contains between 15 and 30 flowers, depending on the variety and on the extent of that year's development.

Photo 1. Calabrian secular olive (*Olea europaea* subsp. *europaea* var. *europaea*) trees

The olive fruit is a drupe (photo 5), botanically similar to the almond, apricot, cherry, nectarine, peach, and plum. The olive fruit consists of an exocarp, a mesocarp and an endocarp. The exocarp represents the 1.5-3.5% of the total fruit; it is free of hairs and contains stomata. The mesocarp represents the 70-80% of the total fruit; it is the tissue that is eaten, and the endocarp is woody and represents the 13-24% of the total fruit and encloses the seed (2-4% of the total fruit).

Quantitatively, the largest constituents of the drupe are water (40-70%) and oil (6-25%). The biochemical composition of olive oil consists of a major portion that includes triacylglycerols and that represents more than 98% of the total oil weight and a minor ones, that is present in very low amount (about 2% of oil weight), including more than 230 chemical compounds such as aliphatic and triterpenic alcohols, sterols, hydrocarbons, volatile compounds and antioxidants (tocopherols and phenolic compounds).

Photo 2. *In vitro* propagation of olive (*Olea europaea* subsp. *europaea* var. *europaea*) trees through micro-grafting

Photo 3. Olive (*Olea europaea* subsp. *europaea* var. *europaea*) leaves: top side and under side

Photo 4. Olive (*Olea europaea* subsp. *europaea* var. *europaea*) inflorescence (raceme)

The phenolic compounds have shown their relevance in the production of virgin olive oil, typical food of the Mediterranean culture because of their bioactive contribution to sensory characteristics, to stability toward autoxidation, and to human health beneficial effects (Muzzalupo *et al.*, 2011, Servili *et al.*, 2004). Olive fruit pulp naturally possesses a bitter taste due to the presence of the glycoside oleuropein (photo 6), (Bianco *et al.*, 1999, 2001; De Nino *et al.*, 2005).

Photo 5. The olive fruit (drupe) at different ripening stages

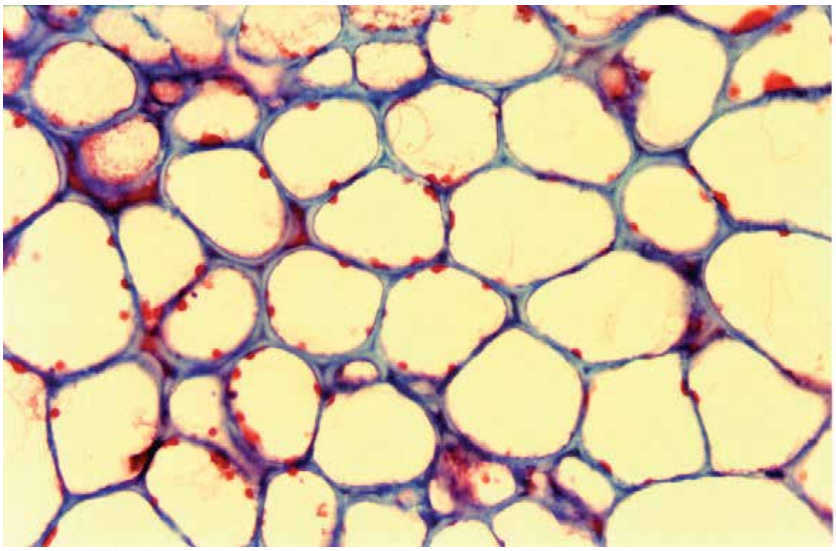


Photo 6. Cytological features of olive fruits. Fruit longitudinal sections stained with safranin O/azur II;

The olive tree and its products can be damaged from many diseases and pests. The most dangerous are the bacterium *Pseudomonas savastanoi* (photo 7), which produce tubercules forms on the branches and stems, the fungus *Cycloconium oleaginum* that damage the leaves and fruits and *Verticillum dahlie* that is destructive for the root apparatus and the growth of the plants. Among phytophagous, most harmful are the olive fruit fly (*Bactrocera olea* Gmelin), the olive moth (*Prays oleae* Bernard) and black scale *(Saissetia oleae* Olivier). Olive fruit fly is the major pest and can cause severe economic damage to olive production, which effect oil extraction and table use (photo 8).

Photo 7. *Pseudomonas savastanoi* olive or tuberculosis

Photo 8. Olive fruit fly (*Bactrocera olea* Gmelin)

Varieties are predominantly diploid ($2n = 2x = 46$) (Minelli *et al.,* 2000). The DNA content is 2.2 pg per 1C nucleus (Bitonti *et al*., 1999), correspondent to a genome size of 2.2 Gbp (De la Rosa *et al*., 2003).

Over the millennia, new varieties have originated by genetic mutation, by spontaneous crossing with a subsequent natural dissemination of stones. Also an important factor in the development of locally specific varietal populations was sexual reproduction, involving populations of local wild *Olea* and those selected to the criteria of local farmers (Breton *et al*., 2006). If agreeable by humans, that new varieties were established by vegetative means. The longevity of the olive tree and the selection of a large number of varieties have contributed to the conservation of its variability and allowed to pass a large proportion of this genetic diversity (Rallo *et al*., 2000). Another factor that has contributed to increasing the biodiversity of this species is the wide genetic variability of olive that has been created and distributed freely without any concern for loyalty to a morphologically defined archetype because the end product is not the whole fruit, such as for most other fruit trees, but the result of squeezing the fruit: the virgin olive oil. This has led, over time, to the formation of polyclonal varieties of heterogeneous phenotype (varieties–populations) rather than the formation of monoclonal varieties. Intra-varietal polymorphisms in fact, have been reported in the literature (Lopes *et al*., 2004; Muzzalupo *et al*., 2009b, 2010) in which the observed differences within the same variety have been suggested as somatic mutations occurring during vegetative propagation.

The problem of characterizing the olive tree germplasm is complicated not only by the richness of its genetic patrimony, but also by the absence of reference standards and a well-defined system of nomenclature that is free from homonymy and synonymy (Bartolini and Petruccelli, 2002). For olive varieties there are still no "standard reference variety" (Roselli and Scaramuzzi, 1974) and only recently, some research Italian projects (*i.e.,* "International Treaty on Plant Genetic Resources for Food and Agriculture - Plant Genetic Resources RGV-FAO", "Improvement and qualification of nursery olive" OLVIVA and "Research and Innovation for the South Olive" RIOM projects) have been raising this issue and are trying

to achieve a "standard certificate" for each variety present in different Italian regions. The extent of this diversity has important implications for both the adaptation of varieties to their local environment and for the optimization of these varieties agronomical performance under a given set of environmental conditions. For example, every initiative promoting olive cultivation should consider the potential repercussions of such action on any local olive varieties. Every region should preserve its own plant material in order to safeguard both the adaptation and productivity of the species and the unique characteristics of the region's olive oil. However, the study of intra-varietal polymorphisms is important since they may have traits that although not considered important in the past, might be important to meet the challenges of modern olive growing (*i.e.*, resistance to low temperatures, salinity tolerance, etc.).

The preliminary work performed in olive tree genomics is currently very far from producing results that are useful for selecting new varieties using molecular tools. This combined with the general lack of prior knowledge regarding the cultivated and wild olive germplasms, has focused attention mainly on the evaluation of the germplasm.

There is a strong need for a means of reliably identifying different olive tree varieties, partly because so many of these varieties are propagated solely via vegetative methods. This would also be of substantial benefit to nurserymen and growers, because the cost of plants represents the major investment in establishing new orchards. At the same time, it is also important to improve the *ex situ* plant germplasm collection in order to characterize adequately all varieties, and to develop future breeding programs.

Morphological and biological characteristics are widely used for descriptive purposes and are commonly used to distinguish olive varieties (Barranco *et al.,* 2000; Cantini *et al.*, 1999; Lombardo *et al.,* 2004). Agronomic characterization has also aided in the classification of different olive varieties (Barranco and Rallo 2000; Lombardo *et al.,* 2004). Morphological characterization of olive varieties is potentially unreliable, because environmental factors strongly influence the plants' morphology. Despite this drawback, the age of trees, their training systems, and the phenological stage of the plants continues to be a key preliminary step in the description and classification of the olive tree germplasm (Lombardo *et al.,* 2004) At the same time, improving *ex-situ* olive plant germplasm collections remains an important objective, which will ultimately prove useful for characterizing all varieties and for developing future breeding programs.

Recently, a multiplicity of molecular markers as been used to characterize and distinguish between olive varieties. In light of these efforts, some combination of enzymatic markers with distinct morphological, physiological, and agronomic characteristics may ultimately provide a method for the reliable and systematic classification of olive tree varieties (Ouazzani *et al.,* 1995). Assessments of microsatellite markers, RAPD profiles, AFLPs, and RFLPs provide direct genotypic information, which has numerous, valuable applications in genetic studies. The main advantages of generating RAPD profiles are the technique's simplicity and low cost (Bogani *et al.,* 1994; Fabbri *et al.,* 1995; Wiesman *et al.,* 1998; Belaj *et al.,* 2001; Muzzalupo *et al.,* 2007a). Nevertheless, RAPD experiments demonstrate poor reproducibility, which hampers comparison between individual studies. Experiments assessing an organism's AFLP markers are more technically demanding than RAPD but are highly effective in detecting DNA polymorphisms (Angiolillo *et al.,* 1999; Baldoni *et al.,* 2000; Muzzalupo *et al.*, 2007a; Owen *et al.,* 2005). In contrast to a plant species' chloroplast DNA

(cpDNA), which occasionally can be insufficiently variable for intra-species comparison (Wolfe *et al.,* 1987; Amane *et al.,* 1999; Lumaret *et al.,* 2000; Besnard *et al.,* 2002), mitochondrial DNA (mtDNA) within a given species varies enormously in terms of organization, size, structure, and gene arrangement (Brennicke *et al.,* 1996). As a result, intra-species mtDNA variation is common in plants, especially in naturally occurring populations (Besnard *et al.*, 2002). Taken together, these distinctive features make mtDNA sequencing a powerful tool for analysing a given plant population's genetic structure and phylogenetic relationships (Cavallotti *et al.,* 2003). Microsatellite markers are ubiquitous, abundant, and highly dispersed in eukaryotic genomes, but are costly to assess experimentally. Once these markers have been ascertained, the data can be readily shared among laboratories. Since not all microsatellites are identical (Baldoni *et al.*, 2009; Rallo *et al.,* 2000; Sefc *et al.,* 2000; Carriero *et al.,* 2002; Cipriani *et al.,* 2002; Muzzalupo *et al.,* 2006, 2009a), however, successful utilization of known microsatellite markers requires prior information regarding the characteristics of a particular genetic locus (Baldoni *et al.*, 2009).

Internal transcribed spacer 1 (ITS-1) sequences, RAPD profiles, and inter-SSR (ISSR) markers have been employed to evaluate the colonization history of *Olea europaea* (Hess *et al.,* 2000). A number of *Olea europaea* retroelements have also been identified (Hernandez *et al.,* 2001), and their copy number has been estimated (Stergiou *et al.,* 2002). Using previously established RAPD profiles (Hernandez *et al.,* 2001; Mekuria *et al.,* 2001) developed SCAR markers linked to leaf peacock spot tolerance. Another method to distinguish inter-variety variability and to characterize clonal variants using single nucleotide polymorphisms (SNPs) in the olive tree genome is also currently under development (Rekik *et al.*, 2011; Reale *et al.,* 2006).

All the aforementioned genetic techniques provide useful information regarding the level of olive tree polymorphism and diversity, which demonstrates their utility for the characterization of germplasm varieties (Belaj *et al.,* 2003).

3.1 Molecular approaches for olive oil quality control

The food crisis situation seen in last years and the controversy about genetically modified organisms (GMO), with a sharp increase in basic food prices, highlights the extreme susceptibility of the current agricultural and food model and the need for more strict food quality control, which should include determination of the origin of the product and the raw materials used in it. That's why a well documented traceability system has become a requirement for quality control in the food chain. The definition of traceability according to the European Council Regulation EEC 178/2002 is the ability to identify and trace a product or a batch of products at all stages of production and marketing. Traceability is important for commercial reasons and plays a considerable role in the assurance of public health.

Olive oil extraction is the process of extracting the oil present in the olive drupes for food use. The oil is produced in the mesocarp cells, and stored in a particular type of vacuole called a lipovacuole (photo 9). Olive oil extraction is the process of separating the oil from the other fruit contents. It is possible to attain this separation by physical means alone, *i.e.* oil and water do not mix, so they are relatively easy to separate.

The modern method of olive oil extraction uses an industrial decanter to separate all the phases by centrifugation. In this method the olives are crushed to a fine paste. This can be

done by a hammer crusher, disc crusher, depicting machine or knife crusher. This paste is then malaxed for 15 to 45 minutes in order to allow the small olive droplets to agglomerate. The aromas are created in these two steps through the action of fruit enzymes. Afterwards the paste is pumped in to an industrial decanter where the phases will be separated.

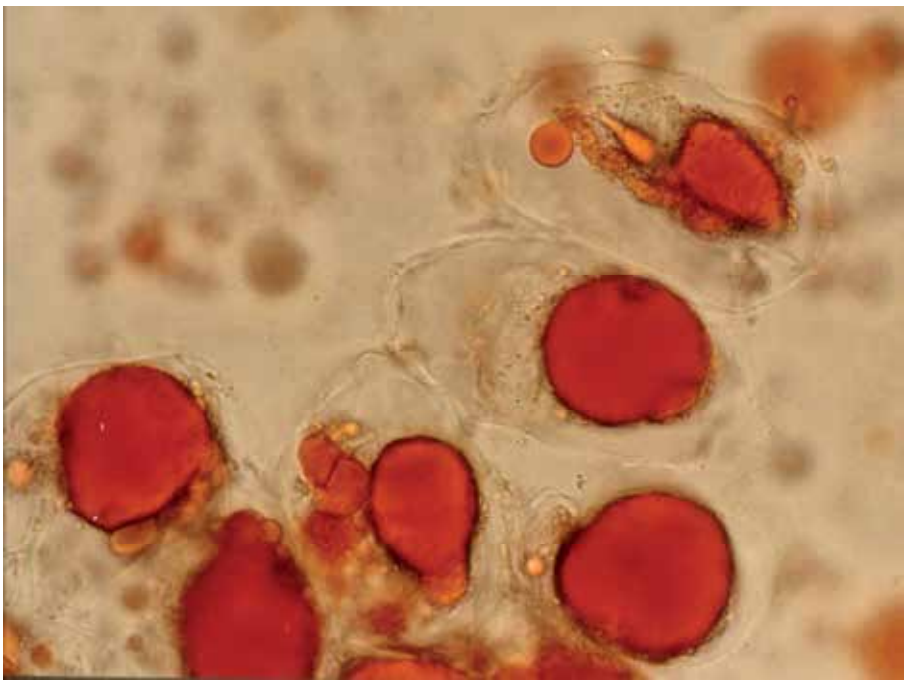


Photo 9. Lipovacuole from olive mesocarp cells stained with sudan IV

The olive oil chemical components are divided, into major and minor compounds that are briefly described below. *Major components*: glycerids correspond to more than 98% of the total weight. Abundance of oleic acid (C18:1 *n*−9), is a monounsaturated fatty acid and present in concentrations between 56 to 84% of total fatty acids, while the most essential polyunsaturated fatty acid in our diet is the linoleic acid (C18:2 *n*−6), ranges from 3 to 21% (Caravita *et al.*, 2007). *Minor components*: amounting to about 2% of the total oil weight, include compounds that are not related to lipids from a chemical viewpoint (tocopherols, polyphenols, chlorophylls, etc.) and compounds from unsaponifiable matter derived from lipids (sterols, phospholipids, waxes, ect.) (Servili *et al.*, 2004).

Almost 84% from the total olive oil production derives from the European Union, especially from Spain, Italy and Greece. The olive oil is a main constituent of the Mediterranean diet. However there has recently been an increase in olive oil consumption internationally, due to greater availability and the current consideration of ist high nutritive and health benefits, including a qualified health claim from Food and Drug Administration (FDA, USA).

Some varieties of olive oil are recognized as being of higher quality because they derive from well-defined geographical areas, command better prices and generally are legally protected. Indeed, the aim of Protected Designations of Origin (PDO), Protected Geographical Indication (PGI) and Traditional Specialty Guaranteed (TSG) is to add value to

certain specific high quality products from a particular origin. Chemical techniques have been employed for the authenticity of olive oils using a high number of variables such as glycerid composition, phenolic fraction, unsaponifiable components monitoring by statistical and mathematical analyses in order to ability the evaluation of the results. Molecular markers allow the detection of DNA polymorphisms and enable to effectively distinguish different varieties in an effective way, without any environmental influence.

When we blend olive oils of the same category, but from different provenances, most chemical analyses are of limited significance. Due to their high variability according to environmental conditions, neither morphological characteristics of different groups, nor the analyses of chemical composition of fatty acid and secondary metabolites can provide reliable results for oil traceability (Ben Ayed *et al*., 2010; De Nino *et al.,* 2005; Papadia *et al*., 2011). For this reason, genetic identity seems to be the most appropriate method for identifying the variety from which the olive oil under study derives. In fact, DNA in oil is not affected by the environment and is identical to the mother tree DNA since the oil containing tissues are formed by diploid somatic cells of the tree (Muzzalupo *et al*., 2007b). However, depending on the molecular markers used correctly, extra alleles can be detected in the oil that do not correspond to the mother tree allele but to the pollinator alleles contained in the embryo, itself located inside the seed (Muzzalupo and Perri, 2002; Ben Ayed *et al*., 2010). The use of DNA based technology in the field of food authenticity is gaining increasing attention. This technique makes use of molecular markers that mostly use polymerase chain reaction (PCR) and are thus easy to genotype. Even in a complex matrix such as olive oil, molecular marker techniques such as RAPDs (random amplified polymorphic DNA), AFLPs (amplified fragment length polymorphism) and SSRs (simple sequence repeat) are very useful in the study of the traceability of olive oil. SNP markers have been recently developed in olive and utilized to study the genetic diversity of olive trees (Reale *et al*., 2006; Rekik *et al.*, 2010).

A recent report by Papadia *et al.*, 2011 reported a systematic effort to obtain genetic characterization by SSR amplification, soil analyses, and ^{1}H-NMR spectra, is carried out in order to make a direct connection between the olive tree variety (genetic information) and the NMR spectra (chemical information) of the extra virgin olive oil produced. The results reported show that a multidisciplinary approach, through the application of multivariate statistical analysis, could be used to set up a method for variety and/or geographic origin certification, based on the construction of a suitable database. Further research will be directed to the growth of an organic genetic/NMR/soil database, in order to improve the prediction ability of the LDA, and furthermore to develop a way to correlate ^{1}H-NMR spectra of commercial extra virgin olive oils with their geographical and genetic origin.

In the following subsections we will discuss the potential of these classes of markers in the oil traceability and in characterization of olive germplasms.

3.1.1 RAPDs (Random Amplified Polymorphic DNA)

In this technique, a PCR amplification of genomic DNA is performed using a set of arbitrary primers (Williams *et al*., 1990). For each primer a large number of bands are generated and each DNA has the presence/absence of a band can distinguish between individuals and each individual is expected to have a specific fingerprint of bands. This molecular technique

has several advantages. It is simple, cheap, it requires small amounts of DNA (Fritsch and Rieseberg, 1996), and it can be applied without prior genetic information about the organism. Besides, it is fast, and does not require radioactivity. However, this analysis has several limitations including dominance, sensitivity to the reaction conditions, uncertain locus homology and the lack of good reproducibility. RAPDs thus combine the advantages of low technical input with almost an unlimited numbers of markers. They have proven to be very useful in the characterization of genetic diversity of plants for which few genomic data are available (Qian *et al.*, 2001; Bandelj *et al.*, 2002). RAPD markers were the first ones to be implemented to study diversity of the species *Olea europaea* (Belaj *et al.*, 2001), to discriminate olive varieties (Khadari *et al.*, 2003; Muzzalupo *et al.*, 2007a), to study inter or intra-variety genetic diversity (Wiesman *et al.*, 1998; Mekuria *et al.*, 2001, Muzzalupo and Perri, 2009; Belaj *et al.*, 2002, 2003; Gemas *et al.*, 2004), to establish genetic relationships between varieties (Belaj *et al.*, 2002, 2003; Besnard *et al.*, 2002; Khadari *et al.*, 2003; Muzzalupo *et al.*, 2007a), and to study genetic differentiation in the olive complex (Besnard *et al.*, 2001; Martins-Lopes *et al.*, 2008). As early as their use in genetic studies RAPD markers has been used for the authentication and traceability of olive oil (Pasqualone *et al.*, 2001; Muzzalupo and Perri, 2002). However, numerous authors (Pasqualone *et al.*, 2001; Sanz-Cortés *et al.*, 2001) concluded the non-reproducibility of RAPD markers in the authentication of olive oil, which resulted in inconsistent electrophoretic patterns. These unsuccessful attempts are due to the bad quality of DNA extracted from oil (Pasqualone *et al.*, 2001; Muzzalupo and Perri, 2002).

3.1.2 AFLPs (Amplified Fragment Length Polymorphism)

AFLP was described by Vos *et al.,* (1995) as a more reproducible alternative to RAPD for the genetic identification of crop plants. This technique is based on the selective PCR amplification of restriction fragments from total digests of genomic DNA. In olive, AFLP markers have been used for genetic diversity studies and variety identification. In fact, amplified fragment length polymorphism technology has been used by Angiolillo *et al.,* (1999) to obtain a large number of markers for olive. This has been used in addressing genetic relationships among wild and cultivated varieties, as well as among *Olea europaea* L. and other species from the genus (within the *Olea* complex). This technique has also been used to study the genetic diversity within and among a range of Spanish and Italian olive varieties (Sanz-Corte´s *et al.*, 2003). Owen *et al.,* (2005) used AFLP markers to evaluate the structure of genetic diversity among common olive varieties cultivated in the Eastern Mediterranean. Additionally, AFLP analysis, as previously described and has been used in genetic variability studies for about 29 varieties (including oil and table olive varieties originating from Tunisia and other Mediterranean countries) of the genus *Olea* using nine AFLP primer combinations (Grati-Kamoun *et al.*, 2006). Different studies (Busconi *et al.*, 2003, Pafundo *et al.*, 2005) have reported that it is possible to use AFLP markers for genotyping olive species. As far as oil traceability is concerned, Busconi *et al.,* (2003) reported that the AFLP fingerprint of olive oil was only partially super imposable with that of the variety from which the oil was made. However, in more recent studies, Pafundo *et al.,* (2005) and Montemurro *et al.,* (2007) concluded that AFLP profiles of DNA purified from leaves and the monovarietal oil of the same variety were comparable. These latter evaluated the possibility of identifying virgin olive oil from ten different varieties by the analysis of AFLP markers using six AFLP primer combinations. For the AFLP as well as for RAPDs, the

quality of DNA isolated from olive oil seems again to be the problem (very low quantity, a high degradation and the richness in polysaccharides and phenolic compounds). Poor quality of DNA is responsible for inconsistent results and low reliability of AFLP profiles due to the inhibition of the restriction enzymes and the DNA polymerase activity.

3.1.3 SSRs (Simple Sequence Repeats)

SSRs are a class of DNA markers that consist of short tandem repeat sequences (2–6 bp), which have become one of the most successful and the most interesting markers for genotype identification due to their good properties; in addition to their high specificity, they are highly polymorphic, codominant, *locus* specific, ubiquitous, widely distributed throughout the genome and easily amenable to automated PCR-based analysis. At present, they are the most reliable DNA profiling methods in forensic investigation (Jobling and Gill, 2004). SSRs also are highly informative and reproducible tools because they use longer primer sequences (Vos, 1995).

In olive SSRs have shown high potential for resolving issues of synonymies, homonymies and misnamings. Many SSRs have been developed in olive and applied with success (Sefc *et al.*, 2000; Carriero *et al.*, 2002; Cipriani *et al.*, 2002; De la Rosa *et al.*, 2003; Sabino Gil *et al.*, 2006). All these characteristics make them ideal markers for applications in analysis of intravariety variability issues (Cipriani *et al.*, 2002; Lopes *et al.*, 2004; Muzzalupo *et al.*, 2009b, 2010), linkage mapping (Wu *et al.*, 2004) and for characterizing olive germplasm resources (Belaj *et al.*, 2004; Montemurro *et al.*, 2007; Muzzalupo and Perri, 2009). Sarri *et al.,* (2006) confirmed the power of SSR markers in the identification of 118 varieties from different Mediterranean countries to study the genetic diversities of olive varieties. A recent report by Muzzalupo *et al.,* (2009a) characterized 211 Italian olive varieties by using 11 loci microsatellite in order to study and to establish relationships of geographically-related olive-tree varieties. Microsatellites are also very useful markers for paternity analysis (Rallo *et al.*, 2000; Diaz *et al.*, 2007; Rekik *et al.*, 2008). Recently microsatellites have become available and reliable molecular markers for the traceability issues to define the olive oil origin and to detect the presence of prohibited varieties (Muzzalupo *et al.*, 2007b; Ben Ayed *et al.*, 2009). Most these publications addressed the optimization of the extraction of high quality DNA from olive oils and to identify the most interesting SSRs markers in variety discrimination. All the studies published so far, showed that the reliability and reproducibility of SSRs profiles is determined by the quality of the DNA extracted from oil (Muzzalupo *et al.*, 2007b; Breton *et al.*, 2004; Bracci *et al.*, 2011; Ben Ayed *et al.*, 2009). In fact, the amount of DNA isolated from olive oil is low and highly degraded by the nuclease present in olive oil (Muzzalupo and Perri, 2002; De la Torre *et al.*, 2004). For this reason, the extraction of DNA from olive oil is a difficult task. Several techniques of DNA preparation and immobilization for subsequent sample analysis have been developed. These methods, utilize such supports as silica, hydroxyapatite, magnetic beads, and spin columns. These supports enable the DNA to be amplified and analyzed using various quantities of oil. In particular, magnetic beads in conjunction with additional processing have proved useful. However, the defined procedure needs 2 x 40 mL of virgin olive oil, and the preparation of DNA regularly necessitates 5 h (Breton *et al.*, 2004). Besides, other authors tried various protocols of DNA extraction from olive oil such as: Wizard kit, CTAB protocol extraction, QIAamp DNA stool extraction Kit. They concluded that the most reproducible results were

obtained when the template DNA was recovered from the olive oil using QIAamp DNA stool extraction Kit (Qiagen) (Muzzalupo *et al.*, 2007b; Testolin and Lain, 2005).

3.1.4 SNP (Simple Nucleotide Polymorphism) and qRT-PCR (Quantitative Real-Time PCR)

SNP detection can be delivered in a number of ways, but the simultaneous detection of multiple SNPs from a single DNA sample is of particular interest. The "ligation detection reaction-universal array" (LDR-UA), was adopted by , to successfully genotype a panel of 49 varieties with respect to 17 SNPs. Out of the 13 amplicons containing these SNPs, 12 were successfully amplified from oil-derived template, and the resulting profiles were fully consistent with those obtained from leaf-derived DNA (Consolandi *et al.*, 2008). qRT-PCR continues to be extensively used for quantifying the amount of a specific sequence in food, with particular interest for GMOs (Marmiroli *et al.*, 2009). PDO oils are typically not monovarietal, so a method for quantifying the components of the mixture is essential if conformity with certification depends on a prescribed proportion of varietal types. So far, application of Real-Time as a tool for olive oil authentication has been explored by Giménez *et al.,* (2010). The authors evidenced that Real-Time PCR is useful to quantify DNA extracted from oil, and thus to assess the yields of different methods of extraction. But the size of amplicon, is critical for the success of analysis. A possibility of utilising qRT-PCR to quantify varieties in PDO oils rests on the use of taqMan probes designed on SNPs specific of varieties entering in the oil composition (Marmiroli *et al.*, 2009).

3.2 Genomics approaches for olive valorisation

The complete sequencing of the genome of *Arabidopsis* in 2000 by the Arabidopsis Genome Initiative (AGI) (Samir *et al.*, 2000) and the emerging sequence information for several other plant genomes, such as rice, *Populus*, *Medicago*, lotus, *Lycopersicon esculenum* and *Zea mays*, represent a valuable tool to determine the function of many genes (Rensink and Buell, 2005; Vij *et al.*, 2006). In the wake of these sequencing approach, plant research enters an exciting period in which genome-wide approaches are becoming an integral part of plant biology, with potentially highly rewarding but as yet unpredictable biotechnological applications. This is reflected in the growing interest of new farms that invest in the development of tools to enhance and expand this wealth of information.

Functional genomics employs multiple parallel approaches, including global transcript profiling coupled with the use of mutants and transgenics, to study genes function in a high throughput mode. The aim of these genome-wide efforts is to link the genome sequences to the phenotypic characters.

The availability of a large volume of genomic data has provided information about the genes content of plants. Partial or complete sequences of cDNAs often provide a firm basis of the dimension of the transcriptome. The all plant expression sequence tags (ESTs) available are organized together with well characterized genes, into non-redundant gene clusters in three main databases (National Center for Biotechnology Information, NCBI; Unigenes, http://www.ncbi.nlm.nih.gov/; The Institute for Genomic Research, TIGR; Gene Indices, www.tigr.org; and Sputnik, http://mips.gsf.de/proj/sputnik) accessible via the Internet. It is worth noting that several companies possess large private EST databases for

various crop plants such as *Zea mays* and soybean; in this case the access can be negotiated on a case by-case basis.

The ESTs are single-pass sequences of 300 to 500 bp determined from one or both ends of randomly chosen cDNA expressed genes. The sequences are sufficiently accurate to unambiguously identify the corresponding gene in most cases. Thousands of sequences can thus be determined with a limited investment. EST information present in public databases is available for a variety of species, including a number of plants (Cooke *et al.*, 1996; Yamamoto and Sasaki, 1997).

ESTs are important for the accurate genome annotation and provide information about gene structure, alternative splicing, expression patterns and transcript abundance (Umezawa *et al.*, 2004). Recent progress in DNA sequencing technology, the rapid growth of EST and cDNA sequence resources and the large amount of genetic variation at the nucleotide level can be exploited to generate various types of molecular markers for variation analysis, marker-assisted selection (MAS) and quantitative trait locus analysis (QTL) for desirable traits and to identify genetic loci involved in phenotypic changes of model and non-model plant species (Lee *et al.*, 2007).

In the absence of the complete genome sequence, EST databases are a good resource for finding genes and for interspecies sequence comparison, and have provided markers for genetic and physical mapping and clones for expression analyses. The relative abundance of ESTs in libraries prepared from different organs and plants in different physiological conditions also provides preliminary information on expression patterns for the more abundant transcripts.

Limitations at the EST approach are represented by the rare transcripts that are induced only under specific condition and consequently they are not present in EST database. In this case the only sure way to gain access to the entire set of genes is to determine the complete genomic sequence. The genomic sequence also provides information on the global structure of the genome, including the relative order of genes on the chromosomes, which is extremely valuable for positional cloning strategies. The major problem with genomic sequences is how to distinguish coding regions from noncoding intergenic sequences and introns. In this case, the comparisons between genomic sequences, ESTs and cDNA sequences can help to assign intron positions for many genes. However, for the genes that do not match sequences in the databases, the coding sequences need to be predicted from the genomic sequence. Therefore, sequencing technology applied to crop species represent the first step to identify the genes involved in the control of important agronomic traits. Rice was the first crop genome to be sequenced (Yu *et al.*, 2002; Matsumoto *et al.*, 2005), after the sequencing of the first model plant genome, Arabidopsis thaliana (Arabidopsis Genome, 2000). Current crop genome sequencing projects are rapidly changing pace with the new technology and researchers are quickly adopting second generation sequencing to gain insight into their favourite genome. Roche 454 technology is being used to sequence the 430 Mbp genome of Theobroma cacao (Scheffler *et al.*, 2009), while a combination of Sanger (old school sequencing) and Roche 454 one of the "2nd generation" technologies of sequencing is being used for the apple genome (Velasco *et al.*, 2009). A similar approach is being applied to develop a draft consensus sequence for the 504 Mbp of grape genome (Velasco *et al.*, 2007) A combined Illumina Solexa and Roche 454 sequencing approach has been used to characterize the genome of cotton (Wilkins *et al.*, 2009). Roche 454 sequencing has been used

to survey the genome of Miscanthus (Swaminathan *et al.*, 2009), while Sanger, Illumina Solexa and Roche 454 sequencing are being used to characterize the genome of banana (Hribova *et al.*, 2009).

3.2.2 Gene identification in crop species

The sequencing and assembly of large and complex crop genomes remains a valuable goal, but at the moment, a significant amount of knowledge can be gained from low coverage shotgun sequencing of these genomes. In this contest, the second generation technologies of sequencing are particularly suitable to know genes and gene promoters in crop plants that are homologous to related species. Therefore, designing polymerase chain reaction (PCR) primers to the read pairs enables the amplification and sequencing of the gene and corresponding genomic region in the target species. This approach to gene discovery offers the potential to identify genes, gene promoters and polymorphisms in a wide range of agronomically important crop species (Bracci *et al.*, 2011).

Microarray represents functional genomic approaches that have revolutionized global gene expression profiling. In fact they allow studying the entire gene complement of the genome in a single experiment (Duggan *et al.*, 1999; Li *et al.*, 2005). At the moment, cDNA and oligonucleotide microarrays have been widely used in plants, such as *Arabidopsis,* rice, maize, strawberry, petunia, ice plants and lima bean, to study and compare global gene expression levels in specific organs and/or tissues under controlled physiological conditions.

In olive, the genomics information present on the international database NCBI, concerning the identification and characterization of functional genes are prevalently based on EST identification and they are predominantly related to pollen allergens and characteristics of olive fruit.

Olea europaea trees are widely distributed throughout the Mediterranean basin and therefore their pollen is one of the most prevalent causes of respiratory allergy such as allergic rhinitis and allergic asthma in the Mediterranean region and some other countries between late April and early June (Kalyoncu *et al.*, 1995). Olive pollen is also responible of allergic inflammation of the upper and/or lower airways that may persist after the pollination season is over (Quiralte *et al.*, 2005).

Allergenic proteins name	Molecular mass (kDalton)	Family
Ole e 1	~ 19	Unknown
Ole e 2	~ 15	Profiling
Ole e 3	~ 9	Polcalcin
Ole e 4	~ 32	Unknown
Ole e 5	~ 16	Cu/Zn superoxide dismutase
Ole e 6	~ 6	Unknown
Ole e 7	~ 10	Lipid transfer protein
Ole e 8	~ 19	Ca++ binging protein
Ole e 9	~ 46	1,3 β glucanase
Ole e 10	~ 10	Carbohydrate binding protein

Table 1. The olive pollen allergens (from Villalba *et al.*, 2007)

At the moment 10 olive pollen allergens have been purified and characterized from *Olea europaea* pollen extract (Table 1). Several of these allergenic proteins, eg, Ole e 6, fail to show any homology to known protein sequences and, therefore, the biochemical function of these gene products remains unknown. Many other allergens belong to well-known families of proteins, such as profilin (Ole e 2), superoxide dismutase (Ole e 5), calciumbinding proteins (Ole e 3 and Ole e 8), lipid transfer proteins (Ole e 7) and 1,3-β-glucanases (Ole e 9) (Villalba *et al.*, 2007).

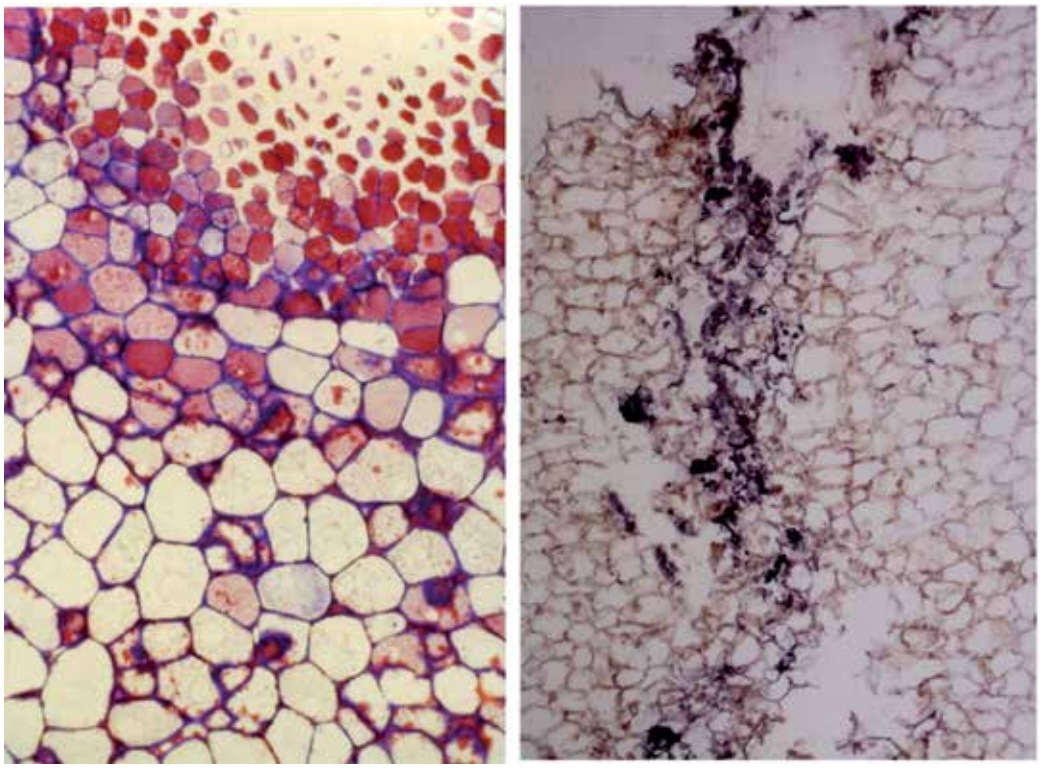


Photo 10. Cross sections of mesocarp olive fruit at level of insect injury (*Bactrocera oleae*, right: sections stained with safranin O/ azur II; left: localization of *OeCHLP* transcripts by *in situ* hybridization with dig-labelled *OeCHLP* antisense probe).

The biochemical composition of olive fruit is variable because it depends on olive variety, soil, climate, and cultivation. The virgin olive oil is overwhelmingly composed of triglycerides (>98%), along with traces of other compounds. The dominant triglyceride fatty acid species are the oleic acids (57-78%) such as palmitic, stearic, linoleic and linolenic acids (Caravita *et al.*, 2007). The other minor constituents such as alcohols, polyphenols, chlorophyll, carotenoids, sterols, tocopherols and flavonoids, contribute to the olive's organoleptic qualities, taste, flavour, and nutritional value (Perri *et al.*, 2002; Servili *et al.*, 2004). These constituents may also serve to distinguish olive oils originating from different regions. Olive oil, especially extra-virgin oil also contains small amounts of hydroxytyrosol, secoiridoids, lignans (Bianco *et al.*, 1999, 2001; De Nino *et al.*, 2005) and other compounds thought to possess anticancer properties (*i.e.*, squalene and terpenoids) (Fabiani *et al.*, 2002;

Owen *et al.*, 2004). In spite of its economical importance and metabolic peculiarities, very few data are available on gene sequences controlling the main metabolic pathways. Particular attention has been paid to the genes encoding the key enzymes involved in fatty acid biosynthesis, fatty acid modification, triacylglycerol synthesis, and fat storage (Hatzopoulos *et al.,* 2002; De la Rosa *et al.,* 2003; Banilas *et al.,* 2005).

In recent years, much attention has turned to the olive fruit. In this contest, the parallel sequencing of different fruit cDNA collections has provided large scale information about the structure and putative function of gene transcripts accumulated during fruit development (Alagna *et al.*, 2009).

A nuclear gene, named *OeCHLP* (*Olea europaea* GERANYLGERANYL REDUCTASE was isolated and characterizated by Bruno *et al.,* (2009). This gene encodes a chloroplastic enzyme involved in the formation of phytolic side chain of tocopherols chlorophyll, and plastoquinones. In olive fruits *OeCHLP* gene expression was enhanced in dark fruit very likely in relation to the increase in mature fruits of the level of total tocopherols suggesting a role in the synthesis of the antioxidant. It is noteworthy that the variations in gene transcript levels that occurred during the ripening of olive fruits depend on the genotype analyzed (Muzzalupo *et al.,* 2011). In this contest, in olive fruits tocopherols confer not only nutritional value (Valk and Hornstra, 2000), but also contribute to product stability and post harvesting shelf life (Goffman and Bohme, 2001) by protecting storage oil from oxidative damage (Sattler *et al.,* 2004). *OeCHLP* was also detected in fruits attacked by *Bactrocera oleae* pathogen as well as in fruits wounded by needle suggesting a role in protection mechanisms related to cell damage and oxidative burst induced by pathogen (photo 8 and 10) (Ebel, 1998; Klessig *et al.,* 2000; Bruno *et al.*, 2009).

4. Conclusion

Although many efforts have been made in the last years, genome studies in *Olea europaea* L. are currently behind those of other crops. Several groups have started to work on the olive genome sequencing and, thanks to the rapid development of the new sequencing technologies; hopefully soon the complete sequence of olive genome will be available. Identification of all genes within a species permits an understanding of how important agronomic traits are controlled, knowledge of which can be directly translated into crop improvement.

The availability of reliable genotype data of olive varieties and oils deriving from them, in publicly accessible curate and regularly update databases will be the challenge for the next few years. Recent advances in DNA sequencing technology are radically changing biological and biomedical research and will have a major impact on crop improvement. The new information on genome sequence will be very useful to identify genes involved in agronomical traits that could be used to improve the nutritional characteristics and the productivity of this crop. A possible application could be, for example, the studies of molecular mechanisms of drought and salinity tolerance of olive, in order to improve the cultivation of this important fruit crop also in the most arid and semiarid areas of the world. The knowledge of genome nucleotide sequences also could be useful to identify new sequence polymorphisms, which will be very useful in the development of many new variety-specific molecular markers and in the implementation of more efficient protocols for tracking and protect olive oil origin.

5. Acknowledgment

The chapter has been possible thanks to the contribution of time, energy and expertise by many individuals of CRA-OLI and Dep. Ecology of UNICAL. The authors would like to take this opportunity to acknowledge their generosity. The authors thank the International Treaty on Plant Genetic Resources for Food and Agriculture - Plant Genetic Resources (RGV-FAO) for financial support.

6. References

Alagna, F.; D'Agostino, N.; Torchia, L.; Servili, M.; Rao, R.; Pietrella, M.; Giuliano, G.; Chiusano, ML.; Baldoni, L. & Perrotta, G. (2009). Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*, Vol.10, pp. 349–353, ISSN 1471-2156.

Amane, M.; Lumaret, R.; Hany, V.; Ouazzani, N.; Debain, C.; Vivier, G. & Deguilloux, M.F. (1999). Chloroplast-DNA variation in cultivated and wild olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, Vol.99, pp. 133-139, ISSN 0040-5752.

Angiolillo, A.; Mencuccini, M. & Baldoni, L. (1999). Olive genetic diversity assessed using amplified fragment length polymorphism. *Theoretical and Applied Genetics*, Vol. 98, pp. 411–421, ISSN 0040-5752.

Baldoni, L.; Cultrera, G.; Mariotti, R.; Ricciolini, C.; Arcioni, S.; Vendramin, G.; Buonamici, A.; Porceddu, A.; Sarri, V.; Ojeda, M.; Trujillo, I.; Rallo, L.; Belaj, A.; Perri, E.; Muzzalupo, I.; Salimonti, A.; Casagrande, A.; Lain, O.; Messina, R. & Testolin, R. (2009). A common protocol for olive genotyping using microsatellite DNA: a consensus list of markers. *Molecular Breeding*, Vol.24, N. 3, pp. 213-231, ISSN 1380-3743.

Baldoni, L.; Pellegrini, M.; Mencuccini, A.; Angiolillo, A. & Mulas, M. (2000). Genetic relationships among cultivated and wild olives revealed by AFLP markers. *Acta Horticulturae*, Vol.521, pp. 275-284, ISSN 0567-7572.

Bandelj, D.; Jakse, J. & Javornik, B. (2002). Characterisation of olive (*Olea europaea* L.) cultivars by RAPD markers. *Acta Horticulturae*, Vol.586, pp. 133–135, ISSN 0567-7572.

Banilas, G.; Moressis, A.; Nikoloudakis, N. & Hatzopoulos, P. (2005). Spatial and temporal expressions of two distinct oleate desaturases from olive (*Olea europaea* L.). *Plant Science*, Vol.168, pp. 547-555, ISSN 0168-9452.

Barranco, D. & Rallo, L. (2000). Olive cultivars in Spain. *HortTechnology*, Vol.10, pp. 107- 110, ISSN 1063-0198.

Barranco, D.; Cimato, A.; Fiorino, P.; Rallo, L.; Touzani, A.; Castañeda, C.; Serafín, F. & Trujillo, I. (2000). *World Catalogue of olive varieties. International Olive Oil Council*. pp. 360, ISSN 1478-7482, Madrid, Spain.

Bartolini, G. & Petruccelli, R. (2002). *Classification, Origin, Diffusion and History of the Olive*, Food and Agriculture Organization of the United Nations, pp. 2-143, Rome, Italy,

Belaj, A.; Caballero, J.M.; Barranco, D.; Rallo, L. & Trujillo, I. (2003). Genetic characterization and identification of new accessions from Syria in an olive germoplasm bank by means of RAPD markers. *Euphytica*, Vol.134, pp. 261–268, ISSN 0014-2336.

Belaj, A.; Rallo, L.; Trujillo, I. & Baldoni, L. (2004). Using RAPD and AFLP markers to distinguish individuals obtained by clonal selection of 'Arbequina' and 'Manzanilla de Sevilla' olive. *HortScience*, Vol.39, pp. 1566–1570, ISSN 0018-5345

Belaj, A.; Trujillo, I.; de la Rosa, R.; Rallo L. & Gimenez M.J. (2001). Polymorphism and discrimination capacity of randomly amplified polymorphic markers in an olive germplasm bank. *Journal of the American Society for Horticultural Science*, Vol.126, pp. 64–71, ISSN 0003-1062.

Ben Ayed, R., Grati-Kamoun, N., Moreau, F. & Rebai, A. (2009). Comparative study of microsatellite profiles of DNA from oil and leaves of two Tunisian olive cultivars. *European Food Research and Technology*, Vol.229, pp. 757–762, ISSN 1438-2377.

Besnard, G.; Baradat, PH.; Chevalier, D.; Tagmount, A. & Berville´ A. (2001). Genetic differentiation in the olive complex (*Olea europaea*) revealed by RAPDs and RFLPs in the rRNA genes. *Genetic Resources and Crop Evolution*, Vol.48, pp. 165–182, ISSN 0925-9864.

Besnard, G.; Khadari, B.; Baradat, PH. & Berville´ A. (2002). Combination of chloroplast and mitochondrial DNA polymorphism to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics*, Vol.105, pp. 139-144, ISSN 0040-5752.

Bianco, A.; Buiarelli, F.; Cartoni, G.; Coccioli, F.; Muzzalupo, I.; Polidori, A. & Uccella N. (2001). Analysis by HPLC-MS/MS of biophenols components in olives and oils. *Analytical Letters*, Vol.36,pp. 1033-1051, ISSN 0003-2719.

Bianco, A.D.; Muzzalupo, I.; Piperno, A.; Romeo, G. & Uccella N. (1999). Bioactive derivatives of oleuropein from olives fruits. *Journal of Agriculture and Food Chemistry*, Vol.47, pp. 3531-3534, ISSN 0021-8561.

Bitonti, M.B.; Cozza, R.; Chiappetta, A.; Contento, A.; Minelli, S.; Ceccarelli, M.; Gelati, MT.; Maggini, F.; Baldoni, L. & Cionini, PG. (1999). Amount and organization of the heterochromatin in *Olea europaea* and related species. *Heredity*, Vol.83, Pp.188–195, ISSN 0018-067X.

Bogani. P.; Cavalieri, D.; Petruccelli, R. & Roselli, G. (1994). Identification of olive tree cultivars by using random amplified polymorphic DNA. *Acta Horticulturae*, Vol.356, pp. 98-101, ISSN 0567-7572.

Bracci, T.; Busconi, M.; Fogher C. & Sebastiani L. (2011). Molecular studies in olive (*Olea europaea* L.): overview on DNA markers applications and recent advances in genome analysis. *Plant Cell Reports*, Vol.30, pp. 449–462, ISSN 0721-7714.

Brennicke, A.; Klein, M.; Binder, S.; Knoop, V.; Grohmann, L.;, Malek, O.; Marchfelder, A.; Marienfeld, J. & Unseld, M. (1996). Molecular biology of plant mitochondria. *Naturwissenschaften*, Vol.83, pp. 339-346, ISSN 0028-1042.

Breton, C.; Besnard, G. & Bervillé, A. (2006). Using multiple types of molecular markers to understand olive phylogeography. In: Zeder, M.A., Decker-Walters, D., Bradley, D., Smith, B. (Eds.), *Documenting Domestication: New Genetic and Archaeological Paradigms*. Smithsonian Press, pp. 141–148, ISBN 9780520246386, Washington DC, USA.

Breton, C.; Claux, D.; Metton, I.; Skorski, G. & Berville, A. (2004). Comparative study of methods for DNA preparation from olive oil samples to identify cultivar SSR alleles in commercial oil samples: possible forensic applications. *Journal of Agriculture and Food Chemistry*, Vol.52, pp. 531-537, ISSN 0021-8561.

Bruno L.; Chiappetta A.; Muzzalupo I.; Gagliardi C.; Iaria D.; Bruno A.; Greco M.; Donato G.; Perri E. & Bitonti MB. (2009) Role of geranylgeranyl reductase gene in organ development and stress response in olive (*Olea europaea*) plants. *Functional Plant Biology*, Vol.36, pp. 370–381, ISSN 1445-4408.

Busconi, M.; Foroni, C.; Corradi, M.; Bongiorni, C.; Cattapan, F.; Fogher, C. (2003). DNA extraction from olive oil and its use in the identification of the production cultivar. *Food Chemistry*, Vol.83, Pp.127–134, ISSN 0308-8146.

Cantini, C.; Cimato, A. & Sani, G. (1999). Morphological evaluation of olive germplasm present in Tuscany region. *Euphytica*, Vol.109, Pp.173-181, ISSN 0014-2336.

Caravita, M.; Benincasa, C.; De Rose, F.; Muzzalupo, I.; Parise, A.; Pellegrino, M.; Perri, E. & Rizzuti, B. (2007). Omega-3/omega-6 fatty acids ratio in olive oils from Italian olive varieties. *AgroFOOD Industry Hi-Tech*, Vol.6, pp. 17-18, ISSN 1722-6996.

Carriero, F.; Fontanazza, G.; Cellini, F. & Giorio, G. (2002). Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L). *Theoretical and Applied Genetics*, Vol.104, pp. 301-307, ISSN 0040-5752.

Cavallotti, A.; Regina, TMR. & Quagliariello, C. (2003). New sources of cytoplasmic diversity in the Italian population of *Olea europaea* L. as revealed by RFLP analysis of mitochondrial DNA: characterization of the *cox3* locus and possible relationship with cytoplasmic male sterility. *Plant Science*, Vol.164, pp. 241-252, ISSN 0168-9452.

Cipriani, G., Marazzo, M.; Marconi, R.; Cimato, A. & Testolin, R. (2002). Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theoretical and Applied Genetics*, Vol.104, pp. 223-228, ISSN 0040-5752.

Consolandi, C.; Palmieri, L.; Severgnini, M.; Maestri, E.; Marmiroli, N.; Agrimonti, C.; Baldoni, L.; Donini, P.; Bellis, G. & Castiglioni, B. (2008). A procedure for olive oil traceability and authenticity: DNA extraction, multiplex PCR and LDR-universal array analysis. *European Food Research and Technology*, Vol.227, pp. 1429–1438, ISSN 1438-2377.

Cooke, R.; Raynal, M.; Laudié, M.; Grellet, F.; Delseny, M.; Morris, PC.; Guerrier, D.; Giraudat, J.; Quigley, F. & Clabault, G. (1996). Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs. *Plant Journal*, Vol.9, pp. 101–124, ISSN 0960-7412.

Costa, F.; Alba R.; Schouten, H.; Soglio, V.; Gianfranceschi, L.; Serra, S.; Musacchi, S.; Sansavini, S.; Costa, G.; Fei, Z. & Giovannoni, J. (2010). Use of homologous and heterologous gene expression profiling tools to characterize transcription dynamics during apple fruit maturation and ripening. *BMC Plant Biology*, Vol.10, pp.349-353, ISSN 1471-2229.

De la Rosa, R.; Angiolillo, A.; Rallo, L.; Guerrero, C.; Pellegrini, M.; Besnard, G.; Bervillé, A.; Martin, A. & Baldoni, L. (2003). A first genetic linkagemap of olive (*Olea europaea* L.) cultivars using RAPD and AFLP markers. *Theoretical and Applied Genetics*, Vol.106, pp. 1273-1282, ISSN 0040-5752.

De la Torre, F.; Bautista, R.; Canovas, FM. & Claros, MG. (2004). Isolation of DNA from olive oil and oil sediments: application in oil fingerprinting. *Journal of Food Agriculture and Environment*, Vol.2, pp. 84–86, ISSN 1459-0255.

De Nino, A.; Di Donna, L.; Mazzotti, F.; Muzzalupo, I.; Perri, E.; Sindona, G. & Tagarelli, A. (2005). Absolute method for the assay of oleuropein in olive oils by atmospheric

pressure chemical ionization tandem mass spectrometry. *Analytical Chemistry*, Vol.77, pp. 5961-5964, ISSN 0003-2700.

Diaz, A.; Martin, A.; Rallo, P. & de la Rosa, R. (2007). Cross-compatibility of the parents as main factor for successful olive breeding crosses. *Journal of the American Society for Horticultural Science*, Vol.132, pp. 830–835, ISSN 0003-1062.

Duggan, DJ.; Bittner, M.; Chen, Y.; Meltzer, P. & Trent, JM. (1999). Expression profiling using cDNA microarrays. *Nature Genetics*, Vol.21, pp. 10-21, ISSN 1061-4036.

Ebel, J. (1998). Oligoglucoside elicitor-mediated activation of plant defense. *BioEssays* 20, 569–576, ISSN 0265-9247.

Fabbri, A.; Hormaza, JI. & Polito, VS. (1995). Random amplified polymorphic DNA analysis of olive (*Olea europaea* L.) cultivars. *Journal of the American Society for Horticultural Science*, Vol.120, pp. 538–542, ISSN 0003-1062.

Fabiani, R.; De Bartolomeo, A.; Rosignoli, P.; Servili, M.; Montedoro, GF. & Morozzi, G. (2002). Cancer chemoprevention by hydroxytyrosol isolated from virgin olive oil through G1 cell cycle arrest and apoptosis. *European Journal of Cancer Prevention*, Vol.11, pp. 351-358, ISSN 0959-8278.

FAO 2010. *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. ISBN 978-92-5-106534-1, Rome, Italy.

Fritsch, P. & Rieseberg, L.H. (1996). The use of random amplified polymorphic DNA (RAPD) in conservation genetics. In: Smith T.B. and Wayne R.K. (eds), *Molecular Genetics Approaches in Conservation*. Oxford University press, pp. 54–73, ISBN 0-19-509526-x, New York, USA.

Gemas, VJ.; Almadanim, MC.; Tenreiro, R.; Martins, A. & Fevereiro, P. (2004). Genetic diversity in the Olive tree (*Olea europaea* L. subsp. *europaea*) cultivated in Portugal revealed by RAPD and ISSR markers. *Genetic Resources and Crops Evolution*, Vol.51, pp. 501–511, ISSN 0925-9864.

Goffman, FD. & Bohme, T. (2001). Relationship between fatty acid profile and vitamin E content in maize hybrids (*Zea mays* L.). *Journal of Agricultural and Food Chemistry*, Vol.49, pp. 4990–4994, ISSN 0021-8561.

Grati Kamoun, N.; Lamy Mahmoud, F.; Rebai, A.; Gargouri, A.; Panaud, O. & Saar, A. (2006). Genetic diversity of Tunisian olive tree (*Olea europaea L.*) cultivars assessed by AFLP markers. *Genetic Resources and Crop Evolution*, Vol.53, pp. 265–275, ISSN 0925-9864.

Green P.S. (2002). A revision of *Olea* (L.). *Kew Bulletin*, Vol.57, pp. 91–140, ISSN 0075-5974.

Guneser, S.; Atici, A.; Cengizler, I. & Alparslan, N. (1996). Inhalant allergens: as a cause of respiratory allergy in east Mediterranean area, Turkey. *Allergol Immunopathol* (Madr). Vol.24, pp. 116-119, ISSN 0301-0546.

Hatzopoulos, P.; Banilas, G.; Giannoulia, K.; Gazis, F.; Nikoloudakis, N.; Milioni, D. & Haralampidis, K. (2002). Breeding, molecular markers and molecular biology of the olive tree. *European Journal of Lipid Science and Technology*, Vol.104, pp. 574-586, ISSN 1438-7697.

Hernandez, P.; de la Rosa, R.; Rallo, L.; Dorado, G. & Martin, A. (2001). Development of SCAR markers in olive (*Olea europaea*) by direct sequencing of RAPD products: applications in olive germplasm evaluation and mapping. *Theoretical and Applied Genetics*, Vol. 103, pp. 788-791, ISSN 0040-5752.

Hess, J.; Kadereit, W. & Vargas, P. (2000). The colonization history of *Olea europaea* L. in Macaronesia based on internal transcribed spacer 1 (ITS-1) sequences, randomly amplified polymorphic DNAs (RAPD) and inter-simple sequence repeats (ISSR). *Molecular Ecology*, Vol.9, pp. 857–868, ISSN 0962-1083.

Hofte, H.; Desprez, T.; Amselem, J.; Chiapello, H.; Caboche, M.; Moisan, A.; Jourjon, MF.; Charpenteau, JL.; Berthomieu, P. & Guerrier, D. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant Journal*, Vol.4, pp. 1051–1061, ISSN 0960-7412.

Hribova, E.; Neumann, P.; Macas, J. & Dolezel, J. (2009). Analysis of genome structure and organization in banana (*Musa acuminata*) using 454 sequencing. In: *Plant and Animal Genomes XVII*, San Diego, CA, USA.

Jobling, M.A. & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Review Genetics*, Vol.5, pp. 739- 51, ISSN 1471-0056.

Kalyoncu, A.F.; Coplu, L.; Selcuk, Z.T.; Emri, A.S.; Kolacan, B.; Kocabas, A.; Akkoclu, A.; Erkan, L.; Sahin A.A. & Baris, Y.I. (1995). Survey of the allergic status of patients with bronchial asthma in Turkey: a multicenter study. *Allergy*, Vol.50, pp. 451-455, ISSN 1398-9995.

Kaul, S.; Koo, HL.; Jenkins, J.; Rizzo, M.; Rooney, T.; Tallon, LJ.; Feldblyum, T. & *et al.,* (2000). *Nature*, Vol.408, pp. 796-815, ISSN 0028-0836.

Khadari, B.; Breton, C.; Moutier, N.; Roger, P.J.; Besnard, G.; Berville´, A. & Dosba, F. (2003). The use of molecular markers for germplasm management in a French olive collection. *Theoretical and Applied Genetics*, Vol.106, pp. 521-529, ISSN 0040-5752.

Klessig, DF.; Durner, J.; Noad, R.; Navarre, DA.; Wendehenne, D. *et al.,* (2000). Nitric oxide and salicylic acid signaling in plant defense. *Proceeding of the National Academy of Sciences of the United States of America,* Vol.97, pp. 8849–8855, ISSN-0027-8424.

Lee, YP.; Yu, GH.; Seo, YS.; Han, SE.; Choi, YO.; Kim, D.; Mok, IG.; Kim, WT. & Sung. SK. (2007). Microarray analysis of apple gene expression engaged in early fruit development. *Plant Cell Reports*, Vol.26, pp. 917-926, ISSN 0721-7714.

Li, L.; Wang, X.; Xia, M.; Stolc, V.; Su, N.; Peng, Z.; Li, S.; Wang, J.; Wang, X. & Deng, X.W. (2005). Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. *Genome Biology*, Vol.6, R52, ISSN 1465-6906.

Lombardo, N.; Godino, G.; Alessandrino, M.; Belfiore, T. & Muzzalupo, I. (2004). *Contributo alla Caratterizzazione del Germoplasma Olivicolo Pugliese*, Istituto Sperimenale per l'Olivicoltura (Ed), Italy, Pp 2-112, Cosenza, Italy.

Lombardo, N.; Madeo, A.; Muzzalupo, I.; Ciliberti, A.; Godino, G. (2006). Osservazioni comparate sulla precocità di entrata in produzione di 70 cultivar di olivo italiane. *Italus Hortus*, Vol.13, pp. 133-136, ISSN 1127-3496.

Lopes, M.S.; Mendonc¸ A.D.; Sefc, M.K.; Gil, F.S. & Da Câmara Machado, A. (2004). Genetic evidence of intra-variety variability within Iberian olive varieties. *HortScience*, Vol.39, pp. 1562–1565, ISSN 0018-5345.

Lumaret, R.; Amane, M.; Ouazzani, N.; Baldoni, L. & Debain, C. (2000). Chloroplast DNA variation in the cultivated and wild olive taxa of the genus *Olea* L. Theoretical and Applied Genetics, Vol.101, pp. 547-553, ISSN 0040-5752.

Marmiroli, N.; Maestri, E.; Pafundo. S. & Vietina, M. (2009). Molecular traceability of olive oil: from plant genomics to Food Genomics. In: Berti L, Maury J (eds) *Advances in olive resources*. Research Signpost, pp. 1–16, ISBN 978-81-7895-388-5, Kerala, India.

Martins-Lopes, P.; Gomes, S.; Santos, E. & Guedes-Pinto, H. (2008). DNA markers for Portuguese olive oil fingerprinting. *Journal of Agriculture and Food Chemistry*, Vol.56, pp. 11786-11791, ISSN 0021-8561.

Matsumoto, T.; Wu, J.Z.; Kanamori, H.; Katayose, Y.; Fujisawa, M.; Namiki, N. & *et al.*, (2005). The map-based sequence of the rice genome. *Nature*, Vol.436, pp. 793–800, ISSN 0028-0836.

Mekuria, G.; Collins, G.; Lavee, S. & Sedgley, M. (2001). Identification of a RAPD marker linked to tolerance to peacock disease (*Spilocaea oleaginea*) in olives (*Olea europaea*). *Acta Horticulturae*, Vol.546, pp. 565-567, ISSN 0567-7572.

Minelli, S.; Maggini, F.; Gelati, MT.; Angiolillo, A. & Cionini, P.G. (2000). The chromosome complement of *Olea europaea* L.: characterization by differential staining of the chromatin and *in situ* hybridization of highly repeated DNA sequences. *Chromosome Research*, Vol.8, pp. 615-619, ISSN 0967-3849.

Montemurro, C.; Pasqualone, A.; Simeone, R.; Sabetta, W. & Blanco, A. (2007). AFLP molecular markers to identify virgin olive oils from single Italian cultivars. *European Food Research and Technology*, Vol.226, pp. 1439–1444, ISSN 1438-2377.

Muzzalupo, I. & Perri, E. (2002). Recovery and characterization of DNA from virgin olive oil. *European Food Research and Technology*, Vol.214, pp. 528–531, ISSN 1438-2377.

Muzzalupo, I. & Perri, E. (2009). Genetic diversity in olive tree varieties from Italy and other countries of the Mediterranean basin as revealed by RAPD and SSR molecular marker. *Advances in Horticultural Science*, Vol.23, pp. 263–275, ISSN 1592-1573.

Muzzalupo, I.; Chiappetta, A.; Benincasa, C. & Perri, E. (2010). Intra cultivar variability of three major olive cultivars grown in different areas of central-southern Italy and studied using microsatellite markers. *Scientia Horticulturae*, Vol.126, N. 3, pp. 324-329, ISSN 0304-4238.

Muzzalupo, I.; Fodale, A.; Mulè, R.; Caravita, M.A.; Salimonti, A.; Pellegrino, M. & Perri, E. (2007b). Genetic relationships of Sicilian olive varieties using RAPD markers. Advances in Horticultural Science, Vol.21, pp. 35–40, ISSN 1592-1573.

Muzzalupo, I.; Lombardo, N.; Musacchio, A.; Noce, M.E.; Pellegrino, G.; Perri, E. & Sajjad, A. (2006). DNA sequence analysis of microsatellite markers enhances their efficiency for germplasm management in a Italian olive collection. *Journal of the American Society for Horticultural Science*, Vol.131, pp. 352–359, ISSN 0003-1062.

Muzzalupo, I.; Lombardo, N.; Salimonti, A. & Perri, E. (2008). Molecular characterization of Italian olive cultivars by microsatellite markers. Advances in horticultural science, Vol.22, pp. 142-148, ISSN 1592-1573.

Muzzalupo, I.; Pellegrino, M. & Perri, E. (2007a). Detection of DNA in virgin olive oils extracted from destoned fruits. *European Food Research and Technology*, Vol.224, pp. 469–475, ISSN 1438-2377.

Muzzalupo, I.; Stefanizzi, F. & Perri, E. (2009a). Evaluation of olives cultivated in Southern Italy by SSR Markers. *HortScience* Vol.44, pp. 582–588, ISSN 0018-5345.

Muzzalupo, I.; Stefanizzi, F.; Perri E. & Chiappetta, A. (2011). Transcript Levels of *CHL P* gene, antioxidants and chlorophylls contents in olive (*Olea europaea* L.) pericarps: a

comparative study on eleven olive cultivars harvested in two ripening stages *Plant Foods for Human Nutrition*, Vol.66, pp. 1-10, ISSN 0921-9668.

Muzzalupo, I.; Stefanizzi, F.; Salimonti, A.; Falabella, R. & Perri, E. (2009b). Microsatellite markers for identification of a group of Italian olive accessions. *Scientia Agricola*, Vol.66, pp. 685–690, ISSN 0103-9016.

Newman, T.; de Bruijn, FJ.; Green, P.; Keegstra, K.; Kende, H.; McIntosh, L.; Ohlrogge, J.; Raikhel, N.; Somerville, S. & Thomashow, M. (1994). Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiology*, Vol.106, pp. 1241–1255, ISSN 0032-0889.

Ouazzani, N.; Lummaret, R. & Villemur, P. (1995). Apport du polymorphisme alloenzymatique à l'identification variétale de l'olivier (*Olea europaea* L). *Agronomie*, Vol.15, pp. 31-37, ISSN 0249-5627.

Owen, CA.; Bita, EC.; Banilas, G.; Hajjar, SE.; Sellianakis, V.; Aksoy, U.; Hepaksoy, S.; Chamoun, R.; Talhook, SN.; Metzidakis, I.; Hatzopoulos, P. & Kalaitzis, P. (2005). AFLP reveals structural details of genetic diversity within cultivated olive germplasm from the Eastern Mediterranean. *Theoretical and Applied Genetics*, Vol.110, pp. 1169-1176, ISSN 0040-5752.

Owen, RW.; Haubner, R.; Wurtele, G.; Hull, WE.; Spiegelhalder, B. & Bartsch, H. (2004). Olives and olive oil in cancer prevention. *European Journal of Cancer Prevention*, Vol.13, pp. 319-326, ISSN 0959-8278.

Pafundo, S.; Agrimonti, C. & Marmiroli, N. (2005). Traceability of plant contribution in olive oil by amplified fragment length polymorphisms. *Journal of Agricultural and Food Chemistry*, Vol.53, pp. 6995–7002, ISSN 0021-8561.

Papadia, P.; Del Coco, L.; Muzzalupo, I.; Rizzi, M.; Perri, E.; Cesari, G.; Simeone, V.; Mondelli, F.; Schena, F.P. & Fanizzi, F.P. (2011). Multivariate analysis of ^{1}H-NMR spectra of genetically characterized extra virgin olive oils and growth soil correlations. *Journal of the American Oil Chemists Society*, Vol.88, N. 10, pp. 1463-1475, ISSN 0003-021X.

Pasqualone, A.; Caponio, F. & Blanco, A. (2001). Inter-simple sequence repeat DNA markers for identification of drupes from different *Olea europaea* L. cultivars. *European Food Research and Technology*, Vol.213, pp. 240-243, ISSN 1438-2377.

Perri, E.; Lombardo, N.; Urso, E.; Muzzalupo, I.; Rizzuti, B.; Pellegrino, M. & Cirillo, N. (2002). Confronto tra olii monovarietali ottenuti con i metodi biologico e convenzionale. In: (Ed.) *Caratterizzazione degli olii vergini di oliva da agricoltura biologica in Calabria*, A.R.S.S.A., pp. 29-32, Regione Calabria, Italy.

Poghosyan, ZP.; Haralampidis, K.; Martinkovskaya, AI.; Murphy, DJ. & Hatzopoulos, P. (1999). Developmental regulation and spatial expression of a plastidial fatty acid desaturase from *Olea europaea*. *Plant Physiology and Biochemistry*, Vol.37, pp. 109-119, ISSN: 0981-9428.

Qian, W.; Ge, S. & Hong, D.Y. (2001). Genetic variation within and among populations of a wild rice *Oryza granulata* from China detected by RAPD and ISSR markers. *Theoretical and Applied Genetics*, Vol.102, pp. 440-449, ISSN 0040-5752.

Quiralte, J.; Llanes, E.; Barral, P.; Arias de Saavedra, JM.; Saenz de San Pedro, B.; Villalba, M.; Florido, JF.; Rodriguez, R.; Lahoz, C. & Cardaba, B. (2005). Ole e 2 and Ole e 10: new clinical aspects and genetic restrictions in olive pollen allergy. Allergy, Vol.60, pp. 360-365, ISSN 1398-9995.

Quiralte, J.; Palacios, L.; Rodríguez, R.; Cárdaba, B.; Arias de Saavedra, JM.; Villalba, M.; Florido, JF. & Lahoz C. (2007). Modelling diseases: The allergens of *Olea europaea* pollen. *Journal of Investigational Allergology and Clinical Immunology,* Vol. 17, Supplement 1: pp. 76-82, ISSN 1018-9068.

Rallo, P.; Dorado, G. & Martin, A. (2000). Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). *Theoretical and Applied Genetics*, Vol.101, pp. 984–989, ISSN 0040-5752.

Reale, S.; Doveri, S.; Díaz, A.; Angiolillo, A.; Lucentini, L.; Pilla, F.; Martín, A.; Donini, P. & Lee, D. (2006). SNP-based markers for discriminating olive (*Olea europaea* L.) cultivars. *Genome*, Vol.15, pp. 1193-1209, ISSN 0831-2796.

Rekik, I.; Grati-Kammoun, N.; Makhloufi, E. & Rebai, A. (2010). Discovery and potential of SNP markers in characterization of Tunisian olive germplasm. *Diversity*, Vol.2, pp. 17-27, ISSN 1424-2818.

Rekik, I.; Salimonti, A.; Grati-Kamoun, N.; Muzzalupo, I., Perri, E. & Rebai, A. (2008). Characterisation and identification of Tunisian olive tree varieties by microsatellite markers. *HortScience* Vol.43, pp. 1371–1376, ISSN 0018-5345.

Rensink, W.A. & Buell, C.R. (2005). Microarray expression profiling resources for plant genomics. *Trends in Plant Science*, Vol.10, pp.603-609, ISSN 1360-1385.

Roselli, G. & Scaramuzzi, F., (1974). Definizione dei termini di cultivar e clone. vol. 2. *Annali dell'Accademia Nazionale di Agricoltura*, pp. 119–126, ISSN 00014443, Bologna, Italy.

Rugini E. (1984). *In vitro* propagation of some olive (*Olea europaea* sativa L.) cultivars with different root-ability, and medium development using analytical data from developing shoots and embryos. *Scientia Horticulturae*, Vol.24, pp. 123-134, ISSN 0304-4238.

Rugini, E.; Biasi, R. & Muleo, R. (2000). Olive (*Olea europaea* var. sativa) Transformation. In: *Molecular Biology of Woody Plants* Vol.2. S.M. Jain and S.C. Minocha (eds), Kluwer Academic Publishers, pp. 245-279, ISBN 978-0-7923-6241-8, Massachusetts, USA.

Sabino Gil, F.; Busconi, M.; Da Camara Machado, A. & Fogher, C. (2006). Development and characterization of microsatellite loci from *Olea europaea*. *Journal of Molecular Ecology Notes*, Vol.6, pp. 1275–1277, ISSN: 1755-0998.

Sanz-cortés, F.; Badenes, M.L.; Paz, S., Iñiguez, A. & Laàcer, G. (2001). Molecular characterization of olive cultivars using RAPD markers. *Journal of the American Society for Horticultural Science*, Vol.126, pp. 7-12, ISSN 0003-021X

Sarri, V.; Baldoni, L.; Porceddu, A.; Cultrera, N.G.M.; Contento, A.; Frediani, M.; Belaj, A.; Trujillo, I. & Cionini, P.G. (2006). Microsatellite markers are powerful tools for discriminating among olive varieties and assigning them to geographically defined populations. *Genome*, Vol.49, pp. 1606–1615, ISSN 0831-2796.

Sattler, SE.; Gilliland, LU.; Magallanes-Lundback, M.; Pollard, M. & DellaPenna, D. (2004). Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *The Plant Cell*, Vol.16, 1419–1432, ISSN 1040-4651.

Scheffler, B.E.; Kuhn, D.N.; Motamayor, J.C. & Schnell, R.J. (2009). Efforts towards sequencing the Cacao genome (*Theobroma cacao*). In*: Plant and Animal Genomes XVII*, San Diego, CA, USA.

Sefc, KM.; Lopes, MS.; Mendonca, D.; Rodrigues dos Santos, M. & Laimer da Camara Machado, A. (2000). Identification of microsatellite loci in olive (*Olea europaea* L.)

and their characterization in Italian and Iberian olive trees. *Molecular Ecology*, Vol.9, pp. 1171-1173, ISSN 0962-1083.

Servili, M.; Selvaggini, R.; Esposto, S.; Taticchi, A.; Montedoro, G. & Morozzi, G. (2004). Health and sensory properties of virgin olive oil hydrophilic phenols: agronomic and technological aspects of production that affect their occurrence in the oil. *Journal of Chromatography A*, Vol.1054, pp. 113-127, ISSN 0021-9673.

Stergiou, G.; Katsiotis, A.; Hagidimitriou, M. & Loukas, M. (2002). Genomic and chromosomal organization of Ty1-*copia*like sequences in *Olea europaea* and evolutionary relationships of *Olea* retroelements. *Theoretical and Applied Genetics*, Vol.104, pp. 926–933, ISSN 0040-5752.

Swaminathan, K.; Varala, K.; Moose, S.P.; Rokhsar, D.; Ming, R. & Hudson, M.E. (2009). A genome survey of *Miscanthus Giganteus*. In: *Plant and Animal Genomes XVII*, San Diego, CA, USA.

Testolin, R. & Lain, O. (2005). DNA extraction from olive oil and PCR amplification of microsatellite markers. *Journal of Food Science*, Vol.70, pp. 108-112, ISSN 0022-1147.

Umezawa, T.; Yoshida, R.; Maruyama, K.; Yamaguchi-Shinozaki, K. & Shinozaki, K. (2004). SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences USA*, Vol.101, pp. 17306-17311, ISSN-0027-8424.

Valk, EE. & Hornstra, G. (2000). Relationship between vitamin E requirement and polyunsaturated fatty acid intake in man. *International Journal for Vitamin and Nutrition Research*, Vol. 70, pp. 31–42, ISSN 0300-9831.

Velasco, R.; Zharkikh, A.; Troggio, M.; Cartwright, D.A.; Cestaro, A.; Pruss, D.; Pindo, M.; FitzGerald, L.M.; Vezzulli, S. & Reid, J. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE*, Vol.2, pp. e1326, ISSN 1932-6203.

Velasco, R.; Zharkikh, A.; Troggio, M.; Salvi, S. & Pindo, M. (2009). Apple genome sequencing and post-genomic program at IASMA research center. In: *Plant and Animal Genomes XVII*, San Diego, CA, USA.

Vij, S.; Gupta, V.; Kumar, D.; Ravi, V.; Raghuvanshi, S.; Khurana, P.; Khurana, J.P. & Tyagi, A.K. (2006). Decoding the rice genome. *Bioessays*, Vol. 28, pp. 421–432, ISSN 0265-9247.

Villalba, M.; Batanero, E.; Monsalve, RI.; Gonzalez de la Pena, MA.; Lahoz, C. & Rodriguez, R. (1994). Cloning and expression of Ole e I, the major allergen from olive tree pollen. Polymorphism analysis and tissue specificity. *The Journal of Biological Chemistry*, Vol.269, pp. 15217–15222, ISSN0021-9258.

Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; Van de Lee, T.; Hornes, M.; Frijters, A.; Pot, J.; Peleman, J.; Kuiper, M. & Zabeau, M. (1995). AFLP: a new technique for DNA Fingerprinting. *Nucleic Acids Research*, Vol.23, pp. 4407-4414, ISSN 0305-1048.

Wiesman, Z.; Avidan, N.; Lavee, S. & Quebedeaux, B. (1998). Molecular characterization of common olive varieties in Israel and the West bank using randomly amplified polymorphic DNA (RAPD) markers. *Journal of the American Society for Horticultural Science*, Vol.123, pp. 837–841, ISSN 0003-1062.

Williams, JG.; Kubelik, AR.; Livak, KJ.; Rafalski, JA. & Tingey, SV. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, Vol.18, pp. 6531-6535, ISSN 0305-1048.

Wolfe, KH.; Li, WH. & Sharp, PM. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear. *Proceedings of the National Academy of Sciences USA*, Vol.84, pp. 9054-9058, ISSN-0027-8424.

Wu, SB.; Collins, G. & Sedgley, M. (2004). A molecular linkage map of olive (*Olea europaea* L.) based on RAPD, microsatellite, and SCAR markers. *Genome*, Vol.47, pp. 26–35, ISSN 0831-2796.

Yamamoto, K. & Sasaki, T. (1997). Large-scale EST sequencing in rice. *Plant Molecular Biology*, Vol.35, pp.135-144, ISSN 1573-5028.

Yu, J.; Hu, S.N.; Wang, J.; Wong, G.K.S.; Li, S.G.; Liu, B.; Deng, Y.J.; Dai, L.; Zhou, Y.; Zhang, X.Q.; Cao, M.L.; Liu, J.; Sun, J.D.; & *et al.,* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). *Science*, Vol.296, pp. 79–92, ISSN 0036-8075.

5

Assessment and Utilization of the Genetic Diversity in Rice (*Orysa sativa* L.)

Jin Quan Li and Peng Zhang
State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University
China

1. Introduction

The basis for raising crop production and improving crop quality is to breed new varieties. The key to breed new varieties are largely depended on the breakthrough of mining the crop germplasm resources. Therefore, the research and utilization of crop genetic diversity plays an important role on crop improvement in the future. Previous researches have indicated that genetic bottleneck effects existed in the procedure of crop domestication and modern breeding, i.e. the allele variation within wild species and landrace would be lost and result in the reduction of gene diversity during domestication and breeding (Tanksly et al., 1997). The narrow genetic basis would lead to cultivars without resistance to new pets and virus and tolerance to bad environment as well as producing the platform effect of yield. These lost alleles in modern cultivars could only trace back to their original landrace and wild species and be recovered. The original landrace are close to cultivars and possess high genetic diversity and many exotic genes, therewith provide useful germplasm resources for crop breeding.

Identification, uses and conservation for the genetic diversity within crop germplasm resources are of importance for their sustainable use in plant breeding. The current rapidly development of bioinformatics, genomics, and molecular biology as well as conventional breeding methods provides useful means to mine the desirable genes in the resources.

Rice (*Oryza sativa* L.) feeds more than 50% of the world's population and is one of the most important crops in the world. Rice genetic resource is the primary material for rice breeding and makes a concrete contribution to global wealth creation and food security. Therefore, understanding its valuable genetic diversity and using it in rice genetic improvement is of importance for raising rice yield and the resistance to biotic and abiotic stress as well as improving rice quality to secure global food supplies. Furthermore, as a model plant of cereal family, two rice genome sequence map have been generated (Goff et al., 2002; Yu et al., 2002) and great progress has been made in gene mining with omics technology. Such researches helps to make use of rice genetic resources however in turn requires to make further insights of rice genetic diversity.

China is well known as an origin center of cultivated rice, with abundant rice genetic resources. As early as 1920-1964, Ying Ting, An academician of Chinese academy of science, collected more than 7128 rice landrace from all over China as well as some main rice

cultivated countries. As far as we know, the collection is one of the earliest collections for rice germplasm resources and therefore we named it as Ting's collection (Lu et al., 2006). Therefore, this chapter aims to explore effective methods on mining the exotic genes within these novel rice germplasm resources.

2. Ting's rice germplasm collection

The Ting's rice germplasm collection consists of 7128 accessions, which was collected and conserved by Academician and Professor Ying Ting during 1920-1964 from 20 different provinces of China as well as from North Korea, Japan, Philippines, Brazil, Celebes, Java, Oceania, and Vietnam (Fig. 1). Most of them are rice landraces and possess high genetic diversity. Due to that it is one of the earliest systematically rice collections in China and covered most of the Chinese rice cultivated regions, it could serve as an representative for the genetic diversity of Chinese rice germplasm resources.

Most accessions were characterized for taxonomical, geographical, morphological and agronomical descriptors, recorded by the previous laboratory of rice ecology of Chinese academy of agricultural sciences, South China Agricultural College and Guangdong academy of agricultural sciences, China (1961-1965). These recorded traits include 20 unordered qualitative traits, i.e. origin of variety, *indica* vs. *japonica*, paddy vs. upland, waxy vs. non-waxy, grain shape, rice color, grain quality, leaf color, leaf margin color, leaf cushion color, auricle color, inner sheath color, outer sheath color, stem color, leaf-green color, stigma color, glume-tip color, sterile lemma color and glume color; 14 ordered qualitative traits, i.e. early- or late-season, type of maturity, shattering habit, awn, awn length, leaf face pubescence, leaf base pubescence, flag-leaf angle, erect vs. bending leaf, compact vs. loose stem, panicle shape, compact vs. loose rachis-braches, sparse vs. dense glume hair, compact vs. loose glume hair; and 15 quantitative traits, i.e. culm length, culm size, thickness of culm wall, the second internode length, number of panicles per plant, panicle length, panicle size, number of seeds per panicle, grain length, grain length/width ratio, grain size, flag leaf length, flag leaf width, length of elongated uppermost internode, grouth duration. These data provide a good basis for studying their phenotypic genetic diversity as well as core collection construction based on the phenotypes.

3. Genetic diversity of phenotypes of Chinese rice germplasm resources

About 6500 accessions of rice germplasm resources from the Ting's collection with well passport data were selected and studied for their genetic diversity of phenotypes. The origin, type and distribution of these accessions are listed in Table 1. All the studied rice accessions were classified as five regions, i.e. South China, Central China, Southwest China, Northwest China, and North China.

The genetic diversity index (I) was calculated by: $I = \frac{-\sum_i \sum_j P_{ij} Log P_{ij}}{N}$, where P_{ij} are the frequency of jth phenotypes for ith traits, and N is the total number of traits.

The average genetic diversity index for the five rice cultivated regions from minimum to maximum are: Southwest China > Central China > Northwest China > South China > North

China for Unordered qualitative traits, North China > Northwest China > South China > Central China > Southwest China for ordered qualitative traits, and South China > North China > Central China > Northwest China > Southwest China for quantitative traits (Fig.2).

Fig. 1. Rice germplasm resources in Ting's collection and their regeneration. Upper left, preparing seeds for sowing; upper middle, sowing in the nursery field; upper right, transplanted in the field; bottom left, measuring the traits and harvest; bottom middle, a global view for the germplasm regenerating field and their genetic diversity; bottom right, seed cool room with air condition, where the rice seeds are conserved in the block jars.

Region	*Indica*	*Japonica*	Paddy	Upland	Waxy	Non-Waxy	Early-seasonal	Late-seasonal
South China	2250	247	2311	58	2304	180	1094	1276
Central China	1450	343	1783	9	1720	73	1415	378
Southwest China	887	768	767	286	715	200	130	338
Northwest China	75	294	363	0	342	21	363	0
North China	70	125	189	6	177	16	195	0
Total	4732	1777	5413	359	5258	490	3197	1992

Table 1. The origin, type and distribution of rice germplasm resources

The results show that the most abundant genetic diversity for rice qualitative traits among the five rice cultivated regions are Southwest China, which is relevant to the geographic location and climate in Yunnan province and Guizhou province of China. The reasons for it might be the high variation in different sea levels (76~2700m) which results in several different climate zone. However, for the genetic diversity in quantitative traits, South and Central China are the highest ones. The reason might be that South and central China are

the main rice production area in our country traditionally and the quantitative traits were greatly improved by the farmers and breeders through thousands of years selection.

The results indicated that the rice germplasm resources from Southwest China might help to raise the rice genetic diversity for qualitative traits, such as grain colors, grain quality, etc. For the improvement of quantitative trains, the rice germplasm resources from South and Central China might contribute to the aim more than other regions.

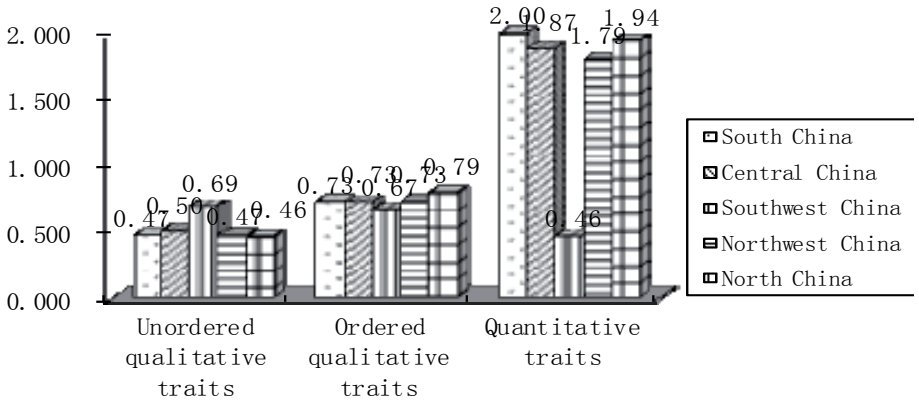


Fig. 2. The average genetic diversity index for rice germplasm resources from different cultivated regions

4. Genome-wide distribution of genetic diversity assessed with SSR markers

A subset containing 150 accessions were taken from the whole collection (described below) and were genotyped with 274 genome-wide distributed SSR markers. Gene diversity for the varieties in the subset is 0.544. Among them, *Indica* rice shows a higher gene diversity (0.484) than that of *Japonica* rice (0.454). Similarly, non-waxy rice shows a higher gene diversity (0.540) than that of waxy rice (0.515). However, early-seasonal rice shows a higher gene diversity (0.546) than that of late-seasonal rice (0.510) in our case.

Cultivated rice has been intensively selected during its domestication and breeding. Consequently, the genomic regions controlling traits of economic importance are expected to be shaped by this selection. Therefore, characterizing the genome-wide distribution of genetic diversity of cultivated rice germplasm which has been selected for different traits, such as waxy vs. non waxy might help to identify the genes controlling these traits. To do so, as one example, gene diversity was calculated for the waxy rice as well as non-waxy rice for each marker separately across the genome. Similarly, a measurement for genetic distance, modified Roger's distance (MRD) between waxy and non-waxy rice was calculated on an individual marker basis.

Our results indicated that gene diversity for waxy and non-waxy rice varied across the genome (Fig.3). A different degree of divergence (as measured by MRD) between these two germplasm types was observed across the genome (Fig.4).

The unequal distribution of genetic diversity across the genome could be explained by the selection history of the different genome regions. Therewith, the genome-wide distribution maps of genetic diversity might be a first step to identify the target genes or regions selected during breeding history. For example, genes related to waxy and non-waxy rice might be present in the most divergent genomic regions between these two germplasm types. Common genes under selection in the breeding program of the both germplasm types (e.g. disease resistant genes) might be present in the genomic regions showing the same level of gene diversity and low MRD.

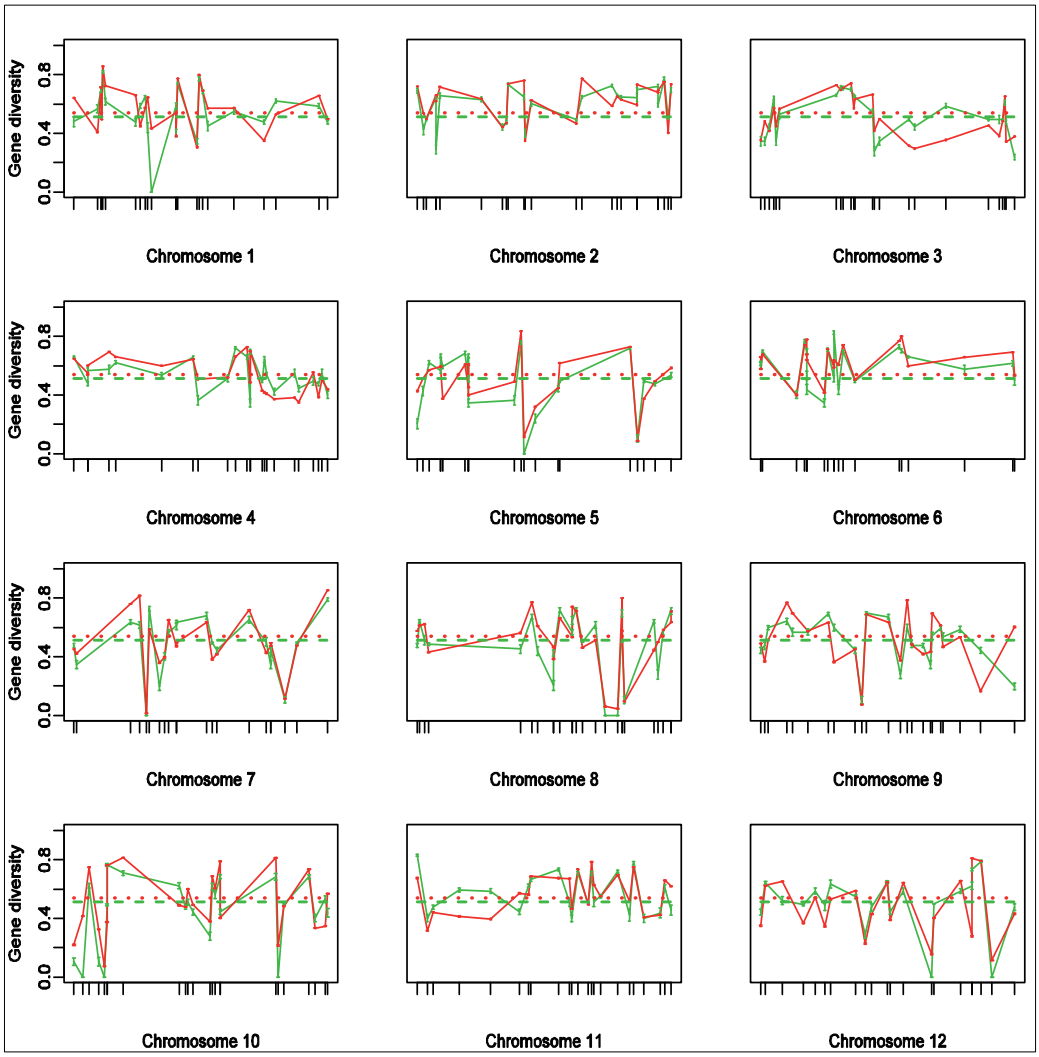


Fig. 3. Green and red lines indicate gene diversity of non-waxy and waxy rice, respectively. Dashed lines indicate the average gene diversity of the corresponding germplasm type. Vertical lines at each point indicate standard error multiplied by 100 which were calculated by bootstrapping across genotypes. Vertical lines at the x axis indicate genetic map positions of the SSR marker on the chromosome.

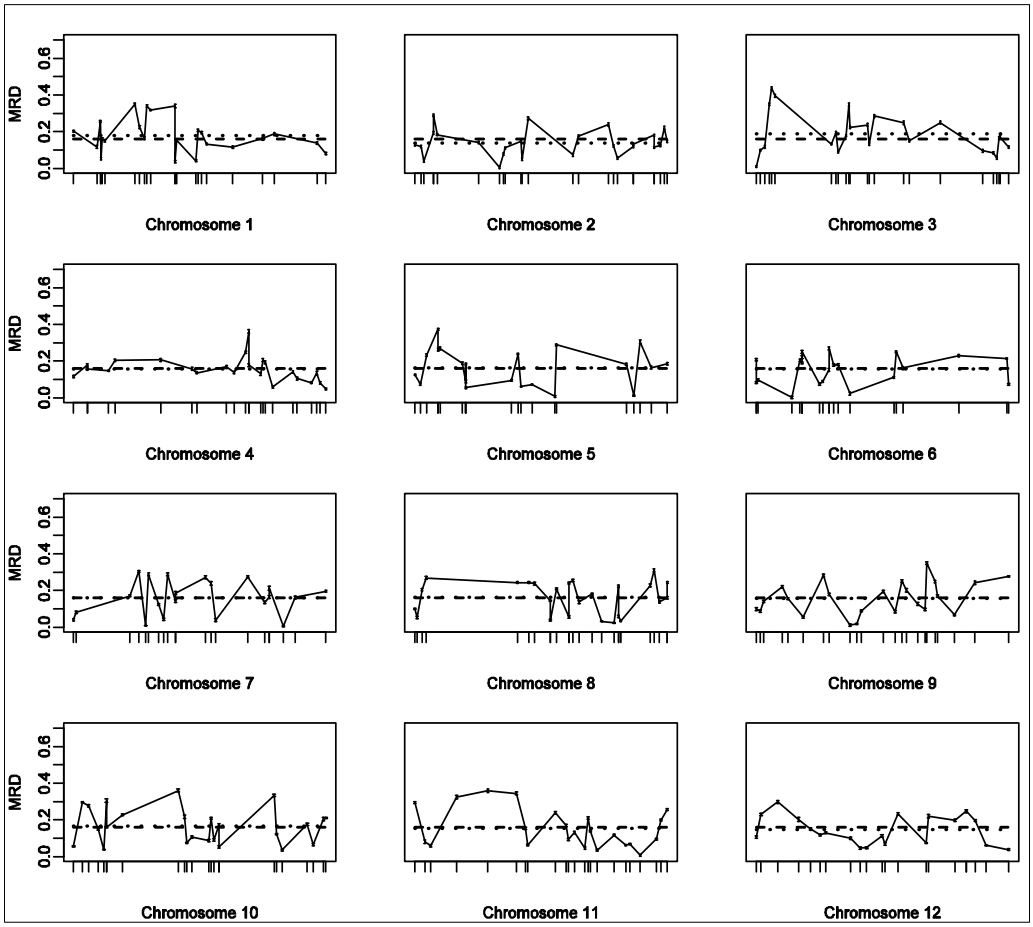


Fig. 4. Modified Roger's distance (MRD) between waxy and non-waxy rice across the genome. Dashed lines indicate average MRD across the genome and dotted lines average MRD for each chromosome. Vertical lines at the x axis indicate genetic map positions of the SSR markers on the chromosome.

5. Constructing a core collection to make use of genetic diversity

A large number of accessions in the germplasm collection (7128 accessions in our case) makes it difficult to choose the most promising ones for utilization. One feasible method is the development of core collections. A core collection is intended to contain, with a minimum repetitiveness, the maximum genetic diversity of a crop species and its wild relatives (Brown, 1989a, 1989b; Frankel and Brown, 1984). The development of a core collection could enhance the utilization of germplasm collections in crop improvement programs and simplify their management. Selection of an appropriate sampling strategy is an important prerequisite to construct a core collection with appropriate size in order to adequately represent the genetic spectrum and maximally capture the genetic diversity in available crop collections. Our studies were aimed to evaluate how sample size, clustering methods, sampling methods, and different data types affected the construction of core

collection and tried to find out an optimal strategy concerning the above factors for core collection construction.

By using three sampling strategies, three kinds of trait data, eight hierarchical clustering methods, and 15 kinds of different sampling proportions were applied to choose the optimal constructing strategies. Analysis of variance (ANOVA) and multiple comparisons were applied to compare different strategies. In order to choose the optimal constructing strategies, 12 evaluated parameters were applied to evaluate the validity of sampling.

The ANOVA analysis showed significant difference for different clustering methods, data types, sample size and sampling methods (Table 2). Furthermore, there were significant interaction effects between these factors except clustering method and sampling method, sampling size and sampling method. The results indicated that these factors as well as their interaction would affect the construction strategy and must be considered carefully.

For different sampling methods, preferred sampling plus multiple clustering and sampling on the degree of variation is better than preferred sampling plus multiple clustering and random sampling, and the completely random sampling is the worst; For the eight clustering methods, clustering analysis with shortest distances has the best of genetic diversity index, average Shano-Weaver index, phenotype retained percentage, and variance of phenotypic frequency; For the three different data types (qualitative trait data, quantitative trait data, intergrated qualitative and quantitative trait data), the core collections constructed by integrated qualitative and quantitative trait data retain the greatest genetic diversity and is the best one. For the sampling rate, the sampling rate of 3.4% ~ 24% is sufficient to retain the greatest genetic diversity of the initial population (Table 4-6).

Finally, a core collection was constructed by using preferred sampling plus multiple clustering and sampling on the degree of variation, clustering analysis with shortest distances, and based on the integrated qualitative and quantitative data. This core collection contains 150 accessions out of 2262 original collection with full recorded data from Ting's collection, accounting for 6.6% of the initial collection.

Origin of variation	df	SS	MS	F value	P value
Clustering method	7	0.07	0.01	29.82**	<.0001
Data type	2	0.49	0.25	720.66**	<.0001
Sampling size	15	0.19	0.01	37.5**	<.0001
Sampling method	1	0.003	0.003	8.78**	0.003
Clustering method × data type	14	0.16	0.01	33.74**	<.0001
Clustering method × sample size	105	0.09	0.0008	2.43**	<.0001
Clustering method × sampling method	7	0.0006	0.0001	0.25	0.97
Data type × sample size	30	0.81	0.03	79.51**	<.0001
Data type × sampling method	2	0.01	0.0038	11.02**	<.0001
Sample size × sampling method	15	0.0035	0.0002	0.68	0.81
Error	615	0.21	0.0003		
Sum of variation	831	3.18			

Table 2. Analyis of variance for Shanno-weaver diversity index of different subsets

Clustering method*	Ratio of phenotype retained	Variance of phenotypic frequency	Average diversity index
1	0.9983A	0.1526C	2.0142A
2	0.9968BC	0.1619B	2.013A
3	0.9973BC	0.1630B	1.9958C
4	0.9977ABC	0.1659A	2.0032B
5	0.9978BA	0.1616B	2.0138A
6	0.9968C	0.1620B	2.0123A
7	0.9970BC	0.1621B	2.0129A
8	0.9972BC	0.1617B	2.0129A
9	0.9512D	0.1636B	1.974D

* Clustering with 1. shortest distance, 2. complete linkage method, 3. median distance, 4.the centroid method, 5. average linkage, 6. the flexible-beta method, 7. the flexible method, 8. Ward's minimum variance method, and 9. completely random method.
** the data with different alphabet show significance at 0.01 level.

Table 3. Multiple comparison for different clustering methods**

Sampling method*	Diversity index For qualitative trait	Average diversity index
1	0.581A	2.005B
2	0.577A	2.013A
3	0.558B	1.974C

* 1. preferred sampling plus multiple clustering and randomly sampling, 2. preferred sampling plus multiple clustering the clustering with sampling on degree of variation, 3. sampling with completely random method
** the data with different alphabet show significance at 0.01 level.

Table 4. Multiple comparison for different sampling methods**

Data type*	Ratio of phenotype retained	Variance of phenotypic frequency	Average diversity index
1	0.996A	0.1509C	2.019A
2	0.9923B	0.18A	1.99B
3	0.9958A	0.1524B	2.017A

*1. integrated qualitative and quantitative trait data, 2. qualitative trait data, and 3. quantitative trait data
** the data with different alphabet show significance at 0.01 level.

Table 5. Multiple comparison for different data types**

6. Association mapping with the rice core collection

Though a large number of exotic genes exist in crop germplasm resources, the rich genetic variations in crop germplasm resources haven't been fully explored and utilized because of being lack of appropriate statistical methods.

In general, conventional method for mining gene from germplasm is linkage mapping. Identifying QTLs by linkage mapping needs to construct one or several segregating

Sample size	Ratio of phenotype retained	Variance of phenotypic frequency	Average diversity index
Original collection	1A	0.1581F	1.994I
1450	1A	0.1581F	1.993I
1350	0.9995A	0.1589EF	1.996HI
1250	0.9995A	0.1593EFD	1.999HG
1150	0.9995A	0.1604EFD	2GF
1050	0.9995A	0.161EFDC	2.003EF
950	0.999BA	0.1624BDC	2.005EDF
850	0.998B	0.1638BAC	2.006ED
750	0.998B	0.1649BA	2.006ED
650	0.998B	0.1651BA	2.008D
550	0.997C	0.1655BA	2.008D
450	0.996C	0.1648BA	2.012C
350	0.993D	0.1655BA	2.018B
250	0.99E	0.1659A	2.02A
150	0.983F	0.1615EDC	2.02A
50	0.996G	0.149G	2.02A

** the data with different alphabet show significance at 0.01 level.

Table 6. Multiple comparison for different sample size**

populations by crossing between parents (e.g., F_2, Double haploid, Backcross population) and linkage mapping would be done in these segregation populations. The accuracy of QTL mapping is dependent largely on selected but limited parents and only two or a few alleles from the parents were detected. Moreover, abundant genetic variation stored in germplasm have not been developed and utilized due to lack of appropriate statistical methods. Provided using conventional QTL linkage mapping method for mining the abundant genetic variations in a large germplasm resources population, it is necessary to make diallel crossing with all studied accessions, which is hard to develop such mapping population and would take much more time, cost, space and analysis.

An alternative, association mapping based on linkage disequilibrium (LD) analysis might be an effective way to identify the function of the gene (or targeted high-resolution QTL), which has been successfully applied in human genetics to detect QTL coding for simple as well as complex diseases (Corder et al., 1994; Kerem et al., 1989). This method uses the LD between DNA polymorphisms and genes underlying traits. LD refers to the non-random combination among different genetic markers. The main mechanism for LD existing in a population is genetic linkage among different loci. Therefore, it is possible to detect QTLs by identifying LD between markers loci and potential QTLs. Through detecting abundant genetic markers loci locating in genome or those nearby candidate genes, the loci which link tightly with QTLs and show correlated to QTLs can be found. The application of association

mapping to plant breeding appears to be a promising approach to overcome the limitations of conventional linkage mapping (Kraakman et al., 2004).

Furthermore, choice of an appropriate germplasm to maximize the genetic diversity and the number of historical recombinations and mutation events (and thus reduce LD) within and around the gene of interest is critical for the success of association analysis (Yan et al., 2011). As described above, core collections are the core subset of the original collections with minimum samples while having the maximum genetic variability contained within the gene pool. Therefore, association mapping with a core collection population helps to catch as more phenotypic variation as possible and would make use of both the advantages of association mapping and core collection, thus could be an effective way to mine and utilize the abundant genetic diversity in the crop germplasm resources.

6.1 Population structure and LD pattern

Population structure is an important component in association mapping analysis because it can reduce both type I and II errors between molecular markers and traits of interest in an inbreeding specie. Moreover, low level of LD could lead to impractical whole-genome scanning because of the excessive number of markers required. Furthermore, the resolution of association studies in a test sample depends on the structure of LD across the genome. Therefore, information about the population structure and extent of LD within the population is of fundamental importance for association mapping.

The rice core collection consisting of 150 varieties were genotyped with 274 SSR markers. Based on the genotyping data, STRUCTURE software was run to detect the number of subgroups within the core collection population and assign the varieties into different subgroups with the membership probability of 0.80 as a threshold. To compare and confirm the STRUCTURE subgroups, a additional principal component analysis was done.

STRUCTURE indicated that the entire population could be divided into two subgroups (i.e. SG 1 and SG 2) (Fig. 5). With the membership probabilities of ≧ 0.80, 111 varieties were assigned to SG 1, 21 varieties were assigned to SG 2 and 18 varieties were retained to the admixed group (AD) (Fig. 5). Principal component analysis confirmed the population structure, i.e. the varieties from SG 1 and SG 2 located in two distinct clusters, and those from AD located between the two subgroups (Fig. 5). The varieties in SG 1 are mainly *indica* rice, and those in SG 2 *japonica* rice, whereas those in AD intermediate (*indica*- or *japonica*-inclined) rice. Furthermore, the varieties from the same cultivated zone were clustered closely.

LD measured as squared correlation of allele frequencies (r^2) between loci pairs in the core collection and different germplasm types were calculated (Table 7). The average r^2 between linked loci (the loci at the same chromosome) varied between 0.0188 and 0.1. Using the 95% quantile of r^2 between unlinked loci pairs as a threshold, 6.23% linked loci pairs were in significant LD. For different germplasm types (*indica*, *japonica*, early-seasonal, late-seasonal, waxy, non-waxy rice), the percentage of loci pairs in LD varied between 5.33 and 6.36%. LD (r^2) against genetic map distance (cM) between linked loci pairs was plotted and a nonlinear regression of r^2 vs. genetic map distance according to Heuertz et al., (2006) was performed (Fig. 6). The LD decays against the genetic distance, which indicated the linkage might be the main reason for the causes of LD. The LD decays to the threshold, i.e. the 95% quantile of r^2 between unlinked loci pairs, at 1.03 cM in the entire collection (Table 7). The cut-off decay

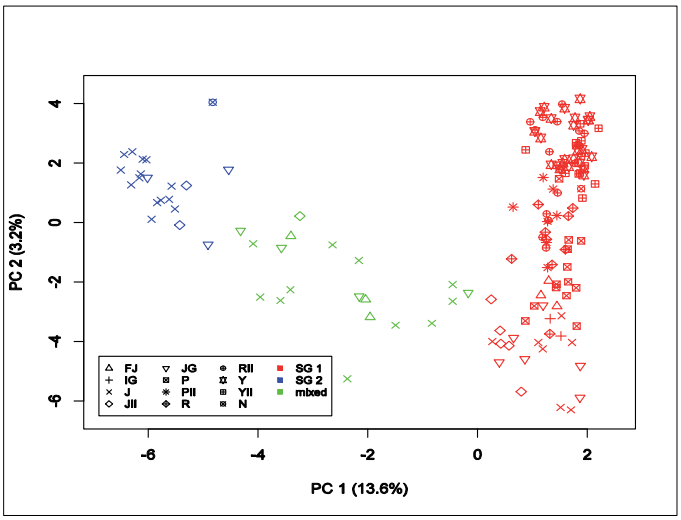


Fig. 5. Principal component analysis for the rice core collection combined with STRUCTURE subgroup assignment. PC 1 and PC 2 refer to the first and second principal components, respectively. The numbers in parentheses refer to the proportion of variance explained by the principal components. Symbols indicate different type of rice, and colors indicate different subgroups from STRUCTURE software. FJ-Foreign *japonica*, IG-glutinous *Indica*, J-early seasonal *Japonica*, J II- late seasonal *Japonica*, JG-glutinous *Japonica*, P-early seasonal *Indica* from Pearl river region or south China, P II-late seasonal *Indica* from Pearl river region, R-early seasonal Red grain rice, R II-late seasonal Red grain rice, Y- early seasonal *Indica* from Yangtze River region, Y II- late seasonal *Indica* from Yangtze River region, and N-Unknown origin.

Type	Sample size	Average R^2	Percentage of loci pairs in LD	Cut-off (cM
Entire collecti	150	0.0330	6.23	1.03
Indica	90	0.0188	5.33	0.89
Japonica	60	0.0570	6.37	1.10
Early-seasonal	86	0.0500	6.22	1.00
Late-seasonal	64	0.0330	5.95	1.04
Waxy	18	0.1000	6.06	0.87
non-waxy	132	0.0350	6.36	1.01

Table 7. Linkage disequilibrium (measured as R^2 value) for linked loci, percentage of linked loci pairs in LD, and the cut-off decay distance for the core collection and different germplasm types.

distances for *indica*, *japonica*, early-seasonal, late-seasonal, waxy, non-waxy rice were 0.89, 1.10, 1.00, 1.04, 0.87, and 1.01cM, which were about 200-500kb physical distance. The results indicated that choice of the core collection could maximize the number of historical recombinations and mutation events and thus reduce LD within and around the gene of interest which is critical for the success of association analysis (Yan et al., 2011). Such short

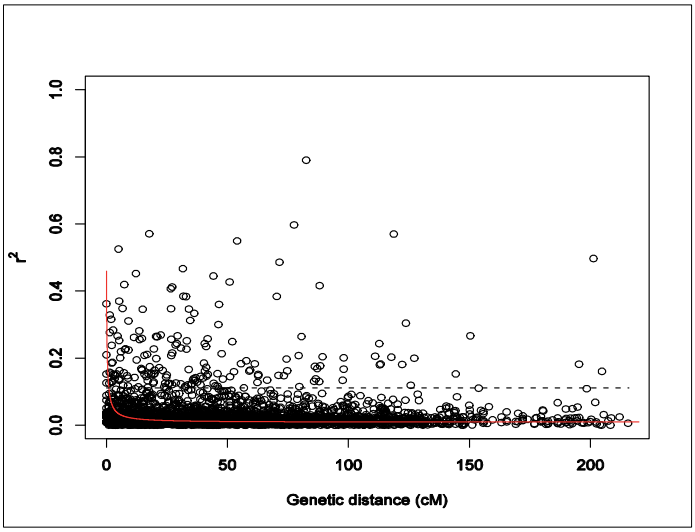


Fig. 6. Plot of linkage disequilibrium measured as squared correlation of allele frequencies (r^2) against genetic map distance (cM) between linked loci pairs in the core collection. The red line is the nonlinear regression trend line of r^2 vs. genetic map distance. The dashed line indicates the 95% quantile of r^2 between unlinked loci pairs.

LD decay distance suggested that fine mapping with a core collection for desirable genes could be possible. However, due to low percentage of linked loci pairs in LD and the quick decay of LD, in turn it indicated that the density of markers for genome-wide association mapping should be greatly increased as compared to our study. Considered for the LD decay distance in the core collection and 1700cM map distance of rice genome, it might at least in theoretically require more than 1700 markers for a genome-wide association mapping with such a core collection. If higher power is needed, the number of required markers could be even more.

6.2 Association mapping

Mining the elite genes within rice germplasm is of importance to the improvement of cultivated rice. Therefore, genome-wide association mapping was applied with the rice core collection using 274 SSR markers.

All of the 150 rice varieties were cultivated at the farm of South China Agricultural University, Guangzhou (23°16N, 113°8E), during the late season (July-November) for two consecutive years (2008 and 2009). The yield related traits (such as grain weight, filled grains, tillers per plants, etc.) were measured for both years. As for an example, the trait yield per panicle (gram) was furthered used for association mapping.

The software STRUCTURE was applied to infer historical lineages that show clusters of similar genotypes and get the Q matrices (Pritchard et al., 2000). Kinship matrix (K) was calculated by software SPAGeDi (Hardy and Vekemans 2002). The quantile–quantile plots of estimated -log10(p) were displayed using the observed p values from marker-trait associations and the expected p values assuming that no associations happened between markers and any trait in the software SAS. Using the TASSEL software and the mixed linear regression model (MLM), association test was performed for the yield trait, incorporating K and Q matrices.

The QQ plot of observed vs. expected p values (Fig. 7) indicated that MLM model incorporating K and Q matrix was suitable for the association analysis for the yield trait. A total of 17 markers in 2008 and 15 markers in 2009 were detected to significantly ($P < 0.05$) be associated with the yield traits (Table 8 and 9). 12 marker-phenotype associations were confirmed by previous researches (either using linkage mapping or association mapping) for 2008 year's results. And it was 7 for the 2009 year's results. Moreover, two marker-phenotype associations were located in the similar position (RM471 with RM218, PSM188 with RM235) for both years. The genetic variants explained by the markers varied between 3.49% and 24.86% in 2008, while it was between 3.02% and 13.87% in 2009. The genetic variants explained by the marker RM346 and PSM336 were more than 15%. It is worth to note that less common marker-phenotype associations were detected in both years, which indicated the yield trait might be easily influenced by the environment. Such problem might be overcome by using the yield data for multiple locations and years. The marker-phenotype associations could be further used in rice breeding by marker-assisted selection.

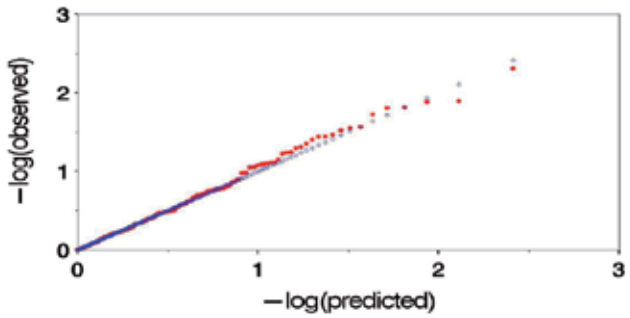


Fig. 7. Plots of observed vs. expected p values for MLM (Q+K) model for the yield trait.

No.	Markers	Chromosome	Map distance (cM)	P value	R^2 (%)
1	RM71	2	49.8	0.0377¥	3.49
2	PSM374	2	83.6	0.0448¥	5.05
3	RM293	3	193.4	0.0268	5.90
4	RM471	4	53.8	0.0202¥	6.39
5	RM559	4	155.8	0.0350	7.02
6	RM469	6	2.2	0.0150¥	10.33
7	RM549	6	42.7	0.0270	7.51
8	RM170	6	2.2	0.0346¥	7.05
9	RM432	7	43.5	0.0021¥§	12.26
10	RM429	7	96.9	0.0119¥	5.16
11	PSM142	7	26	0.0385¥	13.45
12	PSM336	7	80.5	0.0002¥§	16.43
13	RM346	7	47	0.0006¥§	24.86
14	RM201	9	81.2	0.0011§	21.34
15	RM167	11	20.3	0.0133	7.07
16	PSM366	11	89	0.0018¥§	14.44
17	PSM188	12	86.5	0.0059¥	10.34

§The Bonferroni threshold (< 0.0036); ¥supported by previous researches; R^2 represents the genetic variants explained by the marker.

Table 8. Association mapping results for yield per panicle in 2008 using MLM models.

No.	Markers	Chromosome	Map distance (cM)	P value	R^2 (%)
1	RM220	1	28.4	0.0392¥	3.02
2	PSM369	1	170.4	0.0130¥	4.41
3	RM450	2	150.8	0.0270¥	5.17
4	RM218	3	67.8	0.0491	5.62
5	RM153	5	3	0.0341	4.83
6	RM334	5	141.8	0.0188	8.74
7	RM340	6	133.5	0.0152	13.87
8	RM182	7	61	0.0155¥	5.99
9	RM10	7	63.5	0.0358¥	3.14
10	RM257	9	66.1	0.0049	5.71
11	RM242	9	73.3	0.0360	3.13
12	RM304	10	73	0.0126	4.45
13	RM228	10	96.3	0.0282	3.43
14	RM309	12	74.5	0.0299¥	3.36
15	RM235	12	91.3	0.0442¥	13.00

§The Bonferroni threshold (< 0.0036); ¥supported by previous researches; R^2 represents the genetic variants explained by the marker.

Table 9. Association mapping results for yield per panicle in 2009 using MLM models.

7. Discussion and prospect

Rice feeds more than 50% of the world's population and is one of the most important crops in the world. Rice germplasm resource is the primary material for rice breeding and makes a concrete contribution to global wealth creation and food security. Therefore, understanding its valuable genetic diversity and using it in rice genetic improvement is of importance for raising rice yield and the resistance to biotic and abiotic stress as well as improving rice quality to secure global food supplies.

To mine the wide genetic diversity in plant germplasm populations, identification of phenotypic traits might the first and an important step. Besides the agronomic traits, physiological traits, stress-related traits, quality traits, resistant to virus and pets traits, etc. should be furthered studied in details. Based on the full evaluation of phenotypes, a dynamic core collection could be constructed either on a specific target trait or on all the traits so that the core collection could retain as much as genetic diversity with the minimum accessions. The core collection could be furthered studied with high density markers as well as exact measurement of its phenotype.

Association mapping has become a promising approach to mine the elite genes within germplasm population compared to traditional linkage mapping. Association mapping based on a core collection would help to catch as more phenotypic variation as possible. Compared to a natural population or a breeding population with narrow genetic basis, the LD level in a core collection might be low due to its diverse origin. Therefore, more markers might be needed for genome-wide association mapping. However, due to the quick LD decay, fine mapping might be possible with a core collection. To perform a precisely association mapping, multiple replications either locations or years for the phenotypic

identification, exact measurement of the population structure and the kinship should be considered. Furthermore, though a rapid progress has been made in genotyping, a quick, automated, economic genotyping technology (such as SNP array) for a large number of germplasm resources are desirable for association mapping with germplasm resources population. How to effectively combine the linkage and association mapping in plants (such as the nested association mapping in maize, NAM) might be another question which should be concerned. Due to such associations could be further applied in rice breeding with molecular assisted selection, it provide a bright future to make use of the elite genes in the diverse germplasm resources.

A strategy is proposed for exploring and utilization of the wide genetic diversity in plant germplasm populations, i.e. firstly, evaluation of the genetic diversity for germplasm populations at phenotypic and genotypic level; secondly, constructing core collection to achieve the maximum diversity with minimum accessions; thirdly, combining linkage mapping and association mapping to map desire QTL in a large scale and with high resolution; fourthly, developing near isogenic lines to verify and fine map QTLs; finally, cloning desirable genes and make use of them in cultivated plant breeding.

8. Acknowledgment

The authors thank very much to Academician of Chinese academy of science, professor Lu, Yong Gen for his supervision and support for the research. Many thanks are given to Dr. Li, Xiao Ling, Dr. Feng, Jiu Huan, Ms. Zhao, Xing Juan, Mr. Zeng Kai Long, Mr. Deng, Yong Hong, Dr. Xu, Hai Ming for their helps in the experiment and data analysis. The research was supported by Fund of the National Natural Science Foundation of China grant (30700494).

9. References

Brown A H D. 1989a. Core collections: A practical approach to genetic resources management. Genome, 31: 818–824.

Brown A H D. 1989b. The case for core collections. In: Brown A H D, Marshall D R, Frankel O H, Williams J T. The Use of Plant Genetic Resources. Cambridge: Cambridge University Press: 136–156.

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, Haines JL, Pericak- Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 7:180–184

Frankel O H, Brown A H D. 1984. Plant genetic resources today: A critical appraisal. In: Holden J H W, Williams J T. Crop Genetic Resources: Conservation and Evaluation. London, UK: George Allen and Unwin Ltd: 249–257.

Goff SA, Ricke D, Lan TH, et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296: 92–100.

Hardy, O. J., Vekemans, X. (2002) SPAGeDi: a versatile computer program to analysis spatial genetic structure at the individual or population levels. Molecular ecology notes, 2: 618-620

Heuertz M, Emanuele D P, K¨allman T, Larsson H, Jurman I, Morgante M, Lascoux M, & Gyllenstrand N (2006), Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce (*Picea abies* (L.) Karst). Genetics 174:2095 – 105.

Kerem B, Rommens JM, Buchanan JA, Markievicz D, Cox DK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080

Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Lu, Y. G., Li, X. L., Li, J. Q. (2006). Construction of a rice core collection. In: Seminar for the developing strategy of Chinese super rice and 10th annual memory. Shenyang Agricultural University, 124-133

Pritchard JK, S. M., DONNELLY P 2000a Inference of population structure using multilocus genotype data. Genetics: 945-959.

Tanksly, S.B., Mc Couch, S. R. (1997) Seed bank and molecular maps: unlocking genetic potential from the wild. Science, 277(22): 1063-1066

Yan, J. B., M. Warburton and J. Crouch, 2011 Association Mapping for Enhancing Maize (*Zea mays* L.) Genetic Improvement. Crop Science 51: 433-449.

Yu J, Hu S, Wang J, et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79–92.

6

Genetic Diversity and Utilization of Triploid Loquats (*E. japonica* Lindl)

Qiao He, Weixing Wang, Qigao Guo, Suqiong Xiang,
Xiaolin Li and Guolu Liang*
Southwest University,
China

1. Introduction

Eriobotrya is a genus of at least 22 species and 10 varieties or *forma* of evergreen fruit trees in the family of Rosaceae (Lin, 2007). *Eriobotrya* is native to east and southeast of Asia, of which only loquat (*Eriobotrya japonica* Lindl.) is cultivated for its valuable fruit. The hypothesized center of origin and center of diversity of loquat are located in the medium to lower Daduhe River and the southeast slope of Gongga Mountain in Southwestern China (Qiu and Zhang, 1996). Loquat was domesticated in China at least 2000 years ago and has been widely cultivated for fresh and processed fruit, as well as for its medicinal effect (Qiu and Zhang, 1996). Loquat was introduced into Japan, France, England, United States and various Mediterranean countries between 12th to 19th century. Today loquat is mainly distributed between latitudes 20 and 35° north or south from the equator, but it can be cultivated up to latitude 45° under marine climates (Lin et al., 1999). There are more than 30 loquat producing countries in the world and the production is distributed in Asia, Europe, Africa, Australia, and the America. In addition to the utilization of its fruit, loquat flower is a superior honey source and it mostly blooms in fall and early winter. The white flower is aromatic thus is appreciate as ornamental tree as well.

The global planting area was about 130,000 hectares, with the production over 549,220 tons in 2005. The main producing countries are China, Spain, India, Japan and Pakistan, which account for 97% of the planting area and 94% of the output respectively. In China, the planting area of loquat is more than 120,000 hectares with an output of more than 400,000 tons. The production in China is distributed in 20 provinces and the leading producers include Sichuan, Fujian, Chongqing, Zhejiang, Hunan, Guangdong and Guizhou. Many well-known loquat varieties originated from these provinces and are widely planted in China among which include 'Dawuxing', 'Longquan No.1', 'Zaozhong No.6'(Lin 2007a).

Genetic diversity and the relationships among different varieties of loquat are of great importance for the conservation of genetic resources, breeding, national and international exchange of germplasm (He et al., 2011). Research on genetic diversity of loquat based on pomological traits and molecular markers have been widely carried out (Badenes et al.,

* G. L. Liang – Corresponding author
College of Horticulture and Landscape Architecture, Key Laboratory of Horticulture Science for Southern Mountainous Regions, Ministry of Education, Southwest University, Chongqing,China

2000; Cai 2000; Vilanova et al. 2001; Soriano et al. 2005; Dong 2008; Qiao 2008; Watanabe et al. 2008; Gisbert et al. 2009; Yang 2009). These studies significantly enhanced our understanding about the distribution and structure of genetic diversity in loquat germplasm around the world.

2. Germplasm and breeding research of loquat

Loquat is mainly self-compatible, self-incompatibility has only been found in a few varieties (Chen & Chu, 2008). The traditional propagation from seed has provided a range of varieties adaptable to different environments and planting regions. *Ex situ* germplasm collections have been established in China, Japan and Spain (Zheng, 2007; Badenes et al., 2009). Among these collections, the Chinese collections possess highest diversity. There are more than 1000 accessions described in the various Chinese germplasm collections (Zheng, 2007). The largest Chinese collection is located in Fuzhou, with a total of more than 250 accessions. The major European loquat germplasm collection is held at Instituto Valenciano de Investigaciones Agrarias (IVIA) of Spain. The IVIA collection includes more than 74 accessions, of which 49 accessions are Spanish varieties. In addition, Italy and Greece also have small collections of mainly local germplasm (Lin, 2003).

Understanding of genetic diversity in loquat germplasm has been greatly improved through the application of molecular marker technology (He et al., 2011). Using internal transcribed spacer (ITS) region of the 18S–5.8S–26S nuclear ribosomal cistron, Li et al. (2009) analyzed 15 *Eriobotrya* accessions and six close related species. Result of cluster analysis suggested that *E. malipoensis* had closer relationship with *E. japonica* Lindl. This result was further supported by Yang et al. (2009) who analyzed *Eriobotrya* taxa using AFLP markers. However, Zhao et al. (2011) also used ITS region of the 18S–5.8S–26S nuclear ribosomal cistron and assessed the phynogenetic relationship among the *Eriobotrya* species represented by 25 loquat cultivars and seven wild taxa. Their result suggested that the evolution order of the studied taxa was *E. bengalensis* f. *angustifolia*, *E. prinoides var. dadunensis*, *E. prinoides* (*E. bengalensis*), *E. dayaoshanensis* and *E. japonica* Lindl. And *E. bengalensis* f. *angustifolia* was found the likely ancestral taxa of *E. japonica* Lindl.

The intraspecific genetic diversity in loquat has been analyzed by various molecular markers as well. Soriano et al. (2005) assessed genetic relationships among 40 loquat accessions that originated from different countries, including accessions from the European loquat germplasm collection maintained at IVIA in Valencia, Spain. Thirty pairs of microsatellites previously identified in *Malus* × *domestica* (Borkh.) were used. The expected and observed heterozygosities were 0.46 and 0.51 on average, respectively, showing a high level of out crossing behavior in loquat germplasm. However their result also indicates a low degree of genetic diversity in the set of loquat accessions analyzed, in comparison to other members of the Rosaceae family.

He et al. (2011) used apple SSRs that are transferable to loquat and assessed the level of genetic diversity within loquat. They used 39 SSRs transferable to loquat and genotyped 54 loquat accessions from Japan, Spain, and four Chinese provinces, as well as two wild species (*E. prinoides* var. *Daduheensis* and *E. prinoides*). In total they identified 155 different alleles with a mean value of 3.38 per locus, and the mean observed heterozygosity was 0.47. These values indicate a high degree of genetic diversity in the set of Chinese loquat accessions analyzed. Cluster analysis grouped the accessions into cultivated and wild loquats. The

cultivated loquat can be subdivided into three subgroups which generally reflect their geographic origin in China.

The Chinese loquat germplasm is traditionally classified according to the flesh color. Varieties can be classified as either "white flesh" or "orange flesh". The former accounts for about 30 percent of the total Chinese varieties. The texture of the white-fleshed loquat is generally more delicate and tender. Varieties in this group include 'Ruantiaobaisha' and 'Baiyu', both are the leading varieties of Zhejiang and Jiangsu province in China. Several Japanese varieties, such as 'ShiroMogi', canbe placed in the white flesh group. Orange-fleshed varieties, such as 'Mogi', 'Tanaka', and 'Nakasakiwase' account for 95% of the total crop area of Japan. Spanish commercial production depends on four orange flesh varieties, including 'Algerie', 'Magdal', 'Golden Nugget', and 'Tanaka' with 'Algerie' varieties (Badenes et al., 2009).

Crop improvement, both through breeding for new varieties and selection of accessions from existing germplasm, have been carried out in China, Japan and Spain (Zheng, 2007; Shih, 2007; Badenes et al., 2009). However, so far the varieties used in production were mostly farmer selections made by growers in the local areas. Surveys and selection of germplasm accessions have been carried out in China (Zheng, 2007), Mediterranean countries, Turkey and Pakistan (Badenes et al., 2009). Seedless or fewer seeds is one of important objectives in today's fruit breeding programs. Loquat fruit has a relatively smaller fruit than other fruit tree species. The average weight of a loquat fruit is about 30 to 40 grams. Some large-fruited varieties can have a fruit size of 70 grams, and 34 usually with 3 to 4 large seeds. The low edible rate (less than 70%, is an important quality constraint for fresh consumption of loquat (Lin et al, 2007b). Therefore, seedless loquat fruit is highly desired by consumers.

The classical approach to eliminate seeds in many crops is to produce triploids through ploidy manipulation (Ollitrault et al., 2008). And there are three routes to obtain triploidy genotypes, including use of nonreduced megaspore or microspore, crossing induced tetraploids with diploids and in vitro culture of nuclear tissue (Janick, 2011). Tetraploid varities can be spontaneously formed or induced by colchicines (Yahata et al., 2004). In 1978, Chinese researchers of Fruit Research Institute of Fujian Academy of Agricultural Science developed a tetraploid varities 'Min No.3' using colchicine-induced polyploidization. However the tetraploid variety was not well adopted due to the performance in other agronomic traits and quality. Between 1983 and 1985, researchers in the same institute obtained triploid loquat plants through endosperm culture. In early 1990s, the Southern Prefectural Horticulture Institure, Chiba Prefectural Agriculture and Forestry Research Center of Japan hybridized diploid variety "Nagasakiwase" 6 with tetraploidy variety "Tanaka", and developed the first triploid loquat variety "Kibou", as well as a series of triploidy seedlings. From 1997 to present, Liang and his team in the Southwest University of China have made significant progress in identification of natural triploid loquats (Guo et al., 2007).

3. Occurence of natural triploid loquats

Exploitation of triploid plant to induce seedlessness is a promising breeding technique since triploid is a naturally existing ploidy status as found in other Rosaceae species. For example, approximately 10% of apple and pear varieties are triploids although the frequency of natural occurrence is less than one percent. In 1993, Liang et al. discovered that substantial

frequency of natural triploid individuals exists in loquat germplasm. Since then, massive screening have been conducted in Chinese loquat germplasm, which has led to the selection of 352 natural triploid individuals out of 99,542 seeds in 36 varieties, the frequency of occurrence is about 0.35% (Table 1).

Crop	No. 2n × 2n seedlings	Distribution of seedlings, percentage and number			
		Diploid	Triploid	Tetraploid	Pentaploid
Apple	6,825	99.63	0.28 (19)	0.09 (6)	
Loquat	99,542	99.57	0.35 (352)	0.07 (74)	0.01 (10)

Table 1. Comparison between apple and loquat for the frequency of ploidy levels in natural crosses.

4. Morphological characters of triploid loquats

As compared to diploidy loquat, the triploids have stronger growth vigor, characterized by the thick trunk and branches as well as larger leaves. Moreover, phenotypic differences are also found in flowers, triploid plants usually have larger flowers and flower buds than diploid ones. The transverse and length diameters of flowers, flower buds, anthers and ovaries of triploidy loquat were significantly larger than those of diploids. And fruits of triploidy loquat are significantly larger and seedless, with an edible rate of more than 80% (Table2-4, Figure 1). However, the morphological characters of different seedlings have a wide range of variation. For example, the triploid seedlings of Changbai No.1, Q21 has a ovoid-shaped fruits, whereas plant Q27 and Q11 have long ovoid fruits and the fruit weight of Q27 was significantly heavier than that of Q11 and Q21(Table 4).

5. GISH (Genomic *in situ* hybridization) analysis of triploid loquats

GISH is an efficient and accurate technique for the determination of levels and incorporation positions of alien chromatin. This technique has been widely applied to numerous interspecific and intergenomic plant hybrids (Snowdon et al., 1997). The GISH analysis of

Cultivar	Trunk circum. (cm)	No. branches	Annual branch diam (cm)	Leaf		
				Length (cm)	Width (cm)	Leaf index
Dawuxing (2x)	25.5b[z]	7.2a	5.3b	27.5b	7.3b	3.8a
3x seedling	50.0a	3.6b	7.1a	45.8a	15.6a	2.9b
Longquan No.1 (2x)	29.7b	6.0a	6.5b	22.2b	7.0b	3.2b
3x seedling	50.0a	2.4b	8.5a	36.3a	13.4a	2.7a
Jinfeng (2x)	24.9b	7.0a	6.4	25.6b	7.2b	3.5a
3x seedling	45.0a	3.0b	7.5	43.2a	15.6a	2.8b
Zaohong No.3 (2x)	36.0b	7.0a	5.8b	25.7b	7.7b	3.4a
3x seedling	49.0a	4.0b	7.4a	38.7a	14.1a	2.8b

Table 2. Comparison of plant morphology of diploid loquats and their related triploid seedlings. z Mean separation of 2x and related 3x means at 5% level (Liang et al., 2011).

Varieties	Flower width (cm)	Flower length (cm)	No. single flowers	Bud width (mm)	Bud length (mm)	Filament length (mm)
Dawuxing2x	7.34b	4.46b	72.8b	4.50b	4.30b	2.35b
3x seedling	20.20a	14.60a	104.8a	5.85a	6.90a	5.05a
Longquan No.1 (2x)	8.24b	7.18a	68.8b	4.70b	4.90b	2.90b
3x seedling	13.0a	15.0b	82.0a	6.15a	6.70a	3.85a
Jinfeng (2x)	10.00b	9.60b	51.40b	4.77b	4.77b	2.92b
3x seedling	21.80a	18.8a	98.40a	6.85a	7.00a	3.93a
Zaohong No.3 (2x)	11.60b	8.62b	57.6b	4.60b	5.40b	3.00b
3x seedling	20.20a	12.00a	116.2a	7.60a	7.05a	3.75a
	Anther length (mm)	Anther width (mm)	Style length (mm)	Ovary width (mm)	Ovary length (mm)	
Dawuxing2x	1.65b	1.20b	2.50b	1.84b	1.29b	
3x seedling	2.25a	1.55a	4.05a	3.51a	2.91a	
Longquan No.1 (2x)	1.85b	1.30b	3.05b	1.97b	1.35b	
3x seedling	2.60a	1.95a	3.95a	3.56a	2.95a	
Jinfeng (2x)	1.91b	1.93b	2.54b	1.95b	1.35b	
3x seedling	2.57a	2.56a	4.11a	3.62a	2.94a	
Zaohong No.3 (2x)	1.85b	1.70b	2.73b	2.30b	1.44b	
3x seedling	2.56a	2.00a	4.04a	3.60a	2.91a	

Table 3. The comparison of flowers of natural triploid and diploid loquats. The lower cases in the table means of the significant level (P<5%) of multiple range test.

Varieties	Shape[1]	Flesh color[2]	No. seeds	Weight (g)	Length (mm)	Width (mm)	Shape index	Edible portion (%)
Raotiaobaisha2x	R	W	4-6	25.2 (34)[3]	37.8	38.4	0.97	62.0
3x seedling(H324)	LO	OY	0	50.3 (68)	56.2	40.1	1.40	86.0
Jinfeng2x	O	OY	4-6	61.1 (133)	58.2	46.9	1.24	65.0
3x seedling(D425)	LO	OY	0	79.3 (103)	78.5	52.7	1.49	84.9
3x seedling(D327)	LO	OY	0	78.1 (85)	73.8	41.3	1.78	85.6
Dawuxing2x	O	OY	4-6	58.7 (96)	62.5	45.1	0.97	65.0
3x seedling(A322)	LO	OY	0	65.8 (85)	73.0	50.0	1.56	83.5
3x seedling(A313)	LO	OY	0	62.2 (83)	70.0	49.0	1.49	82.5
3x seedling(A35)	LO	OY	0	63.1 (85)	73.0	49.0	1.78	85.2
Changbai No.1 2x	R	W	4-6	36.2(39)	37.0	42.0	0.88	59.6
3x seedling(Q21)	O	W	0	40.0(50.2)	45.0	41.0	1.10	84.9
3x seedling(Q27)	LO	W	0	65.0(78.3)	61.0	46.0	1.33	85.2
3x seedling(Q11)	LO	W	0	41.6(62.5)	61.0	38.5	1.58	84.5

Table 4. Characteristics of fruit in diploid loquats and their related triploids.
[1]LO= long ovoid, O= ovoid, R=roundish; [2]OY= orange yellow, W=white;
[3]Maximum fruit weight (Liang et al., 2011).

Fig. 1. Compariation of morphological characteristics in leaf, flower and fruit between triploid seedling and diploid (left: 3x, right: 2x)

the natural triploid loquat revealed three types of hybrid (Table 5). In the first type, hybrid signals were detected throughout all 51 chromosomes. In the second type, only 34 chromosomes were detected with hybrid signals. In the third one, only areas around the centromere of all the 51 chromosomes showed hybrid signals. These results showed that these triploids were either homogenous or heterogenous triploids. The different types of hybrid signals also indicated the genetic diversity of natural triploid loquats. Wang (2008) used GISH analysis on artificial triploid loquats, the results showed that the source of hybrid somatic chromosomes can be accurately distinguished using genome DNA of one parent as probe, revealing 17 chromosomes from male parent, and 34 from maternal parent. No significant variation in chromosome structure, such as interchange and inversion was found.

Type of GISH	Individuals	Genome composing
I	Dawuxing (A322, A376), Longquan No.1 (B4-331, B316, B349, B352, B378), Jinfeng (D425), Zaohong No.3 (E39), Ganlu No.1 (I315), Huangrou (G320), Luzhou No.6 (C321)	+++
II	Dawuxing (A368, A379), Longquan No.1 (B347, B372, B374, B441)	++−
III	Dawuxing (A332), Longquan No.1 (B38, B329, B333, B338, B339, B345, B350, B351, B356, B375), Zaohong No.3 (E310), Ruantiaobaisha (H324), Longquan No.5 (K381, K459), Longquan No.6 (J367), Donghuzao, Jianyangtezao, Xiangzhong No.11, Zaozhong No.6	+++

Table 5. The results of GISH and composing of genomes of different varieties of natural triploid loquats. "+, -" means of different genomes. (Liang, 2006)

6. Molecular marker analysis of natural triploid loquats

Molecular marker is an effective technique to assess genetic diversity both dominant and co-dominant molecular markers such as ISSR (inter-simple sequence repeat), AFLP (amplified fragment length polymorphism), MSAP (methylation sensitive amplified polymorphism) and SSR (simple sequence repeat) have been used in genetic diversity analysis of loquat germplasm, including natural triploid loquats.

6.1 ISSR (Inter-simple sequence repeat) analysis of natural triploid loquats

Liang (2006) used ISSR markers to analyze genetic diversity of loquat. Result based on twelve ISSR primers showed that similarity of 'Dawuxing', 'Longquan No.1', 'Longquan No.5', 'Longquan No.6' ranged from 0.65 to 0.98, 0.64 to 0.95, 0.76 to 0.96 and 0.83 to 0.93 respectively. As showed by the amplification pattern of primer 835, unique bands were detected in some seedlings (Figure 2, Table 6), indicating significant genetic differentiation among the triploidy accessions.

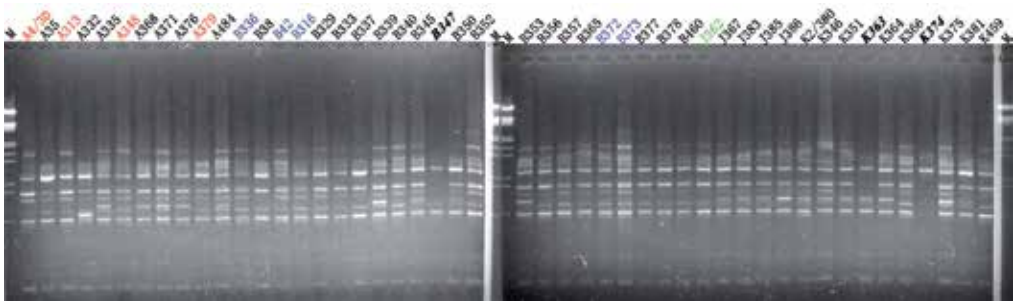


Fig. 2. The amplification pattern of ISSR primer 835. (M: λDNA/*Hind* III + *EcoR* I marker)

Band size	Exist	Missing
1500bp	A4/30(2x), A313, A335, A348, A368, A371, A376,A484	A35, A332, A379
770bp	K346, K366, K375, K381, K459; J367, J386	K2/380, K351, K363, K364, K374; J362(2x), J383, J385
700bp	A335, A368, A371, A376, A484; B339, B345, B350, B353; K2/380, K346, K351, K363, K364, K375, K381; J367, J383, J385, J386	A4/30(2x), A35, A313, A332, A348, A379; B336(2x), B38, B329, B333, B337, B340, B347, B352, B356, B357, B365, B377, B378, B460; K366, K374, K459; J362(2x)
660bp	A4/30(2x), A35, A313, A335, A348, A368, A371, A376, A379, A484	A332
550bp	B336(2x); K2/380	B38, B329, B333, B337, B339, B340, B345, B347, B350, B352, B353, B356, B357, B365, B377, B378, B460; K346, K351, K363, K364, K366, K374, K375, K381, K459

Table 6. The characteristic bands of primer 835.

6.2 AFLP (amplified fragment length polymorphism) analysis of natural triploid loquats

Wang (2008) assessed the effectiveness of AFLP markers in loquat using 6 natural triploid loquats and their maternal parents. With results indicated that, 12 pairs of AFLP primers amplified 2454 bands, as contrasted with maternal parent, there were 112 added bands and 96 lost bands of triploids, the number of amplified bands also differed among the clones of triploid loquats (Table 7). A369 has the greatest difference from its parents (25 additional bands and 19 missing sites), followed by A348 (20 new bands and 22 missing sites), A35 (22 new bands and 14 missing sites), A368 and A322 have the same number of different sites

(A368 has more missing bands and A322 has more new bands), and A313 has 12 new bands and 10 missing bands (Wang et al., 2011).

Types	Numbers of polymorphic bands						Numbers of polymorphic loci and their ratios[a]		Total bands
	A348	A368	A313	A35	A322	A369			
Added bands	20	14	12	22	19	25	112	53.9%	2454
Missing bands	22	18	10	14	13	19	96	46.1%	
Total	42	32	22	36	32	44	208	8.5%[b]	

Table 7. Statistics of polymorphic bands of natural triploids by AFLP analysis. [a] the ratios of one type of polymorphic bands and total polymorphic bands; [b] the ratios of polymorphic bands and total bands.

6.3 MSAP (methylation sensitive amplification polymorphism) analysis of natural triploid loquats

A total of 3879 bands were amplified with 12 pairs of primers within the group of six natural triploid loquat clones and their maternal plant, in which, 363 bands were fullly methylated and 241 bands were hemimethylated. The methylation ratios of six triploid lines were between 12.9% and 18.3%, 15.8% on average, full methylation ratios were between 7.6% and 11.7%, 9.7% on average, hemimethylation ratios were between 4.7% and 6.1%, 5.6% on average. All these belong to four patterns of methylation, monomorphism, demethylation, hyper-methylation and hypo-methylation, the number of sites and frequency were 251 and 29.2%, 171 and 19.9%, 334 and 38.9%, 103 and 12.0% respectively, and all of them exisited in all triploid lines (Table 8). All these indicted that, during the process of genome recombination and triploidization, a great number of hyper-methylation, demethylation, hypo-methylation and maintained methylation were proceeded (Wang, 2008).

	2x	A348	A368	A313	A35	A322	A369
Total methylated bands[a]	573	551	567	583	539	531	535
Full methylated bands[b]	44	69	43	59	63	51	53
Hemimethylated bands[c]	37	32	30	34	33	25	31
Total methylated bands	81	101	73	93	96	76	84
Full methylation ratios(%)	7.7%	12.5%	7.6%	10.1%	11.7%	9.6%	9.9%
Hemimethylation ratios(%)	6.5%	5.8%	5.3%	5.8%	6.1%	4.7%	5.8%
Total methylation ratios(%)	14.1%	18.3%	12.9%	16.0%	17.8%	14.3%	15.7%

Table 8. Genomic DNA methylation of natural triploid loquats and their female parent. [a] including full methylated and hemimethylated sites; [b] Full methylation denotes 5′-C^{m}CGG-3′in double strands; [c] Hemimethylation denotes 5′-^{m}C CGG-3′in single strand.

6.4 SSR (simple sequence repeat) analysis of natural triploid loquats

Fifty five pairs of polymorphism primers were screened, and a total of 135 alleles were detected with ten clones of 'Dawuxing'. The allele with 222 base pairs of CH01h02 was only found in the triploids. And there's three alleles of 238 bp, 236 bp and 230 bp with primer Hi15h12 of A322 (Figure 3). Similar results were obtained by Watanabe et al. (2008). New alleles emerged as compared diploid and each one of triploid individuals, indicating foreign genes maybe introgressed along with the formation progress of triploid individuals. All ten clones were completely distinguished from each other, the highest SM similarity coefficient was between A2x and A313, with 0.926, and the contrast one was between A313 and A332, with the similarity coefficient of 0.496. Principal component analysis divided 10 strains into three groups (Figure 4), Group I, including the A484, A376, A379 and A368 of four lines, A35, A322 and A332 are three lines constitute the group II, group III consists of A2x, A313 and A484 (He, 2010).

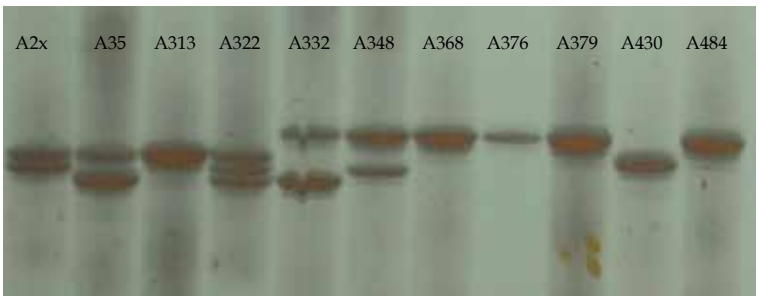


Fig. 3. The amplification pattern of SSR primer Hi15h12.

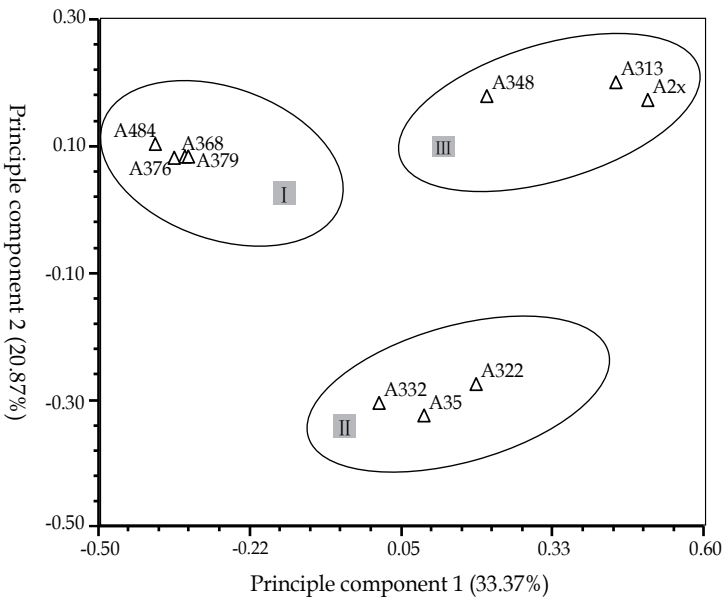


Fig. 4. Two dimensional plot of the principal components analyzed of 10 loquat individuals with 55 primer pairs, using the similarity matrix obtained with SM coefficient.

7. Molecular characterization of artificial triploid loquats

Seven artificial triploid loquats were obtained by sexual hybridization between tetraploid "Jiefangzhong" and diploid "Hunanzaoshu". Eleven ISSR primers amplified 1989 bands, to contrast with parents, the added, lost, agnate and maternal bands were 1, 4, 9 and 19 respectively, indicating that the genomes of hybrid progenies are more similar to maternal parent which provided more chromosomes, the added and lost bands different from both parents , suggesting substantial genome variations among the artificial triploid loquats.

Similar result was observed using AFLP analysis. A total of 3122 bands were amplified with 12 pairs of AFLP primers. The numbers of added, lost, agnate and maternal bands were 82, 58, 49 and 105 respectively. All these indicated that substantial degree of genome variation occurred druing the process of triploid formation (Table 9).

DNA methylation analysis of artificial triploid loquat individuals and their parents showed that, a total of 5302 bands were amplified with 12 pairs of primers, of which 605 bands were full methylated and 233 bands were hemimethylated. Total methylation ratios of triploidy F1 hybrids were between 13.2% and 17.8%, 15.4% on average, full methylation ratios were between 10.5% and 12.2%, 11.7% on average, hemimethylation ratios were between 2.2% and 5.5%, 3.9% on average. Relative to parental plants, transmutation tendency of toal methylation and full methylation ratios of the artificial triploids was not significant. However, the hemimethylation ratios in the artificial triploids all decreased. The frequencies of the five types of methylation, demethylation, hyper-methylation, hypo-methylation and intermediate pattern were 24.2%, 28.8%, 38.5%, 6.6% and 1.9% respectively (Table 10). Hyper-methylation occurred mainly during the development of artificial triploids, then demethylation, accompanied hypo-methylation, and methylation maintain and intermediate pattern that methylation state maintained between parent plants.

Types	Patterns of band			Numbers of polymorphic bands							Numbers of polymorphic loci and their ratios[a]			Total bands
	4x	2x	3x	H1	H2	H3	H4	H6	H8	H19				
Added bands	-	-	+	14	5	10	14	11	18	10	82		27.9%	
Lost bands	+	+	-	7	7	6	9	9	12	8	58		19.7%	
Agnate bands	-	+	+	1	5	5	4	6	6	5	32	49	16.7%	3122
	+	-	-	1	7	2	3	2	2	0	17			
Maternal bands	+	-	+	12	4	11	10	11	11	13	72	105	35.7%	
	-	+	-	8	4	4	5	3	3	4	33			
Total				43	32	38	45	42	52	40	294		9.4%[b]	

Table 9. Statistics of polymorphic bands of artificial triploids by AFLP technique. [a] the ratios of one type of polymorphic bands and total polymorphic bands; [b] the ratios of polymorphic bands and total bands.

	Parents		Artificial triploid loquats						
	4x	2x	H1	H2	H3	H4	H6	H8	H19
Total methylated bands[a]	567	596	599	547	592	614	609	599	579
Full methylated bands[b]	65	66	70	63	74	67	64	68	71
Hemimethylated bands[c]	35	35	15	29	30	14	27	13	32
Total methylated bands	100	101	85	92	104	81	91	81	103
Full methylation ratios(%)	11.5%	11.1%	11.7%	11.5%	13.5%	10.9%	10.5%	11.4%	12.3%
Hemimethylation ratios(%)	6.2%	5.9%	2.5%	5.3%	5.1%	2.3%	4.4%	2.2%	5.5%
Total methylation ratios(%)	17.6%	16.9%	14.2%	16.8%	17.6%	13.2%	14.9%	13.5%	17.8%

Table 10. Genomic DNA methylation of artificial triploid loquats and their parents. [a] including full methylated and hemimethylated sites; [b] Full methylation denotes 5′-C^{m}CGG-3′in double strands; [c] Hemimethylation denotes 5′-^{m}C CGG-3′in single strand.

8. Future research prospect

Loquat originated from China, and there's diverse loquat germplasm, including related species need to be investigated. Most of the selected natural or artificial triploidy seedlings were leading varieties, which only covers a very small fraction of the loquat diversity. Additional populations or germplasm groups should be used for breeding triploid loquat, and the hybridization between polyploidy and diploidy accessions should be explored. Furthermore, detailed characterization of all triploid seedlings should be carried out using new generation of molecular tools. As the genome sequencing of loquat will be finished, techniques based on transcriptomics and comparative genomics will enrich the genetic research of triploid loquats.

9. Acknowledgment

This work was mainly supported by the National Natural Science Foundation of China (Grant No. 30170656, 30971986); Transformation Fund for Agricultural Science and Technology Achievements Project, The Ministry of Science and Technology, China (Grant No. 2010GB2E020600); Special Fund for Agro-scientific Research in the Public Interest, Ministry of Agriculture, China (Grant No. 201003073); Key Scientific and Technological Research Project, Chongqing Science and Technology Commission, China (Grant No. CSTC, 2009AB1183); The Natural Science Foundation, Chongqing Science and Technology Commission, China (Grant No. CSTC, 2010BB1129); The Fundamental Research Funds for the Central Universities (Grant No. XDJK2009C123, XDJK2009C127); The Ph.D. Programs Foundation of Southwest University (Grant No. SWU111016, SWU111017) and the 211 Project of Southwest University.

10. References

Badenes M.L., Martínez-Calvo J., & Llácer G. (2000). Analysis of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.). *Euphytica*, Vol. 114, No. 3, (August 2000), pp. 187–194, ISSN 1573-5060

Badenes M.L., Lin S.Q., Yang X.H., Liu C.M., & Huang X.M. (2009). Loquat (*Eriobotrya* Lindl.), In: *Genetics and Genomics of Rosaceae*, Folta K.M. & Gardiner S.E., (Ed.), 525-538, Springer, ISBN 978-0-387-77490-9, New York, USA

Bian Yu. (2010). Analysis on genetic variation and DNA methylation of different loquat ploidies. Master Dissertation, Southwest University, Chongqing, China

Cai L.H. (2000). Allozyme analysis of genetic diversity, interspecific relationship and cultivar identification in genus *Eriobotrya*. PhD. Dissertation, Huazhong Agricultural University, Wuhan, China

Chen S.B., & Chu C.R. (2008). The self-compatiblity test of Guanyu (Baisha) loquat, *South China fruit,* Vol. 37, No. 5, (October 2008), pp. 43-44, ISSN 1007-1431

Dong Y.N. (2008). RAPD Analysis of genetic diversity of seedlins from minia ture seeds in loquat (*Eriobotrya japonica* Lindl). Master Dissertation, Sichuan Agricultural University, Ya'an, China

Einest J. (1952). Spontaneous polyploidy in cultivated apples. *Proc. Amer. Soc. Hort. Sci.* 59: 291-302

Gisbert A.D., Martínez-Calvo J., Llácer G., Badenes M.L., & Romero C. (2009). Development of two loquat [*Eriobotrya japonica* (Thunb.) Lindl.] linkage maps based on AFLPs and SSR markers from different Rosaceae species. *Mol Breeding*, Vol. 23, No. 3, (April 2009), pp. 523–538, ISSN 1572-9788

Gisbert A.D., Romero C., Martínez-Calvo J.M., Leida C., Llácer G., & Badenes M.L. (2009). Genetic diversity evaluation of a loquat (*Eriobotrya japonica* (Thunb) Lindl) germplasm collection by SSRs and S-allele fragments. *Euphytica*, Vol. 168, No. 1, (July 2009), pp. 121-134, ISSN 1573-5060

Guo Q.G., Li X.L., Wang W.X., He Q., & Liang G.L. (2007). Occurrence of natural triploids in loquat. *Proceedings of the second international Symposium on Loquat*, pp. 128-128, ISBN 978-90-66055-40-7, Guangzhou, China, April 1-6, 2006

He Q. (2010). Genetic diversity analysis and cultivar indentification of loquat (*Eriobotrya japonica* Lindl.) based on SSR. PhD. Dissertation, Southwest University, Chongqing, China

He Q., Li X.W., Liang G.L., Ji K., Guo Q.G., Yuan W.M., Zhou G.Z., Chen K.S., van de Weg W.E., & Gao Z.S. (2011). Genetic diversity and identity of Chinese loquat cultivars/accessions (*Eriobotrya japonica*) using apple SSR markers, *Plant Mol Biol Rep*, Vol. 29, No. 1, (March 2011), pp. 197–208, ISSN 1572-9818

Janick J. (2011). Predictions for loquat improvement in the next decade. *Proceedings of the third International Symposium on Loquat*, pp. 25-29, ISBN 978-90-66056-83-1, Antakya, Hatay, Turkey, May 3-6, 2010

Li P., Lin S.Q., Yang X.H., Hu G.B., & Jiang Y.M. (2009). Molecular phylogeny of *Eriobotrya* lindl. (loquat) inferred from internal transcribed spacer sequences of nuclear ribosome. *Pak. J. Bot*., Vol. 41, No. 1, (February 2009), pp. 185-193, ISSN 2070-3368

Liang G.L. (2006). Selection of natural triploid loquats and studies on their genetic characters and genome analysis. PhD. Dissertation, Southwest University, Chongqing, China

Liang G.L., Wang W.X., Xiang S.Q., Guo Q.G., & Li X.L. (2007). Genomic in situ hybridization (GISH) of natural triploid loquat seedlings. *Proceedings of the second*

international Symposium on Loquat, pp. 97-100, ISBN 978-90-66055-40-7, Guangzhou, China, April 1-6, 2006

Liang G.L., Wang W.X., Li X.L., Guo Q.G., Xiang S. Q., & He Q. (2011). Selection of large-fruited triploid plants of loquat. *Proceedings of the third International Symposium on Loquat*, pp. 95-100, ISBN 978-90-66056-83-1, Antakya, Hatay, Turkey, May 3-6, 2010

Lin S.Q., Sharpe R.H., & Janick J. (1999). Loquat: botany and horticulture. In: *Horticultural Reviews*, Janick J., (Ed.), 233–276, John Wiley & Sons, Inc., ISBN 978-0-470-65075-2, Oxford, UK

Lin S.Q. (2003). The world newest trends of production and scientific research of loquat, brief introduction of I international symposium on loquat, *News of citrus and subtropical fruit* Vol. 19, No. 3, (March 2003), pp. 8-10, ISSN 1008-2905

Lin S.Q. (2007a). World loquat production and research with special reference to China. *Proceedings of the second international Symposium on Loquat*, pp. 37-43, ISBN 978-90-66055-40-7, Guangzhou, China, April 1-6, 2006

Lin S.Q., Huang X.M., Cuevas J., & Janick J. (2007b). Loquat: an ancient fruit crop with a promising future. *Chronica Horticulturae*, Vol. 47, No. 2, (June 2007), pp. 12-15, ISBN: 978-90-6605-247-5

Ollitrault P., Dambier D., Luro F., & Froelicher Y. (2008). Ploidy manipulation for breeding seedless triploid Citrus. In: *Plant Breeding Reviews*, Janick J., (Ed.), 323-352, John Wiley & Sons, Inc., ISBN 978-0-470-17152-3, Hoboken, NJ, USA

Qiao Y.C. (2008). Construction of molecular genetic linkage map and genetic diversity of genus Eriobotrya. PhD. Dissertation, SouthChina Agricultural University, Guangzhou, China

Qiu W.L. & Zhang H.Z. (1996). *Fruit flora of China (longan and loquat)*, China Forestry Press, ISBN: 978-75-0381-604-8, Beijing, China

Shih J C. (2007). Loquat production in Taiwan. *Proceedings of the second international Symposium on Loquat*, pp. 55-60, ISBN 978-90-66055-40-7, Guangzhou, China, April 1-6, 2006

Snowdon R.J., Köhler W., Friedt W., & Köhler A. (1997). Genomic *in situ* hybridization in *Brassica* amphidiploids and interspecific hybrids. *Theor Appl Genet*, Vol. 95, No. 8, (December 1997), pp. 1320-1324, ISSN 1432-2242

Soriano J.M., Romero C., Vilanova S., Llácer G., & Badenes M.L. (2005) Genetic diversity of loquat germplasm (*Eriobotrya japonica* (Thunb) Lindl) assessed by SSR markers. *Genome,* Vol. 48, No. 1, (February 2005), pp. 108–114, ISSN 1480-3321

Vilanova S., Badenes M.L., Martínez-Calvo J., & Llácer G. (2001). Analysis of loquat germplasm (*Eriobotrya japonica* Lindl) by RAPD molecular markers. *Euphytica*, Vol. 121, No. 1, (August 2001), pp. 25-29, ISSN 1573-5060

Wang W.X. (2008). Genome variation and DNA methylation analysis of natural and artificial triploid loquats. PhD. Dissertation, Southwest University, Chongqing, China

Wang W.X., Xiang S.Q., Guo Q.G., Li X.L., He Q., & Liang G.L. (2011). Genome variation and genetic diversity of triploid loquats. *Proceedings of the third International Symposium on Loquat*, pp. 55-58, ISBN 978-90-66056-83-1, Antakya, Hatay, Turkey, May 3-6, 2010

Watanabe M., Yamamoto T., Ohara M., Nishitani C., & Yahata S. (2008) Cultivar differentiation identified by SSR markers and the application for polyploidy loquat

plants. *J Jpn Soc Hortic Sci*, Vol. 77, No.4, (October 2008), pp. 388–394, ISSN 0013-7626

Yang C. (2009). Studies on mechanism of embryo abortion and genetic diversity of seedlings from degenerated seeds by ISSR in loquat (*Eriobotrya japonica* Lindl.). Master Dissertation, Sichuan Agricultural University, Ya'an, China

Zhang H.Z., Peng S.A., Cai L.H., & Fang D.Q. (1990). The germplasm resources of the genus *Eriobotrya* with special reference on the origin of *E. japonica* Lindl. *Acta Horticulturae Sinica*, Vol. 17, No. 1 (April 1990), pp. 5–11, ISSN 0513-353X

Zhao G.J , Yang Z.Q, Chen X.P., & Guo Y.H. (2011) Genetic relationships among loquat cultivars and some wild species of the genus *Eriobotrya* based on the internal transcribed spacer (ITS) sequences. *Scientia Horticulturae*, Vo. 130, No. 4, (October 2011), pp. 913-918, ISSN 0304-4238

Zheng, S.Q. (2007). Achievement and prospect of loquat breeding in china. *Proceedings of the second international Symposium on Loquat*, pp. 85-92, ISBN 978-90-66055-40-7, Guangzhou, China, April 1-6, 2006

7

Genetic Diversity in *Aechmea fulgens* (Bromeliaceae: Bromelioideae) Revealed by Molecular Markers

Clébia Maria Alves de Almeida, Alexandre Gomes da Silva, Maria Tereza dos Santos Correia, Luciane Vilela Resende and Márcia Vanusa da Silva
Federal University of Pernambuco
Federal University of Lavras
Brazil

1. Introduction

The Brazilian Atlantic Forest biome contains a high level of genetic diversity and endemic species. In the Atlantic Forest of Northeastern Brazil (Figure 1), habitat loss and fragmentation are the major threats to biodiversity. The degree of degradation of these areas is even greater than that observed in other regions of Brazil (Kimmel et al. 2008). These remnants are only 2% of the original forest area and are mostly restricted to private areas. These forest remnants shelter many native bromeliad species that are threatened with extinction.

Theoretically, population changes associated with habitat degradation lead to an erosion of genetic variation and increased interpopulation genetic divergence due to increased random genetic drift, elevated inbreeding, and reduced gene flow, where remnant/population size and isolation fall below critical levels. These changes are also predicted to affect population viability in the short and long term. In the short term, disturbed plant populations are expected to suffer increased disease and pest susceptibility, loss of incompatibility alleles, and fixation of deleterious alleles. In the long term, loss of genetic variation is expected to reduce the ability of populations to respond to changing selection pressures (Lowe et al. 2005).

The objective of this work was to study the genetic diversity and structure of *Aechmea fulgens* populations through SSR and ISSR molecular markers. Three populations from different remnants of Atlantic Rainforest in Pernambuco state, located in the northeastern region of Brazil, were comparing regarding the level of polymorphism and other populational parameters.

1.1 Bromeliad history

Bromeliads are typical plants of the American continent and probably originated in the Andean region. The species spread out throughout the millennia and reached the tropical

forest about 200,000 years ago. In this taxonomic group, 2700 species were described, all native to the American continent except for a single species, *Pitcairnia feliciana*, which is found in West Africa (Benzing 2000).

Fig. 1. General view of an Atlantic Forest remnant in Northeastern Brazil. Igarassu, Pernambuco

Both terrestrial and epiphyte bromeliad are found even in the most desert and arid regions of the world. Bromeliaceae is subdivided into three sub-families: Bromelioideae, Tillandsioideae and Pitcairnioideae. It is estimated that about 40% of the species dwell in Brazil, many of which are endemic and concentrated mainly in the Atlantic Forest biome (Leme and Marigo 1993).

The bromeliads, being mostly epiphytes, depend on the forest conservation. The devastation process of the Atlantic Forest biome is the main cause of the genetic erosion of the bromeliad species. Only 18.06% of this forest is maintained in form of dispersed fragments in its associated ecosystems (Fundação SOS Mata Atlântica 2001).

The dynamics of tropical forest destruction have led to serious alterations in the ecosystems that compose the respective biomes due in particular to the high fragmentation of the habitats and to genetic erosion (Heringer and Montenegro 2000), making the establishment of germplasm conservation programs necessary and urgent.

1.2 *Aechmea fulgens*

The *Aechmea* genus has significant representation in Pernambuco state where there are 22 species distributed in different environments (Forzza et al. 2010). The *Aechmea fulgens* Brongn. species deserves attention as a plant with economic value because of its ornamental

potential. These species belongs to the Bromelioidea subfamily, Bromeliaceae family, and has epiphyte and soil habit and also occurs on rock outcrops. It can be found in the Atlantic Rainforest from Pernambuco, Bahia as far as Rio de Janeiro. In Atlantic Rainforest the bromeliads account for about 30% of the food resources used by hummingbirds and bats (Sazima et al. 1999). In spite of its great importance for tropical forests, little is known about the reproductive biology of the Bromeliaceae family (Balke et al. 2008). Some species of this family have mechanisms that favor cross pollination, but most are self incompatible. These processes are frequently associated to pollination by vertebrates (Fleming et al. 2009).

The ornamental value and lack of commercial production of the species has led to aggression against the native forests, from where great quantities are collected for commercialization on the domestic and foreign markets. Anthropological action and habitat fragmentation are the greatest causes of erosion of the biodiversity of tropical forests (Schaffer and Prochnow 2002).

The preservation of genetic diversity is one of the main objectives of conservation programs and the first step is knowledge of the distribution of this diversity within and among natural populations (Han et al. 2009). The most important characteristics in population genetics include understanding the gene flow and the genetic diversity and structure, because they are considered essential in the formulation of management and conservation strategies.

1.3 Molecular markers

The polymerase chain reaction (PCR) based on molecular markers such as microsatellite (SSR) and inter-simple sequence repeats (ISSR) are being extensively used to study the genetic diversity in a number of plant species at the species as well as cultivar level. To date, no report is available on applications of molecular markers in studies on the genetic diversity in *Aechmea fulgens* populations. In this investigation, a first attempt was made to study the genetic diversity among three populations of *A. fulgens* using SSR and ISSR markers.

The Microsatellites or SSR (simple sequence repeats) and ISSR (inter-simple sequence repeats) marker methodologies have been widely used in studies with plant populations. ISSR involves amplification of DNA segments present at an amplifiable distance in between two identical microsatellite repeat regions oriented in opposite direction. The technique uses microsatellites as primers in a single primer PCR reaction targeting multiple genomic loci to amplify mainly ISSR of different sizes. ISSRs exhibit the specificity of microsatellite markers, but need no sequence information for primer synthesis enjoying the advantage of random markers ISSR markers usually show high polymorphism although the level of polymorphism has been shown to vary with the detection method used. SSR markers often present high levels of inter- and intra-specific polymorphism, particularly when tandem repeats number is ten or greater. SSRs have been the marker of choice in most areas of molecular genetics as they are highly polymorphic even between closely related lines, require low amount of DNA, can be easily automated for high throughput screening, can be exchanged between laboratories, and are highly transferable between populations. SSRs are mostly codominant markers, and are indeed excellent for studies of population genetics and mapping (Semagn et al. 2006).

2. Material and methods

2.1 Species under study

Aechmea fulgens (Figure 2) is a bromeliad native to the Atlantic Rainforest. In addition to its ecological importance, it has economic importance as a plant with ornamental value. The specie was used in this study because of its wide distribution in the Atlantic Rainforest remnants in Pernambuco state and because of its ornamental potential. As certification that the species under study was indeed *Aechmea fulgens*, a specimen from each collection location was identified and placed in the herbarium at the Instituto Agronômico de Pernambuco (IPA 57000.250).

The species and the collection locations were selected following a survey carried out in the herbariums at the Instituto Agronômico de Pernambuco (IPA), the Pernambuco Federal University (UFPE) and the Pernambuco Federal Rural University (UFRPE). The survey considered: species common to the various regions, most recent data collection and importance of the species. The plant material was collected in remnants of the Atlantic Rainforest in three areas located in Pernambuco - region of Northeastern Brazil (Figure 3). Pieces of young leaves were collected from 20 individuals of each population.

2.2 Study areas

2.2.1 Charles Darwin Ecological Refuge (07º49' S - 34º56' W) – located in the region of Igarassu city on the northern coast of Pernambuco state (Brazil), with an area of approximately 60 hectares, whose vegetation is in the process of recomposition.

2.2.2 Tapacurá Ecological Station (08º04' S - 35º13' W) - located in eastern Pernambuco state in the region of São Lourenço da Mata city, with about 700 hectares. The mission of the Tapacurá Ecological Station is to carry out research in the fields of botany, zoology and ecology. Work at the station is aimed at developing habitats for conserving forest resources and fauna in the Brazilian Atlantic Forest.

2.2.3 Mata do Estado (07º 35'S - 35 º 30'W) - located in the region of São Vicente Férrer city, in north Pernambuco state, with area of approximately 600 hectares . Of the areas studied, this has suffered most from anthropological action.

2.3 Isolation of genomic DNA

Equal amounts (0.2 g dry weight) of leaf tissue from each sample were placed in porcelain mortars refrigerated with liquid nitrogen and ground to a fine powder. Total genomic DNA was extracted according to the CTAB method described by Doyle and Doyle (1990). RNA was removed by RNaseA treatment. DNA samples were spectrophotometrically quantified at 260 nm. The A260/280 reading to each DNA ranged from OD 1.8 to 2.0. After quantification, the DNA was diluted with ultra-pure water to a working concentration of 20 nanograms (ng) μL^{-1} for PCR analysis.

2.4 SSR assays

First 12 pairs of primers were selected, developed by Sarthou et al. (2003) and Boneh et al. (2003) for species of the *Pitcairnia, Tillandsia* and *Guzmania* genera, to test the clarity and

Fig. 2. *Aechmea fulgens*. Population of bromeliads in inner of an Atlantic Forest remnant in Igarassu - Pernambuco – Brazil (on the right) and inflorescence (on the left)

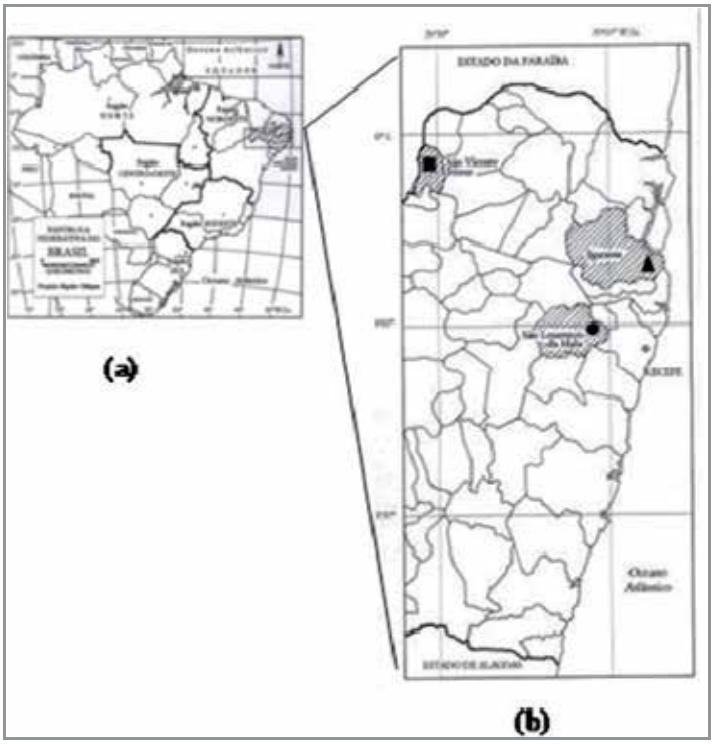


Fig. 3. (**a**) Pernambuco State – Brazil; (**b**) Sampling areas for *Aechmea fulgens* populations: ▲Igarassu, north coast; •São Lourenço da Mata and ■SãoVicente Férrer, area of forest north.

reproducibility for the analyses. Five pairs of primers were chosen for the amplification reactions that were performed in a TGradient 4.20 Whatman – Biometra (Table 1). With programmed cycles of 5 min at 94°C, followed by 30 cycles of 1 min at 94°C (denaturation); 1 min at 52°C (annealing); and 1 min at 72°C (extension), followed by 7 min at 72°C (final extension of the fragments). Each 20 µL reaction contained 10 mM Tris-HCl (pH 8.0), 2 mM $MgCl_2$, 100 µM of each triphosphate deoxynucleoside (DNTPs), 0.2 µM of each pair of primers (right and left) from Invitrogen, one unit of the *Taq* DNA polymerase enzyme and 20 ng DNA. The amplification products were resolved in 2.5% agarose gel stained with SyBr Gold (Molecular Probe).

2.5 ISSR assays

The ISSR primers used (Table 2) were selected from a set produced by the University of British Columbia, Vancouver, Canada for *Sphagnum angermanicum* and *Pogonatum dentatum*. The DNA was amplified in a TGradient 4.20 Whatman – Biometra. With the final volume of 25 µL each reaction contained 10 mM Tris-HCl (pH 8.0), 2 mM $MgCl_2$, 100 µM of each triphosphate deoxynucleoside (DNTPs), 0.2 µM primer from Operon Technologies, one unit of the *Taq* DNA polymerase enzyme and 20 ng DNA. The amplification program was: 94°C for 5 min, 94°C for 30s followed by 30-35 cycles at 50-55 °C, (depending on the primer used). The amplified DNA fragments were separated by electrophoresis in 2.0% agarose gel and stained with SyBr Gold (Molecular Probe).

2.6 Data analysis

2.6.1 ISSR and SSR markers

The ISSR marker fragments visualized in the gels were computed as presence (1) or absence (0). The results of these data generated a binary matrix that was analyzed by the POPGENE Program v. 1.31 (Yeh et al. 1999) to estimate some populational parameters: the percentage of polymorphic loci (*P*); *F* statistics by Wright (Φst); Nei genetic diversity (*h*) and the Shannon index (*I*). The Fst values; Φst were used to estimate the gene flow among the populations Nm = 1/4 (1/Fst – 1). The distance between pairs of populations was measured by the Nei genetic distance (Nei 1978). These measurements were then used to construct the dendrogram, using the UPGMA grouping method (Unweighted Pair Group Method with Arithmetic Average). The genetic distance matrix was compared to the geographic distance matrix by the Mantel test to verify whether they were associated. The values of the geographic distance were transformed in $\log_{10}$ for greater proximity to the normal distribution. The Mantel test was carried out by the GENES Program, using 1000 random permutations. The analysis of molecular variance (AMOVA) was used to analyze the distribution of the total within and among genetic variation of the populations, from ISSR marker data using the Program GENES v. 2005.6.1 (Cruz 2001).

The SSR markers were analyzed for the same populational parameters. Because they were co-dominant markers, the data entry, in the genes program, was computed with zero (for absence of fragments), one (for heterozygosity) and 2 (for homozygosity).

3. Results

The primers used to amplify the microsatellite loci (Table 1) generated, depending on the primer, from one to five fragments per locus with sizes ranging from 110 to 450-base pairs

(bp). The transferability of the molecular markers was validated among the *Guzmania* and *Pitcairnia* and *Aechmea* genera by amplifying loci of the same size in species of the referred genera of bromeliad. The percentages of polymorphic loci raged, respectively, from 80% to 90.3% for the SSR and ISSR markers and the number of alleles per loci ranged from 1.6 to 1.8 for the three populations (Igarassu, São Lourenço da Mata and São Vicente Férrer) and for the two molecular markers used. Figures 3 and 4 represent an example of the polymorphism obtained with SSR and ISSR primers, respectively.

Primer	Derived specie	Sequence (5′ – 3′)	Size (base pairs)
PgI	*Pitcairnia geyskesii*	F:TTGAGCCATGAACAATAGGG R:AGAATTCTAGTGGCAGTCCTC	450- 350
PgII	*Pitcairnia geyskesii*	F:GAGGATGAAGGATTTCCAAGG R:ACCGTCCCACGATAAGAGC	300
PgIII	*Pitcairnia geyskesii*	F:AACCATTACATGCACCCTCAC R:TCACTGGGGAAGCCATAGAG	200 -100
TGI	*Tillandsia fasciculata e Guzmania monostachya*	F:CGTACGAAGGTAAGCACAA R:CCGTTGAAGAGGTTAGAGG	250- 110
TGII	*Tillandsia fasciculata e Guzmania monostachya*	F:AATGAGTTTCAGTTTTAGAAGC R:CCAAGAAAAGAACGGATCA	250 -110

Table 1. SSR primers, derived species, respective base sequences and amplified fragment size in samples of *Aechmea fulgens.*

Table 2 shows the polymorphism and size of the fragments generated by the ISSR markers and reports the ISSR primers that were used and the percentage of polymorphism produced for the set of populations. The number of fragments ranged from eight to 24 and the fragment sizes ranged from 350-2000 bp. The total molecular genetic diversity for each population was observed both by the percentage of polymorphic loci (*P*) and the Shannon-Wiener (*I*) diversity index and the Nei genetic diversity (*h*).

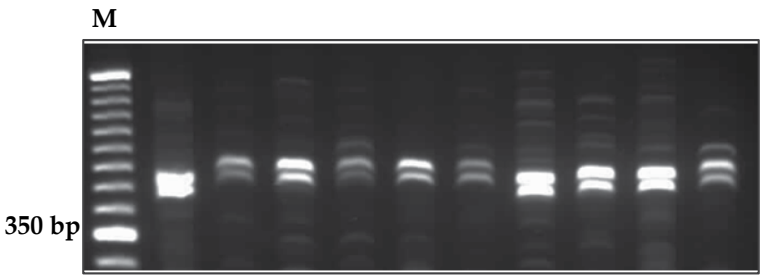


Fig. 3. SSR marker profiles of 10 samples of *A. fulgens* generated by primer PgI in 2.0% agarose gel. M – 50 bp DNA ladder

Primers UBC	Sequence (5′ → 3′)	Number of fragments	Number of polymorphics fragments	Polymorphism (%)
808	$(AG)_8C$	18	17	94
810	$(GA)_8T$	17	15	90
813	$(CT)_8T$	12	10	88
834	$(AG)_8YT$	12	10	88
842	$(GA)_8YG$	10	10	100
845	$(CT)_8RG$	9	9	100
848	$(CA)_8RG$	7	6	88
890	$VHV(GT)_7$	8	7	88
Total		**93**	**84**	**90,3**

The AMOVA analysis, based on 60 ISSR polymorphic loci, showed that 9.1% of the genetic variability was among population and 90.9% within population and the Φst value was, therefore, equal to 0.0903. The Fst (0.040) value detected for the data based on the SSR markers also suggested no differentiation among populations. The estimated gene flow was 2.5 for the analyses with ISSR markers and 5.74 for SSR loci.

Table 2. ISSR primers used and their respective base sequences, number of amplified fragments and polymorphism produced in samples of *Aechmea fulgens*.

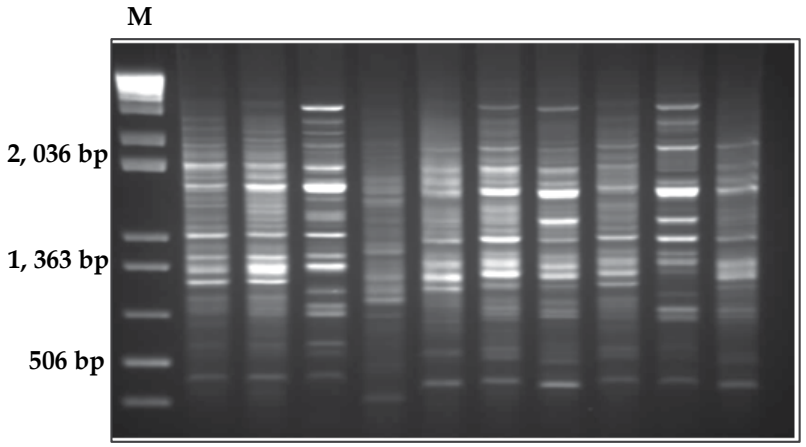


Fig. 4. ISSR marker profiles of 10 samples of *A. fulgens* generated by primer UCB 842 in 2.0% agarose gel. M - 1 kb Plus DNA ladder.

4. Discussion

4.1 Genetic diversity

Many studies of plants confirm the influence of plants' mating mode on the population genetic structure. Populations of clonal plants are expected to be dominated by one or a few locally adapted genets as distinct from plant population with predominant sexual propagation (Jogaitė et al. 2005). According to Murawski & Hamrick (1990), the mean

proportion of polymorphic loci for monocotyledons is 40.3%. These authors examined 18 isoenzyme loci in *Aechmea fulgens* populations and detected 21.1% mean polymorphism. The ISSR markers used in this study therefore generated much higher values than those reported by Murawski & Hamrick (1990) for monocotyledons. The same primers used in this study generated similar results in *Grevillea* (Proteaceae) (Pharmawati *et al*., 2004) and *Cycas guizhouensis* (Cycadaceae) (Xiao et al. 2004). In these studies the number of fragments ranged from 8-23 (UCB 808), 7-10 (UCB 842) and 8-24 (UCB 890). Pharmawati et al. (2004) obtained fragments for the oligonucletides UCB 808 (359-1558 bp) and UCB 890 (369-1385 bp), with sizes close to those detected in this study, that ranged from 350-2000 bp. The percentage of polymorphism (90.3%) was close to that reported by Cavallari et al. (2006), 89.8%, who studied populations of *Ecolirium biflorum* (Bromeliaceae) with RAPD markers.

Microsatellite molecular markers, widely used in plants, are considered to be very reliable because they are highly specific, usually producing a single fragment, frequently polymorphic at the infrageneric level (Oliveira et al. 2006). However, the number of fragments detected depends on the size of the genome, the frequency of the sequences repeated in the genome and the detection method (Barth et al. 2002). In the present study, the primers used to amplify the microsatellite loci (Table 1) generated one to five fragments per loci, depending on the primers. The allele pattern produced by the five pairs of primers used in this study permitted the characterization of the *Aechmea fulgens* species with microsatellites developed for use in studies of other species belonging to different genera.

According to Zane et al. (2003), microsatellites can be transferred among close species, or even among close genera. The transferability of these markers among related species is a consequence of the homology of the regions flanked by the microsatellites and the size of the region between the pair of primers responsible for amplification by PCR. The possibility of using microsatellites developed for one species in the genetic assessment of another is very useful because it reduces costs considerably. Boneh et al. (2003) suggested that the microsatellites they developed for the *Tillandsia* and *Guzmania* genera could be applied to other species within the genus and even to close genera. According to the phylogenetic genetic analysis of the bromeliad family reported by Terry et al. (1997), the *Aechmea fulgens* genus is a very different from those referred to by Boneh et al. (2003) but two primers used by the these authors were useful in the characterization of *Aechmea fulgens*.

The percentage of polymorphic loci (*P*) detected in this study was high compared with tropical tree species, for the three populations studied. Auler et al. (2002) reported a mean of 43% of polymorphic loci and 1.8 alleles per loci in *Araucaria angustifolia* populations with isoenzymatic markers, and considered the values to be high because these populations had a good conservation `status´ compared to other more degraded populations (20% and 1.4).

The total molecular genetic diversity for each population can be observed both by the percentage of polymorphic loci (*P*) or by diversity indexes of Shannon-Wiener (*I*) and Nei's genetic diversity (*h*).

The values detected for the Shannon-Wiener diversity index in this study were, respectively, for the SSR and ISSR markers: Igarassu: $I = 0.55$ and 0.53; São L. da Mata: $I = 0.51$ and 0.51; São V. Ferrer: $I = 0.51$ and 0.56. The Nei genetic diversity ranged from 0.35 to 0.39 in the three populations and for both markers used. Measures of allelic richness are suitable for

assessing short-term diversity loss, but for longer-term loss, involving commoner alleles, measures of gene diversity are more appropriate (Lowe et al. 2004).

4.2 Genetic distances

The Nei genetic distances, calculated for the pairs between populations, ranged from *approximately* 2% to 7% for the SSR and 4% to 7% for the ISSR markers (Tables 3 e 4). The dendrograms among populations distances for both the molecular markers formed two distinct groups: populations 1 and 3 and population 2 (Figure 5 A and B), where the geographic closer populations were in different groups, suggesting absence of association between geographic distance and genetic distance. This result was confirmed by the Mantel test that showed absence of correlation between the Nei genetic distance matrix and geographic distance for the two markers used. It is important point out that the two types of molecular markers gave the same grouping in spite of their different natures.

As the Nei distance value was proportional to the time since divergence and to the genetic substitution rate by loci and generation (Dias 1998) the values of the genetic distances detected in this study and the grouping formed indicated that the populations did not diverge a long time ago suggesting an explanation for the high within population values and lower among population values.

4.3 Genetic structure

The total genetic diversity revealed that the greatest part of the variation was present within the populations. The results of high within population diversity were in line with several studies with natural plant populations (Auler et al., 2002; Zucchi et al., 2003; Souza et al., 2004; Xiao et al., 2004; Galeuchet et al., 2005; Cavallari et al., 2006). Galeuchet et al. (2005) used seven microsatellite loci and detected Fst = 0.02 in *Lychnis flos-cuculi* L. (Caryophyllaceae). The authors suggested that the low differentiation among populations studied was primarily due to the significant level of historic gene flow (*Nm*) when the populations were more connected and second by the contemporary gene flow and accidental anthropogenic seed dispersion.

The *Nm* value determines whether the gene flow alone can produce substantial genetic variability among locations. If *Nm* is greater than 1.0, the gene flow will be high enough to prevent differentiation due to drift (Moraes & Derbyshire, 2002). In the present study, the estimated gene flow among the populations was high (2.5 and 5.7 for the ISSR and SSR markers, respectively) that together with the high within population genetic variation suggested low impact from the habitat fragmentation on the genetic diversity of *Aechmea fulgens* in the three environments studied. The *Nm* values, according to Galeuchet et al. (2005), should reflect the historic gene flow, because although bromeliads are probably pollinated by birds and/or bats that have a great flight range, most present clonal propagation. Furthermore, the genetic distance dendrograms (Figure 5 A and B) indicated no divergence among the populations studied.

Foré et al. (1992) studied the genetic structure after fragmentation in 15 *Acer saccharum* (Aceraceae) populations and reported the isolation of remnants with levels of genetic diversity. The authors observed no divergence among the remnants (Fst = 0.03) and that the gene flow was greater after fragmentation, in this case, due to the greater incidence of wind

and that fragmentation did not result in the isolation of the populations. Souza et al. (2004), however, disagreed with Foré et al. (1992) because they observed high genetic divergence within fragmented populations of *Chorisia speciosa* St. Hil. (Bombacaceae) that indicated, according to the authors, that fragmentation led to genetic drift and to the increase in within population divergence. On the other hand, Nei (1978) stated that when there is reduction in the population size, even though the number of alleles is reduced, the degree of heterozygosity and the genetic diversity can remain as high as in the original population. This is due in most cases to the fact that the alleles lost by drift are rare and contribute little to the level of heterozygosity.

Population (place)	Igarassu	São L. da Mata	São V. Ferrer
Igarassu	-	0.03	0.02
São L. da Mata	9 km	-	0.07
São V. Ferrer	74 km	98 km	-

Table 3. Nei's genetic distance matrix (higher diagonal) and geographical distance (lower diagonal) between pairs of populations of *A. fulgens*. SSR marker data.

Population (place)	Igarassu	São L. da Mata	São V. Ferrer
Igarassu	-	0.06	0.04
São L. da Mata	9 km	-	0.07
São V. Ferrer	74 km	98 km	-

Table 4. Nei's genetic distance matrix (higher diagonal) and geographical distance (lower diagonal) between pairs of populations of *A. fulgens*. ISSR marker data.

The genetic structure of *A. fulgens* populations indicates that in situ conservation strategies for this species should be based on the establishment of several small-scale protected areas chosen so as to represent the species' genetic variability. A similar *in situ* conservation schedule was previously suggested by Laguna et al. (1998) for endemic, rare or threatened plant species. Moreover, considering that the samples collected in this study provide a snapshot of the species' distribution area as a whole in Northeastern Brazil and that the genetic variability detected is low, *ex situ* conservation should also be considered in conservation strategies d if *A. fulgens* germplasm represents its overall genetic variability.

5. Conclusions

Nowadays, the fragmentation and isolation of natural habitats is one of the main threats to the persistence of many animal and plant species. In this study, we used molecular tools to investigate for the first time the genetic background of the *Aechmea fulgens* populations gene pool. The genetic among population differentiation was small suggesting that the historic gene flow among the populations was still high or that fragmentation did not happen long enough ago to result in pronounced differentiation. Therefore drastic genetic consequences were not shown on the *A. fulgens* populations in the three forest remnants studied.

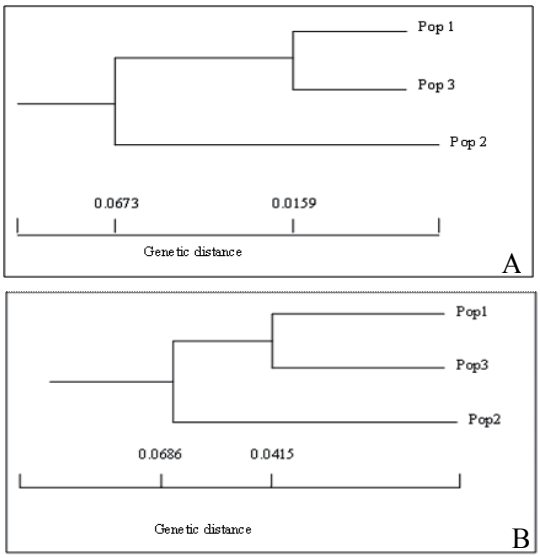


Fig. 5. Nei (1978) Genetic distance dendrogram among the three *Aechmea fulgens* populations. Pop 1: Igarassu; Pop 2: São L. da Mata; Pop 3: São V. Ferrer; A - SSR markers; B – ISSR markers.

Moreover, considering that the samples collected in this study provide a snapshot of the species' distribution area as a whole in Northeastern Brazil and that the genetic differentiation detected is low, *ex situ* conservation should be considered in conservation strategies these bromeliads. Despite the possibility of *ex situ* conservation, *in situ* conservation must be considered as a priority. More extensive studies must be carried out to delineate areas protected by law that effectively help to preserve these species. Furthermore, the potential of the ISSR and SSR markers in studies of population structuring in *A. fulgens* was also demonstrated, this possibly constituting a useful aid in the studies of other bromeliads species.

6. Acknowledgments

The authors thank the Empresa Instituto Agronômico de Pernambuco (IPA), Pernambuco, Brazil, for support and infrastructure and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/PADCT), Brazil, for the funding.

7. References

Auler, N.M.F.; Reis, M.S.; Guerra, M.P. & Nodari, R.O. (2002). The genetics and conservation of *Araucaria angustifolia*: I. Genetic structure and diversity of natural populations by means of non-adaptive variation in the state of Santa Catarina, Brazil. *Genetic and Molecular Biology,* Vol. 25, No. 3, pp. 329-338, ISSN 1415-4757

Balke, M.; mez-Zurita, J.O.; Ribera, I.; Viloria, A.; Zillikens, A.; Steiner, J.; Garcı´a, M.; Hendrich, L. & Vogler, A.P. (2008). Ancient associations of aquatic beetles and tank bromeliads in the Neotropical forest canopy. *Proceedings of the National Academy of Sciences*, Vol. 105, pp. 6356–6361, ISSN 1091-6490

Barth, S.; Melchinger, A.E. & Lübberstedt T.H. (2002). Genetic diversity in *Arabidopsis thaliana* L. Heynh. investigated by cleaved amplified polymorphic sequence (CAPS)

and inter simple sequence repeat (ISSR) markers. *Molecular Ecology,* Vol.11, pp.495-505, ISSN 0962-1083

Benzing, D. H. (2000). *Bromeliaceae: profile of an adaptive radiation*. Cambridge University Press, Cambridge

Boneh, L.; Kuperus, P. & Van Tienderen, H. (2003). Microsatellites in the bromeliads *Tillandsia fasciculata* and *Guzmania monostachya*. *Molecular Ecology Notes,* Vol. 3, pp.302-303, ISSN 1471-8278

Cavallari, M.M.; Forzza, R.C.; Veasey, E.A.; Zucchi M.I. & Oliveira, G.C.X. *(*2006). Genetic variation in three endangered species of Encholirium (Bromeliaceae) from Cadeia do Espinhaço, Brazil, detected using RAPD markers. *Biodiversity* and *Conservation* Vol.15, pp.4357-4373, ISSN 0960-3115

Cruz, C.D. (2001). *Program GENES*: A Software in the Area of Genetics and Experimental Statistics. Version 2005.6.1 for Windows. Universidade Federal de Viçosa <http://www.ufv.br/dbg/genes/genes.htm >

Dias, L.A.S. (1998). Análises Multidimensionais, In: *Eletroforese de Isoenzimas e Proteínas afins: fundamentos e aplicações em plantas e microrganismos*, Alfenas, A.C. (ed), pp. 201-317, Editora UFV, ISBN 85-7269-027-1, Viçosa, Minas Gerais, Brasil.

Doyle, J. J & Doyle, J, L (1990). Isolation of plant DNA from fresh tissue. *Focus,* Vol. 12, pp.13 -15.

Fleming, T.H.; Geiselman, C. & Kress, W.J. (2009). The evolution of bat pollination: a phylogenetic perspective. *Annals of Botany*, Vol.104, No.6, pp.1017-1043, ISSN 0305-7364

Foré, A.S.; Hickey, R.J.; Vankat, J.L.; Guttman, S. & Shaefer R. (1992). Genetic structure after forest fragmentation: a lands ecology perspective on *Acer sacharum*. *Canadian Journal of Botany,* Vol. 70, pp.1659-1668, ISSN 0008-4026

Forzza, R.C.; Costa, A.; Siqueira Filho, J.A. & Martinelli, G. (March 2010). Bromeliaceae, In: Lista de Espécies da Flora do Brasil, 22.03.2010, Available from http://floradobrasil.jbrj.gov.br/2010/FB005753/

Fundação SOS Mata Atlântica. (2001). SOS Mata Atlântica, http://www.sosmatatlantica.org.br, Assessed 09 August 2011

Galeuchet, D.J.; Perret, C. & Fischer, M. (2005). Microssatellite variation and structure of 28 populations of the common wetland plant, Lychnis flos-cuculi L., in a fragmentaed landscape. *Molecular Ecology*, Vol. 14, pp.991-1000, ISSN 0962-1083

Han, Y.C.; Teng, C.Z.; Wahiti, G.R.; Zhou, M.Q.; Hu, Z.L. & Song, Y.C. (2009). Mating system and genetic diversity in natural populations of Nelumbo nucifera (Nelumbonaceae) detected by ISSR markers. *Plant Systematics and Evolution*, Vol. 277, pp. 13–20, ISSN 0378-2697

Heringer, H. & Montenegro, M. M. (2000). *Avaliação e ações prioritárias para a conservação da biodiversidade da Mata Atlântica e Campos Sulinos*, MMA/SBF, ISBN 85-87166-09-3, Brasília

Jogaitė, V.; Naujalis, J. R.; Lazutka, J. R. (2005). Genetic diversity in fragmented population of herb-Paris *(Paris quadrifolia* L., Trilliaceae). *Biologija,* No. 2, pp. 5–12.

Kimmel, T.; Gottsberger, G. & Piechowski, D. (2008). The History of Fragmentation of the Lowland Atlantic Forest of Pernambuco, Brazil. In: *The Fragments Project*, Rodal, M. J. N.; Lins e Silva, A. C. B. & Gottsberger, G. (eds), 1-4, Bioremediation, Biodiversity and Bioavailability, Vol.2 (Special Issue 1), ISSN 1749-0596

Laguna, E. (1998). *Flora endémica, rara o amenazada de la Comunidad Valenciana*. Generalitat Valenciana, Conselleria de Medio Ambiente, ISBN 84-482-1879-5, Valencia, España

Leme, E. M. C. & Marigo, L. C. (1993). *Bromélias na natureza*, Salamandra (1ª. Edition), ISBN-13: 9788585352035, Rio de Janeiro

Lowe, A. J.; Harris. S. A. & Ashton, P. (2004). *Ecological Genetics: Design, Analysis and Application*, Blackwells, Oxford

Lowe, A. J.; Boshier, D.; Ward, M.; Bacles, C. F. E. & Navarro, C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, Vol. 95, pp. 255–273, ISSN 0018-067X

Moraes, P. L. R. & Derbyshire, M. T. V. C. (2002). Estrutura genética de populações naturais de *Cryptocarya aschersoniana* mez. (Lauraceae) através de marcadores isoenzimáticos. *Biota Neotropica*, Vol.2, pp. 2-17, ISSN 1676-0603

Murawski, D. A. & Hamrick, J. L. (1990). Local genetic and clonal structure in the tropical terrestrial bromeliad, *Aechmea magdalenae*. *American Journal of Botany,* Vol.77, pp.1201-1208, ISSN 0002-9122

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics,* Vol.89, pp. 583-590, ISSN 0016-6731

Oliveira, E. J.; Pádua, J. G.; Zucchi, M. I.; Vencovsky, R. & Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology,* Vol.29, pp. 294-307, ISSN 1415-4757

Pharmawati, M.; Yan, G. & McFarlane, I. J. (2004). Application of RAPD and ISSR markers to analyse molecular relationships in *Grevillea* (Proteaceae). *Australian Systematic Botany,* Vol.17, pp. 49–61, ISSN 1030-1887

Sarthou, C.; Boisselier-Dubayle, M. C.; Lambourdière, J. & Samadi, S. (2003). Polymorphic microsatellites for the study of fragmented populations of *Pitcairnia geyskesii* L. B. a specific saxicolous species of inselbergs in French Guiana. *Molecular Ecology Notes,* Vol.3, pp. 221-223, ISSN 1471-8278

Sazima, M.; Buzato, S. & Sazima, I. (1999). Bat-pollinated flower assemblages and bat visitors at two Atlantic Forest Sites in Brazil. *Annals of Botany,* Vol.83, pp. 705-712.

Schaffer, W.B. & Prochnow, M. (2002). *A Mata Atlântica e você,* Apremavi, ISBN 978-85-88733-05-3, Brasília, Brasil.

Souza, L. M. F. I.; Kageyama, P. Y. & Sebbenn, A. M. (2004). Genetic structure in fragmented populations of *Chorisia speciosa* St. Hil (Bombacaceae). *Scientia Forestalis,* Vol.65, pp. 70-79, ISSN 1413-9324

Semagn, K.; Bjørnstad, Å. & Ndjiondjop, M. N. (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology*, Vol. 5, No. 25, pp. 2540-2568, ISSN 1684–5315

Terry, R. G.; Brown, G. K. & Olmstead, R.,G. (1997). Examination of subfamilial phylogeny in Bromeliaceae using comparative sequencing of the plasticid locus *ndhF*. *American Journal of Botany,* Vol.84, pp. 664-670, ISSN 0002-9122

Xiao, L. Q.; Ge, X. J.; Gong, X.; Hao, G.; Zheng, S. X. (2004). ISSR Variation in the Endemic and Endangered Plant Cycas guizhouensis (Cycadaceae). *Annals of Botany (London)*, Vol.94, pp.133-138, ISSN 0305-7364

Yeh, F. C.; Boyle, T.; Yang, R. C. (1999). POPGENE, the user friendly shareware for population genetic analysis, version 1.31. University of Alberta and Centre for International Forest Research.

Zucchi, M. I.; Brondani, R. P. V.; Pinheiro, J. B.; Chaves, L. J.; Coelho, A. S. G. & Vencovsky, R. (2003). Genetic structure and gene flow in Eugenia dysenterica DC in the Brazilian Cerrado utilizing SSR markers. *Genetics and Molecular Biology,* Vol.26, pp. 449-457, ISSN 1415-4757

Zane, L.; Bargelloni, L. & Patarnello, T. (2003). Strategies for microsatellite isolation: a review. *Molecular Ecology*, Vol.11, pp.1-16, ISSN 0962-1083

Part 2

Genetic Diversity Analysis

8

Genetic Diversity in Tomato (*Solanum lycopersicum*) and Its Wild Relatives

Guillaume Bauchet and Mathilde Causse
Institut National de la Recherche Agronomique (INRA),
Unité de Génétique et Amélioration des Fruits et Légumes (GAFL),
France

1. Introduction

Tomato, ranking 1st in the world for vegetables, accounts for 14% of world vegetable production (over 100 million metric tons/year $ 1.6 billion market; (Food and Agriculture Organisation [FAO] 2010). Tomato is a rich source of micronutrients for human diet. It is also an acknowledged model species for research on fruit development and metabolite accumulation. The major goals of tomato breeders (higher productivity, better tolerance to biotic and abiotic stresses and increased sensory and health value of the fruit) require a good comprehension and management of tomato genetic resources diversity.

Due to its Latin American origin and related domestication history, cultivated tomato has faced several bottlenecks over ages. This led to a drastic reduction of its genetic diversity. Explorations of tomato center of origin permitted major advances in the characterization of its diversity. In parallel, *ex situ* plant conservation initiatives bloomed, ensuring the collection and conservation of landraces and wild species through development of seed banks. Thus, unraveling the genetic potential of tomato's wild relatives for breeding purpose emerged. In parallel, the ecological and taxonomic diversity of tomato turned it into a model species for evolutionary studies. Since the mid-20th century, new methods such as controlled hybridization allowed crossing between wild and cultivated tomato. Modern genetics and breeding methods contributed to understand the genetic control of agronomical traits but also accentuated the progress. If successful, the accuracy to introgress agronomic traits of interest from wild relatives into cultivated tomato was not always straightforward. This was notably due to inherent linkage between "favorable" and "unfavorable" effects of introgressed fragments.

The advent of molecular biology in the 80's raised great hopes in terms of characterization of the genetic diversity present in both wild and cultivated compartments. Also, great expectations emerged since the development of molecular techniques to "pinpoint" genomic regions involved in targeted traits. Dissection of the genetic control of complex traits, using ad hoc techniques from quantitative genetics, was possible, leading to the identification of key alleles involved in diverse agronomic traits, originating from several wild relatives.

Today the tomato genome is fully sequenced. A new step in the knowledge on tomato diversity with the so called "-omics" and next generation sequencing techniques is coming.

These technologies and related data analysis allow a complete and combined reading of genomes and related levels of expression (transcriptome, proteome, metabolome) in a high throughput way. Among the new approaches, QTL mapping techniques in natural populations or genome wide association studies will facilitate the genetic characterization of complex traits and germplasm management of both wild and cultivated tomatoes.

In this chapter we will first show how tomato diversity evolved from its early domestication until today. We will discuss how valuable tomato genetic resources are, and that investigating natural variation not only highlights existing diversity -which is of critical use for cultivated tomato improvement- but can also provide insights into the evolution and genetic bases of complex traits. In the last part, we will present how molecular markers have completed our view.

2. Diversity of the tomato clade species

Tomato belongs to the large and diverse *Solanaceae* family also called Nightshades which includes more than three thousand species. Among them, major crops arose from Old world (Eggplant from Asia) and New world (pepper, potato, tobacco, tomato from South America). The *Lycopersicon* clade contains the domesticated tomato (*Solanum lycopersicum*) and its 12 closest wild relatives (Peralta and Spooner 2005). The radiation of tomato clade has been estimated to 7.8 (Nesbitt and Tanksley 2002) and to 2.7 Million years between *S. lycopersicum* and *S. pennellii*(Kamenetzky et al. 2010). First detailed studies on this group of wild relatives were made by Charles Rick and colleagues since the 40's. Tomato clade species are originated from the Andean region, including Peru, Bolivia, Ecuador, Colombia and Chile (Figure 1). On Figure 1, *lycopersicon* species distributions are defined according to geographic data from the Tomato Genetics Resource Center, UC Davis http://tgrc. ucdavis.edu/Data/Acc/dataframe.aspx? start=GIS_ dataoption. aspx&navstart=nav.html. Their growing environments range from near sea level to 3,300 m altitude, from arid to rainy climate and from Andeans Highlands to the coast of Galapagos Islands (*S. cheesmaniae; S. galapagense*). Their habitats are often narrow and isolated valleys where they were adapted to particular microclimates and various soil types. Their very large range of ecological conditions contributed to the diversity of the wild species. This broad variation is also expressed at the morphological, physiological, sexual and molecular levels (Peralta and Spooner 2005). Over times, several phylogenetic classifications have been proposed and several adjustments occurred. Being first classified in the *Solanum* genus, the group turned to a specific genus, *Lycopersicum* (Miller, 1731). It recently got renamed *Solanum* within an updated classification (Peralta and Spooner 2001). Taxonomic, ecological, reproductive, breeding specificities for each member of the clade are listed in Table 1 and reviewed by Peralta and colleagues (Peralta, Spooner et al. 2007). The first classification was morphology based (Luckwill 1943). Later molecular data confirmed tomato membership of Linnaeus classification, but also improved subtaxa classification (Spooner 2008). The tomato clade is an interesting example for research on plant biodiversity, notably, on evolution, adaptation, human domestication and nutrition perspectives (Peralta and Spooner 2007). Nowadays, across South America, populations of wild tomatoes are being severely reduced. Their natural habitats are shrinking due to urban development and intensive agriculture as well as goat herding in the highlands, as recently documented by a botanical expedition in Peru. (Grandillo, Chetelat et al. 2011).

Fig. 1. Geographic distribution of wild species in *Solanum* section *lycopersicon*.

Many studies were conducted on evolutionary aspects of the lycopersicon clade. The mating system was extensively studied, using the clade as a model to study its effects on species variation (Bedinger, Chetelat et al. 2011). Mating system has played a key role in evolution of wild tomatoes, varying from allogamous self-incompatible, to facultative allogamous, to autogamous and self-compatible (Table 1). Flower stigma exertion and gametophytic incompatibility system contribute in greater outcrossing and genetic diversity. All the species of the clade are intercrossable (Table 1), but with a variable success rate (Rick, Fobes et al. 1977a; Rick, Fobes et al. 1979). Fruit color discriminate the wild relative species. Most of the latter carry green fruits, with the exception of the two species from the Galapagos (with yellow and orange fruits) and *S. pimpinellifolium*, which is the only wild relative species with red fruits. *S. pimpinellifolium* fruits are round, small, weighing only few grams. These fruits are edible and the species referred as the currant tomato. The plant presents a reduced apical dominance and prostrate growth habit resulting in a large shrub with inflorescence carrying many flowers and fruits (Paran and van der Knaap 2007). *S. pimpinellifolium* undergone bottleneck only recently with a drastic reduction of its natural habitats and is now an endangered species (Biodiversity-International 2006). *S. lycopersicum* var *cerasiforme* fruit is larger than *S. pimpinellifolium* and is commonly round and red. This subspecies of tomato is referred to as the "cherry tomato". It has been proposed as the direct ancestor of cultivated tomato because of its diversity, its wide spread occurrence in central America and its close genetic relationship with cultivated tomato (Rick and Chetelat 1995). The modern cultivated tomato, *S. lycopersicum*, is cosmopolite. It has spread all around the world and is now cultivated under a broad range of environments and conditions.

3. Tomato domestication in South America

Domestication is a special type of species diversification, distinct from species divergence through natural selection in the wild (Darwin and Wallace 1858). Domesticated species differ from wild and relative species for a set of traits known as the domestication syndrome (Doebley, Gaut et al. 2006). Domestication is often controlled by a limited number of chromosomal regions with major phenotypic effect (Purugganan and Fuller 2009). In tomato, edible fruits, attractive red color and fruit size increase are characterizing this process.

The domestication time of tomato is unclear. It is supposed to be due to a recent divergence from S. *pimpinellifolium*. The first hypothesis supports Peru as the center of origin and domestication (de Candolle 1882). This hypothesis gives emphasis on botanical evidences and has been complemented by botanical, linguistic and historical aspects. It was further supported by other colleagues (Müller 1940a; Müller 1940b; Luckwill 1943) and recent molecular studies (Nesbitt and Tanksley 2002). Nevertheless, very little and unclear archeological evidences are available to clearly support this hypothesis (McMeekin 1992). The second hypothesis supports that domestication occurred primarily in Mexico in the Vera Cruz Puebla area (Jenkins 1948), as there is no evidence for pre-Colombian cultivation of tomato in South America but good evidences in Mexico. Referring to Guilandini (1572), Jenkins also argued that tomato name comes most probably from the Mexican Nahua people word "Tomatl" that described "plants bearing globous and juicy fruit" (Sahagún 1988). Based on its

Subsection	Species New Nomenclature [previous nomenclature]	Geographic distribution and habitat	Mating system / Cross compatibility S.lycopersicum	Fruit color	Genetic polymorphism
Arcanum	*S. arcanum* [*L. peruvianum*]	Northern Peru, 100-2500m, Coastal and inland Andean valleys, in lomas, dry valleys, and on dry rocky slopes	SI[a], facultative AL[b] / UI[c], EL[d]	green with dark green stripes	Intermediate
	S. chmielewskii [*L. chmielewskii*]	South Peru to North Bolivia native, 1500-3000 m, dry and drained areas	SC[b], facultative AL / reciprocal	green with dark green stripes	Intermediate
	S. neorickii [*L. parviflorum*]	South Equador-south center Peru native, 1300-3000m, rocky, humid and well drained areas	SC, highly AT[d] / reciprocal	green with dark green stripes	Low
Neolycopersicon	*S. pennellii* [*L. pennellii*]	Peruvian cost native, 0-2000 m, dry and rocky hillsides.	SI usually, SC populations in southern parts /reciprocal	green	High
Eriopersicon	*S. habrochaites* [*L. hirsutum*]	South west Equador to south center Peru native , 500-3300, forest regions	SI, SC populations in southern parts / UI	green with darker green stripes	High
	S. chilense [*L.chilense*]	South Peru to North Chili, 0-3000 m, dry river bed	SI, AL / UI, EL	green to whitish	High
	S. huaylasense [partly *L. peruvianum*]	Peru, 1700-3000m, rocky slopes around Callejon de Huaylas	SI, AL / UI, EL	green with purple green with dark green stripes	High
	S. peruvianum [*L. peruvianum*]	Central Peru to northern Chile,0-600 m, lomas formations and occasionally in coastal deserts	SI, AL / UI, EL	green to greenish white,sometimes	High
	S. corneliomulleri [partly *L. peruvianum*, known as *L. glandulosum*]	Southern Peru,1000-3000 m, Middle to high elevations on the western slope of the Andes, lower slopes on the edges of landslides	SI, AL / UI, EL	green with dark green stripes	High
Lycopersicon	*S. chmielewskii* [*L. chmielewskii*]	Galápagos islands endemic species,0-1300 m, From sea shore to volcanic area .	SC, AT / reciprocal	yellow, orange	Low
	S. galapagense [partly *L. chmielewskii*]	Galápagos islands endemic species, sea shore.	SC, AT / reciprocal	yellow, orange	Low
	S. pimpinellifolium [*L. pimpinellifolium*]	South Equador-North Peru native, under 1000 m, south valleys of the pacific coast.	SC, AT, facultative AL / reciprocal	red	Intermediate
	S. lycopersicum var. *cerasiforme** [*L. esculentum* var. *cerasiforme*]	Adventive worldwide in tropics and subtropics, probably native from Andeans region	SC, AT, facultative AL / reciprocal	red	Low
	S. lycopersicum [*L. esculentum*]	Probably Equador-Peru, nowadays widely spread, various range of habitats	SC, AT	red	Very Low

(a) Self-incompatible (b) Self compatible (c) Allogamous (d) Autogamous * cults group

(1) Unilateral incompatibility (2) Embryo lethality (embryo-rescue technique required)

Table 1. Principal features of the *lycopersicon* subsection (*Solanum* sect. *Lycopersicon*) Data are compiled from Peralta *et al*. 2007, Moyle *et al*. 2008, Grandillo *et al*. 2011

center theory, Harlan suggested that biloculed domesticated forms found in south Mexico and Guatemala are the oldest cultivated types (Harlan 1971). Quoting Sahagun, Diez argued that tomato was totally "integrated" in the Aztec civilization food consumption in XVI century, contrary to South American Incas (Diez and Nuez 2008). Nevertheless, two authors identified Quechua names possibly referring to tomato: "pirca" (Horkheimer 1973) and "pesco-tomate" (Yakovleff 1935). However, botanists consider the origins of tomato domestication as unsolved (Peralta and Spooner 2007b). These authors mention recent evidences showing that the Mexican hypothesis is not supported by comparative data, as South American and Mexican tomato accessions share similar isozymes (Rick and Fobes 1975) as well as molecular markers (Villand, Skroch et al. 1998). So far, no evidence appears to be enough conclusive and tomatoes may have been domesticated independently in both areas. To go further a more extensive analysis of molecular polymorphism in the wild and cultivated tomatoes is needed. This would allow investigating demographic scenarios and estimating the parameters of these scenarios (bottleneck intensity, ancestral population size, migration rates) using Markovian model implemented in tools such as IM* program (Hey and Nielsen 2004) or ABC[1] methodology (Beaumont, Zhang et al. 2002; Lopes and Beaumont 2010). Very recently, this approach has been implemented to infer past demography and ecological parameters of two tomato wild relatives, *S. chilense* and *S. peruvianum* (Tellier, Laurent et al. 2011).

Many authors consider *S. lycopersicum* var. *cerasiforme* as ancestral form of the cultivated tomato. It is present in both Mexico and Peru, on the contrary to *S. pimpinellifolium* which is absent from Mexico. If we assume that *S.l. cerasiforme* results from direct domestication from *S. pimpinellifolium*, a consequence of this domestication is that *S.l. cerasiforme* suffered a decrease of its population effective size during domestication (Bai and Lindhout 2007). Subsequent changes occurred for domestication traits such as growth habit, mating system, gigantism and fruit morphological diversity. Notably a change from exerted to inserted stigmas is responsible for the change from partial allogamy to strict autogamy. Selection for self-pollinating as well as shortening of the stigma compared to close wild relatives such as *S. pimpinellifolium* has allowed a yield increase (Rick 1977b). This "selfing syndrome" (Sicard and Lenhard 2011) is striking in tomato where a mutation in gene controlling stigma length has been identified in cultivated germplasm (Chen, Cong et al. 2007).

4. Early cultivation in Europe and in the world

Probably only a few tomato seeds were brought back from Mexico to Europe, leading, after domestication, to a new genetic bottleneck. George McCue has extensively reviewed the history of tomato diverse uses, tracking back the first references by country upon the time (McCue 1952). Most remote reference available comes from Petrus Matthiolus, an Italian Physician (1544). Due to its botanical closeness with toxic *Solanum* species common in Europe (Mandrake, Belladonna), tomato was for long mostly used as an ornamental. Two centuries later, it was referred as a cultivated plant in Italy by Saccardo (1769). Southern Europe was precursor in use of tomato for human consumption. In France, Bois at first mentions it as ornamental (1760). The same author reported it as vegetable seeds sold in the catalogue of the seeds of the "Maison grainière Andrieux Vilmorin" (1778). Lamarck mentioned it in 1798. Extensive consumption in Spain is described by Quer (1784).

[1]Approximate Bayesian Computation

Progressively, following South-North gradient, tomato consumption reached Northern Europe (Sabine 1819). Similarly in USA, Bartram (1766) reported tomatoes being used as food plants. Boyd (1784) mentioned that The David Landreth Seed Co. started to sell tomato seeds for vegetable consumption. Selection for diverse fruit shapes and local adaptation probably rapidly occurred through bulk selection. The crop gained in economic importance by the end of XIXth century with the establishment of tomato breeding programs. Most of the plant material at that time can be considered as landraces: selected for subsistence agriculture environments, producing low but relatively stable yield. At the end of the XIXth century, tomato cultivars (nowadays called landraces or heirlooms) were open pollinated from which seeds were saved by the farmers from a year to the other. Selection of new genotypes within heterogeneous cultivars (or selection of chance variance) resulted from spontaneous mutations, natural outcrossing or recombination of pre-existing genetic variation. Thus, *S. lycopersicum* found in Europe a secondary centre for diversification (García-Martínez, Andreani et al. 2006). In the XIXth century, establishment of commercial routes and colonies contributed to spread the species worldwide (Diez and Nuez 2008). In United States, prior to 1850 and "Trophy" the first commercially successful variety, no breeding programs were effective (Smith 2000). On an evolutionary perspective, domestication and implementation of breeding programs induced physiological changes. Artificial selection has reduced the genetic diversity of the crop which suffered a new bottleneck.

5. Tomato breeding in the XXth century: Seeking for diversity and intensive production

After domestication and adaptation to North hemisphere growing conditions for two centuries, the crop started the XXth century with benefits of two major scientific discoveries: The rediscovery of Mendel pioneering work to set up the basis of experimental methods on the use of plant hybridization (Mendel 1866). Second are established domestication concept (Darwin and Wallace 1858) and selection theory (Darwin 1859).

This context has seen the emergence in public institutes of plant germplasm banks, starting point for collecting existing genetic diversity, preserving and valorizing it, following the pioneer work of Nikolai Vavilov (1887-1943) (Kurlovich, Rep'ev et al. 2000). Later on, he was followed by Charles Rick (1915-2002) who dedicated his life to discover, collect and characterize exotic tomato germplasm (Tanksley and Khush 2002). Today, more than 83,000 tomato accessions are stored in seed banks worldwide, ranking 1st among vegetable species collected (FAO 2010). The main collections in the world are: In USA, the Tomato Genetic Resources Center in California (TGRC), (www.tgrc.ucdavis.edu) and the USDA[2] collection (www.ars.usda.gov), the World Vegetable Center in Taiwan (www.avrdc.org) and several Europeans collections. The establishment of tomato resource collections made major contributions to understand the distribution of its diversity around the world. Nevertheless the lack of coordination and conflicting passport data is a pitfall for an efficient tomato germplasm management. Efforts are now made to coordinate national initiatives in global or regional approaches. Since 2007, The European Cooperative Programme for Plant Genetic Resources (www.ecpgr.cgiar.org) is a collaborative project between most European countries for long-term conservation and utilization of plant genetic resources in Europe.

[2] United States Department of Agriculture, Geneva

This project is based on large network of national centers for tomato genetic resources including COMAV[3] (Spain), CGN[4] (Netherlands), INRA (France), IPK[5] (Germany), Vavilov Institute (Russia) and others. These institutions share their germplasm informations through a database (http://documents.plant.wur.nl/cgn/pgr/tomato/). More recently, in the context of a European Solanaceae project (EU-SOL, www.eu-sol.wur.nl), a collection of more than 6,000 domesticated tomato accessions was established and phenotyped, accompanied by an *ad hoc* database (Finkers, de Weerd et al. 2011). Finally, since 2008, a world initiative, is conducted under the Plant Biodiversity Inventories (www.nhm.ac.uk/research-curation/projects/solanaceaesource/). Project aim is to produce a worldwide taxonomic monograph of the species occurring within the plant genus *Solanum.* As well, tomato is part of long term collection of plant species project, launched by the Svalbard Global Seed Vault initiative (Food 2008).

Due to its broad use for food consumption and adaptation to many environmental conditions (from Alaska summers to tropical conditions) and different crop systems, tomato experienced an important phenotypic diversification. Hundreds of past and present cultivated varieties are now available. Cultivars are dedicated to two main markets, processing and fresh market. Processing tomatoes are cultivated as a field crop, whereas fresh market tomatoes are grown outdoor or indoor (heated and non-heated greenhouses). Breeding objectives have evolved over time, with the evolution of production systems. Nevertheless, three main objectives remain: adaptation to growth constraints, disease and pest resistances and fruit productivity and quality. Wild species were first used as source of adaptation to biotic stress. Disease resistance selection started in United States early XXth century. The first *fusarium* wilt –resistant cultivar "Tennessee red" was released in 1912. Early 1920's breeders used hybridization with selection in segregating generations. By the mid-30's, plant breeders developed technical procedures to improve selection, such as pedigree selection. Later, existing or emerging private companies enhanced their development with the release of F1 hybrid varieties. Selection for disease resistance was successful as dominant resistance genes were found in the wild relatives for most of the diseases and pests. Modern cultivars can cumulate up to 12 different disease resistance genes which all derive from wild species. Wild germplasm has been primarily used as a source of major resistances.

Processing tomato industry was developed to provide North American and European households canned tomato, tomato paste and ketchup. Processing tomato varieties differ from fresh market ones in their pulp volume. Their growing conditions are dramatically different from fresh tomatoes (open field, mechanical harvesting). Thus the main criteria for processing tomatoes are fruit firmness, plant type with short fruit set period to produce a high percentage of ripe fruits simultaneously. Compact fruit set was obtained from a natural mutation discovered at the beginning of the XXth century, named *sp* (for self-pruning), conferring to the plant a determinate growth. This mutation was introduced in the well known "Roma" variety, whose long fruit type became a specific trait in processing tomatoes. In the 60's, VF145 was the first variety mechanically harvested. This cultivar has been the major cultivar for tomato ketchup industry for more than a decade in California. Apart from *sp*, several other mutations (detected in cultivated tomato or through interspecific

[3] Centro de Conservación y Mejora de la Agrodiversidad Valenciana, Valencia
[4] Center for Genetic resources, Wageningen
[5] Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben

hybridization) were used in tomato breeding. For example, the jointless (Szymkowiak and Irish 2005; Quinet, Kinet et al. 2011) *j2* allele was introgressed from *S. Cheesmani*i. In combination with *sp*, *j2* has been bred into many processing varieties, allowing a large scale mechanical harvest of tomato fruits. Major genes used in tomato breeding are listed in Table2.

Today, after a rapid evolution towards very firm fruits and long shelf life varieties (with the major success of the variety *Daniela*, which carried the spontaneous *rin* mutation), consumers request more diverse texture and tastes (Causse, Friguet et al. 2010). The fresh tomato market faces rapid developments and diversification (Navez 2011). New products and varieties are emerging always faster and their life cycle gets shorter, 5 years in average (Bai and Lindhout 2007). Consumption trend is for broader and diverse choice of fruit types. After the development of truss and cherry type tomatoes, new cultivars resembling to old heirloom varieties are developed. If access to allelic diversity is a must to improve fruit quality, choices in breeding objectives are critical to maintain organoleptic fruit values. Strong associations are often made by consumers between morphology and sensory values. Association of "Oxheart" or "Marmande" fruit shape with a pleasant texture is a good example. This link can be lost through modern breeding (Casals, Pascual et al. 2011). Improved content in potential health beneficial components such as anti-oxidants (lycopene, vitamin C) is also promoted. This can be obtained thanks to specific mutations like *hp* (Lieberman, Segev et al. 2004) or, again, by the introgression of genes from wild relatives.

Farmers and breeders have shaped diversity over years in an ever-evolving process that is hard to track and to record. Intensive breeding of crop varieties by modern science has increased the genetic erosion which started with domestication. Nevertheless the introgression from wild relatives allowed major progress and introduced a new source of diversity. Charles Rick observed that crosses between wild and cultivated species generated a large diversity of novel phenotypic diversity. Rick's work represents milestones for the modern use of genetic diversity in tomato. It led to uncover positive transgressive variation within interspecific progenies. This encouraged a greater use of exotic germplasm and thus larger gene pools to unlock causal polymorphism.

6. Biotechnology as a source of new diversity

Many natural mutants discovered in cultivated tomato have been extensively studied and characterized (http://tgrc.ucdavis.edu/), but their amount is limited. Thus reverse genetic techniques were developed aiming to discover gene function by analyzing the phenotypic effects of specific variants of targeted gene sequence. This approach is complementary to classical ("forwards") approaches, as they allow silencing or promoting the expression of targeted gene. They can also be used to generate genetic diversity within DNA sequences.

A delayed ripening tomato, named *Flavr-savr*® tomato, with a reduced expression of a cell wall protein (a polygalacturonase), obtained by an antisense construction, was the first transgenic plant released on the fresh market (Kramer and Redenbaugh 1994; Sanders and Hiatt 2005). This transformation significantly improved fruit shelf life and storage quality. Nevertheless, it was a commercial failure. Few years later, consumer's concerns about GMO[6],

[6] Genetically Modified Organism

Gene Name	Plant trait	Gene product	Phenotype	Chromosome(s)	Germplasm source	Reference
*Ae**	abiotic stress	bHLH protein	Iron-uptake response in roots	6	*S. lycopersicum*	Ling et al. (2002)
*chlorosis**	abiotic stress	Nicotianamine synthase	Iron uptake	1	*S. pennellii*	Ling et al. (1999)
*asc**	biotic stress	Novel protein of Longevity Assurance Gene (LAG1) family	*Alternaria alternata* (stem canker) resistance	3	*S. lycopersicum*	Brandwagt et al. (2000)
*Ve1**	biotic stress	Tdl/intolerulin-1-receptor-nucleotide-binding-leucine-rich-repeat protein (TIR–NB	*Xanthomonas campestris* pv. vesicatoria resistance	3	*S. pennellii*	Balbiros et al. (2001); Schornack et al. (2004)
*I-3**	biotic stress	Tdl/intolerulin-1-receptor-nucleotide- binding site-leucine-rich-repeat (TIR–NBS–	*Fusarium oxysporum f. sp. lycopersici* resistance	7	*S. pennellii*	Hamming et al. (2004)
*mi-3**	biotic stress	Nucleotide-binding site-leucine-rich-repeat protein (NBS–LRR)	GNBV, GISV, TCSV	9	*S. peruvianum*	Brommonschenkel et al. (2000)
*mi-1**	biotic stress	Coiled-coil-nucleotide-binding site- leucine-rich repeat protein (CC–NBS–LRR)	Root-knot nematode resistance	6	*S. peruvianum*	Zhong et al. (1999)
*tmv-1, tmv-2**	biotic stress	Unknown	Tobacco mosaic virus resistance	2, 9	*S. peruvianum*	Lanfermeijer et al. (2005)
*pto**	biotic stress	Serine threonine protein kinase	*Pseudomonas syringae* pv. tomato resistance	5	*S. pimpinellifolium*	Martin et al. (1993)
*hero**	biotic stress	Protein (NBS–LRR) Nucleotide-binding-site-leucine-rich repeat	Nematode resistance	4	*S. pimpinellifolium*	Ernst et al.(2002)
*cf-2**	biotic stress	Leucine-rich repeat protein	Cladosporium fulvum resistance	6	*S. pimpinellifolium*	Dixon et al. (1996)
*I-2**	biotic stress	Coiled-coil-nucleotide-binding site-leucine-rich repeat protein (CC–NBS–LRR)	*Fusarium oxysporum f. sp. lycopersici* resistance	11	*S. pimpinellifolium*	Ori et al. (1997)
ph-2, ph-3	biotic stress	Unknown	Phytophthora infestance resistance	9, 10	*S. pimpinellifolium*	Monsan et al. (1998) Chunwongse et al. (1998)
Sy-1	biotic stress	Unknown	Tomato yellow leaf curl virus resistance	6	*S.chilense, S. habrochaite*	Zamir (1994)
style2.1	flower development	Transcription factor regulating cell elongation	Flower style lenght	2	*S. lycopersicum*	Chen et al. (2007)
S	flower development	S-ribonucleases (pistil), S-locus F-box (SLF) proteins (pollen)	Control unilateral interspecific incompatibility	2	*S. pennellii*	Li et al. (2010)
hp1, hp2*	fruit color	Damaged DNA binding protein 1 (DDB1)	High pigment, immature fruit dark green	1,2	*S. lycopersicum*	Yen et al. (1997); Liebesman et al. (2004)
*tangerine**	fruit color	CRTISO, Carotenoid isomerase	Carotenoid desaturation, orange fruit	10	*S. lycopersicum*	Isaacson et al. (2002)
cr	fruit color	Chromoplast-specific lycopene beta cyclase Cyc-B	Crimson, increase lycopene	6	*S. lycopersicum*	Ronen et al. (2000)
*beta old-gold**	fruit color	Lycopene -cyclase	Carotene synthesis lycopene increase, orange fru	6	*S. pennellii*	Ronen et al. (2000)
fgr	fruit content	Unknown	Increase fructose level	4	*S. habrochaite*	Levin et al. (2000)
*Inv-5-2, 5**	fruit content	Apoplastic invertase (LIN5)	Increases sugar content and tomato yield	9	*S. pennellii*	Fridman et al. (2004)
*cnr**	fruit development	SQUAMOSA promoter binding protein (SBP) box transcription factor	Colorless non-ripening mutant	2	*S. chmielewskii*	Manning et al. (2006)
*Gr/Nr-2**	fruit development	Novel protein, block ethylene perception	Green ripe/never ripe mutant, canker tums red	1	*S. lycopersicum*	Barry and Giovannoni (2006)
rin1	fruit development	PLENA subfamily of MADS-BOX genes	Yellow-orange fruit, reduced carotenoids	7	*S. lycopersicum*	Mão et al. (2009)
*rin**	fruit maturity	MADS-box transcription factor	Ripening inhibitor	5	*S. chmielewskii, S. pennellii*	Vrebalov et al. (2002)
nor	fruit maturity	MADS-box transcription factor	Ripening inhibitor	10	*S. lycopersicum*	Lanahan et al. (1994)
nr	fruit maturity	Unknown	Ripening inhibitor	9	*S. lycopersicum*	Yen et al. (1995)
*ovate**	fruit shape	Novel protein with bipartite nuclear localization signal	Growth repressor, Pear shaped fruit	2	*S. lycopersicum*	Ku et al. (1999); Liu, J. et al. (2002)
*sun**	fruit shape	IQ67 domain-containing protein	Elongated fruit shape	7	*S. lycopersicum*	van der Knaap et al. (2004); Xiao et al. (2008)
*lcn**	fruit size	3' end of WUSCHEL (homeodomain protein)	Fruit size and locule number increase	2	*S. lycopersicum*	Munos et al. (2011)
*fas**	fruit size	YABBY like transcription factor	Fruit size and locule number increase	11	*S. lycopersicum*	Cong et al. (2008)
*fw2.2**	fruit size	CNR, similar to human oncogene c-H-ras p21	Major fruit weight locus	2	*S. pennellii*	Alpert and Tanksley (1996); Frary (2000)
*fs8.1**	fruit size	Unknown	Fruit size, Imparts blocky, elongated shape	8	*S. pimpinellifolium*	Ku et al. (2000)
*Clv**	leaf development	Unknown	Cyngate leaf veins	5	*S. pennellii*	Jones et al. (2007)
*lr**	plant development	VHIID protein	Lateral suppressor	7	*S. lycopersicum*	Schumacher et al. (1999)
*sft**	plant development	Flotigen precursor	Regulates transition and maintenance of flowers	3	*S. lycopersicum*	Molinero-Rosales et al. (2004)
ms.s	plant development	F-box protein involved in transcriptional co-activation with the transcription factor	Inflorescence branching and floral identity	2	*S. lycopersicum*	Lippman et al. (2008)
*j,j2**	plant development	MADS-box transcription factor	Jointless, flower abscission zone development	11,12	*S. lycopersicum, S. chmielewskii*	Zhang et al. (1994); Mao et al. (2000)
*sp**	plant development	Ortholog of CENTRORADIALIS and TERMINAL FLOWER1	Self pruning	6	*S. pennellii*	Pnueli et al. (1998)
*sw4.1**	seed development	ABC transporter	Seed weight	4	*S. lycopersicum S. pimpinellifolium*	Orsi et al. (2009)

* cloned genes

Table 2. List of genes characterized through molecular techniques with their related function and germplasm origin

as well as the high engineering cost, stopped further commercial developments. In the research field, transformation with *Agrobacterium thumefaciens* is still widely used for the functional characterization of specific genes. For instance, transformed tomato plants were produced to enable study of endotoxins genes (Zhang, Buehner et al. 2006) plant disease resistance genes, abiotic stress genes or to produce molecules useful in human medicine (Sharma, Singh et al. 2008).

TILLING (Target Induced Local Lesion In Genomes), a mutagenesis technique, has experienced important development (Comai and Henikoff 2006). Early days of this technique were in the 50's (Rick 1991). It is now widely used for reverse genetics to generate and identify induced point mutations in genomes. A chemical reagent (Ethylmethane Sulphonate) is used to induce genetic mutations. Collections of tomatoes carrying artificially induced genetic variants, called mutant libraries are currently available (Menda, Semel et al. 2004; Minoia et al. 2010) or under development (Okabe, Asamizu et al. 2011). In contrast to transgenic methods, mutagenesis is random, cost effective and is not submitted to GMO regulation. TILLING allows generating variants in cultivated genetic background (Piron, Nicolaï et al. 2010) and thus transfer rapidly interesting mutations into cultivars (Gady, Hermans et al. 2009). Application of TILLING technique to screen for natural variation within tomato germplasm collection is now performed (Rigola, van Oeveren et al. 2009).

7. Molecular markers offer a new vision of tomato diversity

Natural genetic diversity is the fuel of evolution. No evolutive forces or adaptation to environment changes can apply without it (Alonso-Blanco, Aarts et al. 2009). Consequently it is a vital characteristic for species adaptation in general and for crop breeding in particular. Genetic variation occurs both within cultivated tomato (intraspecific) and between wild species (interspecific). Tomato breeding for adaptation to specific growing areas is in progress for more than two centuries now (Stevens and Rick 1986). Since the early days of quantitative genetics, initiatives were developed to improve the understanding of trait inheritance. Attempts to construct genetic maps based on interspecific crosses (*S. pimpinellifolium* x *S. lycoperpersicum*) and to map disease resistance genes are performed for years (Langford 1937). A linkage map showing the distribution of agronomic trait with Mendelian inheritance, based on linkage between two or three mutations, was proposed (Butler 1952). Nevertheless, the lack of polymorphic and neutral markers was strongly limitant. Development of isozymes allowed a first evaluation of wild germplasm (Rick and Fobes 1975) and introgression diagnostic (Tanksley, Medina-Filho et al. 1981), but isozyme marker scarcity and their low polymorphism was still limitant. This limitation was progressively overcome since the 80's thanks to the discovery of several molecular marker types.

7.1 Ecological and evolution in wild tomato related species

Molecular studies provide important clues into ecological and evolutionary questions in wild tomatoes species. In speciation process, hybrid sterility is frequently due to dysfunctional interactions between loci that accumulate between different lineages. A "snowballing effect" characterizes loci controlling such reproductive barrier and hybrid sterility that should accumulate faster than linearly with time. Such "snowballing" effect has been recently described within distinct populations derived from crosses of *S. lycopersicum* with *S. pennellii*,

S. habrochaites and *S. lycopersicoides* (Moyle and Nakazato 2010). However, further investigations are suggested to confirm these results (Stadler, Florez-Rueda et al. 2011).

Tellier and colleagues quantified the number of adaptive and deleterious mutations and the distribution of fitness effects of new mutations within housekeeping genes in 4 species, *S. arcanum, S. chilense, S. habrochaites* and *S. peruvianum*. Little evidence for adaptive mutations was shown but strong purifying selection in coding regions was detected (Tellier, Fischer et al. 2011). This suggests that closely related species with similar genetic backgrounds but contrasted environments differ in the frequency of deleterious fitness effects.

The west coastal area between the Andes and the ocean, from Ecuador to Chile is widely recognized as the center of origin of the species from the Solanum sect. *lycopersicon*. This area covers a wide range of geographical conditions. Complex geography and ecology of Andes had a major impact in species divergence and hybridization between *S. pimpinellifolium* and *S. lycopersicum* (Nakazato and Housworth 2011). The two species present a distinct lineage, separated by the Andes. They hybridize extensively in north and central Ecuador. Nakazato and colleagues demonstrated using molecular markers and geographic information system (GIS) data that *S. lycopersicum* has likely experienced a severe population bottleneck during the colonization of the eastern Andes followed by a rapid population expansion. In plant, resistance genes and homologs (RGA) tend to be highly variable. Caicedo et al (2004) studied the geographic distribution of a RGA family Cf-2 (see Table 2) within and among plant populations of *S. pimpinellifolium*. They underlined that the geographical distribution of RGA diversity has been primarily shaped by demographic factors and selective pressure (Caicedo and Schaal 2004; Caicedo 2008). The authors underlined the reduction of natural habitat. This phenomenon is also observed on Galapagos Islands. The endemic species *S. cheesmanii* shows a reduction of its population due to human activity. Differentiation within *S. cheesmanii* was also observed (Nuez, Prohens et al. 2004) as well as hybridization with the two introduced species *S. lycopersicum* and *S. pimpinellifolium* (Darwin, Knapp et al. 2003).

7.2 Diversity analysis among wild and cultivated germplasm

Allelic richness (number of different alleles segregating in the population) is used to measure the genetic diversity and is considered as a key parameter for genetic resources management. It reveals past fluctuations in population size (Nei, Maruyama et al. 1975). Molecular differences between more than 200 Peruvian and Ecuadorian *S. pimpinellifolium* accessions were highlighted by Zuriaga and colleagues. Climate and genetic data were highly correlated. Thus the non-uniform nature of climates between the two countries is shown to be an important factor. Highest diversity was found in North Peru, lowest on Galapagos Islands. Authors stressed the fact that interspecific variation between *S. pimpinellifolium* and *S. lycopersicum* was indicating a very close relatedness between the two species (Zuriaga, Blanca et al. 2009).

Cherry tomato accessions show typically a large genetic diversity and an intermediate fruit size *between S. pimpinellifolium* and large cultivated ones. Botanists postulate that cherry tomato accessions are feral plants (also called revertant) or a possible genetic admixture of wild and cultivated germplasm (Rick and Holle 1990; Peralta et al. 2007a). Recently molecular analysis of the structure of a large set of accessions of wild *S. pimpinellifolium*, cherry tomato and cultivated accessions showed that domesticated and wild tomatoes have

evolved as a species complex with intensive hybridization. This highlighted the admixture position of *S. lycopersicum* var. *cerasiforme* (Ranc, Muños et al. 2008) which is illustrated on figure 2 using a data from Ranc et al (2010) and analyzed using Structure 2.0 (Pritchard et al. 2000) output data. Accessions display clustering patterns (circled) following two phenotypic traits: fruit size and stigma insertion. Structuration effect of those domestication traits can be observed. The emergence of molecular markers has allowed quantifying with accuracy the diversity within germplasm material. The first molecular diversity studies on cultivated tomato revealed the very low polymorphism compared to wild species, whether it was based on RFLP[7] (Miller and Tanksley 1990), SSR[8] (Bredemeijer, Cooke et al. 2002; He, Poysa et al. 2003) AFLP[9] (Park, West et al. 2004; Berloo, Zhu et al. 2008), SSAP[10] (Tam, Mhiri et al. 2005) or SNP[11] (Yang, Bai et al. 2004; Labate and Baldo 2005). However, Bredemeijer et al (2002) characterized 500 cultivated lines from European lines and showed that it was possible to distinguish them all from each other using a set of 20 SSR markers. When comparing old varieties (or landraces) to modern hybrids, a higher level of molecular diversity in landraces is usually observed (Mazzucato, Papa et al. 2008; van Berloo, Zhu et al. 2008).

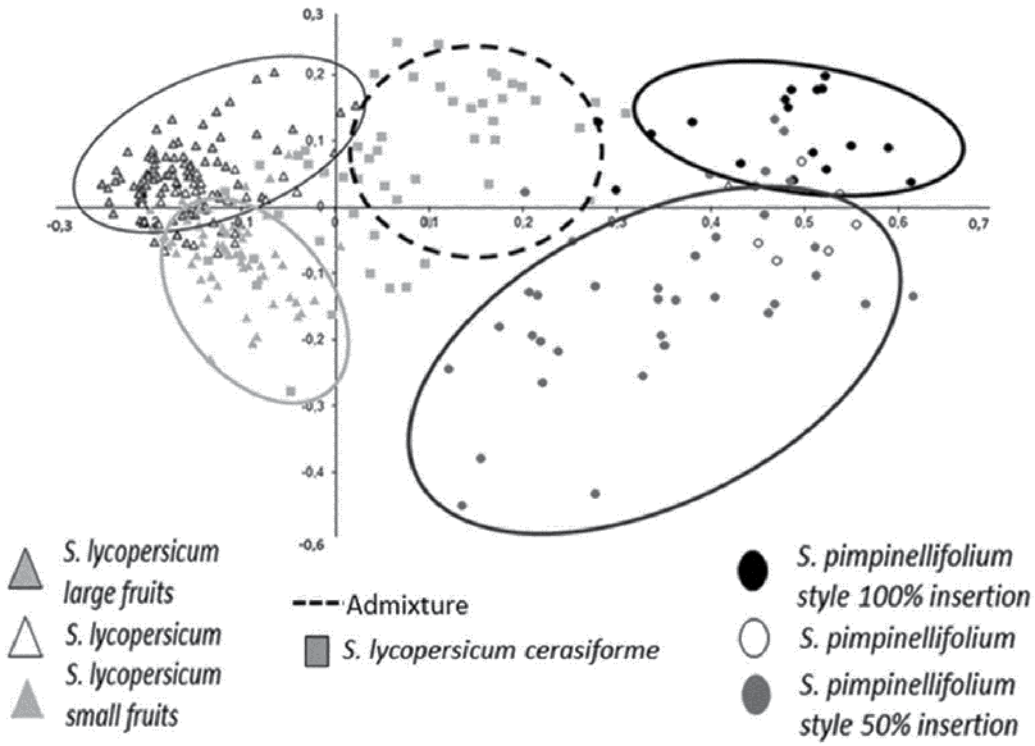


Fig. 2. Principal Coordinate Analysis of 318 accessions tomato core collection.

[7] Restriction Fragment Lenght Polymorphism
[8] Simple Sequence Repeats
[9] Amplified Fragment Lenght Polymorphism
[10] Sequence-Specific Amplification Polymorphism
[11] Single Nucleotide Polymorphism

If interspecific populations for genetic analyses and diversity studies answered to many questions, it has left a void in the understanding of genotypic variation within tomato breeding programs which focus on intra-specific populations (Van Deynze, Stoffel et al. 2007). The recent discovery of SNP markers, first detected in EST (expressed sequenced tag) sequences (Van Deynze et al. 2007; Jimenez-Gomez and Maloof 2009) then in non-coding sequences (Labate et al. 2009) provided access to a higher level of polymorphism. Labate and colleagues estimated parameters of diversity among *S. lycopersium* accessions, first using the SNP detected in 50 loci that were resequenced in a diversity panel of 31 accessions. In a second investigation, multilocus estimates of polymorphism were obtained and led to rejection of the neutral equilibrium model of evolution within the studied collection (Labate, Robertson et al. 2009). Public germplasm are potential allele mining sources for crop improvement as illustrated by previous authors who sampled among US seed banks 30 accessions from the five continents. The study confirmed that history of crossing with wild tomato species and distribution among different environments across the world has spread allelic variation (Labate, Sheffer et al. 2011).

Molecular markers have proven their efficiency in sampling and maximizing allelic richness (Schoen and Brown 1993) through the development of nested core collections (McKhann, Camilleri et al. 2004). Such nested core collections (from 8 to 96 accessions) were constructed in tomato, capturing most of the molecular and phenotypic variation present in a set of 360 constituted of wild, feral and cultivated accessions (Ranc, Muños et al. 2008).

7.3 Use of molecular diversity to dissect phenotypes

Molecular markers allowed the construction of high density genetic maps of the tomato genome (Tanksley, Ganal et al. 1992). This permitted the dissection of quantitative traits into Mendelian factors or QTL (Quantitative Trait Loci) (Paterson, Lander et al. 1988; Tanksley et al. 1992). This strategy also opened the way to investigate physical mapping and molecular cloning of genetic factors underlying quantitative traits (Paterson, Damon et al. 1991). Moreover, *Lycopersicon* varieties and related species are all diploid and chromosomally collinear, making genetic dissection straightforward. The first gene cloned by positional cloning was the *Pto* gene, confering resistance to *Pseudomonas syringae* (Martin, Brommonschenkel et al. 1993). Since then, interspecific crosses with each wild species were performed. Due to the low genetic diversity within the cultivated compartment (Miller and Tanksley 1990), most of the mapping populations are based on interspecific crosses between a cultivar and related wild species from the lycopersicon group (as reviewed by Foolad (2007) ;Labate, Grandillo et al.(2007); Grandillo et al (2011)) or from lycopersicoides (Pertuzé, Ji et al. 2002) and juglandifolia group (Albrecht, Escobar et al. 2010). However, maps based on intraspecific crosses have proved their interest notably on fruit quality aspects (Saliba-Colombani, Causse et al. 2001). All those populations allowed discovering and/or characterizing a myriad of major genes (Table 2) and QTLs involved in various traits.

Rapidly, molecular breeding strategies were set up and implemented to "pyramid" genes of interest for agronomical traits, notably using Advanced Backcross QTL method (AB-QTL) (Tanksley, Grandillo et al. 1996). Using this approach with a *S. lycopersicum x S. pimpinellifolium* progeny, in which agronomical favorable QTL alleles were detected, Grandillo and colleagues showed how a wild species could contribute to improve

cultivated tomato (Tanksley, Grandillo et al., 1996). Introgression Lines (IL) derived from interspecific crosses allowed to dissect the effect of chromosome fragments from a donor (usually from a wild relative) introgressed into a recurrent elite line. IL offer the possibility to evaluate the agronomic performance of a specific set of QTL (Paran, Goldman et al. 1995). IL was used as a base for fine mapping and positional cloning of several genes and QTL of interest. The first IL library was developed between *S. pennellii* and *S. lycopersicum* (Eshed and Zamir 1995; Zamir 2001). QTL mapping power was increased compared to biallelic QTL mapping population, and was again improved by the constitution of sub-IL set with smaller introgressed fragments. This progeny was successful in identifying QTLs for fruit traits (Causse, Duffe et al 2004); anti-oxidants (Rousseaux, Jones et al. 2005), vitamin C (Stevens, Buret et al. 2007) and volatile aromas (Tadmor, Fridman et al. 2002). The introgression of a QTL identified in these IL has allowed plant breeders to boost the level of soluble solids (brix) in commercial varieties and largely increased tomato yield in California (Fridman, Carrari et al. 2004). Such exotic libraries were thus designed with several species, involving *S. pimpinellifolium* (Doganlar, Frary et al. 2002), *S. habrochaites* (Monforte and Tanksley 2000; Finkers, van Heusden et al. 2007) and *S. lycopersicoides* (Canady, Meglic et al. 2005).

Introgression lines were also used to dissect the genetic basis of heterosis (Eshed and Zamir 1995). Heterosis refers to phenomenon where hybrids between distant varieties or crosses between related species exhibit greater biomass, speed of development, and fertility than both parents (Birchler, Yao et al. 2010). Heterosis involves genome–wide dominance complementation and inheritance model such as locus–specific overdominance (Lippman and Zamir 2007). Heterotic QTL for several trait were identified in tomato IL (Semel and Nissenbaum, 2006). A unique QTL was shown to display at the heterozygous level improved harvest index, earliness and metabolite content (sugars and amino acids) in processing tomatoes (Gur, Osorio et al. 2010; Gur, Semel et al. 2011) Furthermore, a natural mutation in the SFT gene, involved in flowering (Shalit, Rozman et al. 2009), was shown to correspond to a single overdominant gene increasing yield in hybrids of processing tomato (Krieger, Lippman et al. 2010).

Metabolite detection is an approach of choice to identify compounds involved in fruits quality traits. Metabolite QTL (mQTL) can be now identified for non-volatile metabolites like sugars, pigments or volatiles compounds (Bovy, Schijlen et al. 2007). This was done on several interspecific populations, notably on *S. lycopersicum x S. Chmielewskii* (Do, Prudent et al. 2010) and intraspecific crosses (Saliba-Colombani et al. 2001; Causse, Saliba-Colombani et al. 2002; Zanor, Rambla et al. 2009). Recent technologies allowed screening for diversity in a wide range of components on whole genomes. This can be done in a targeted way to better characterize known metabolites (Tieman, Taylor et al. 2006) or untargeted manner to identify new metabolites (Tikunov, Lommen et al., 2005). Further than identify and quantify compounds, metabolomics can be of great help to decipher biosynthetic pathways (Keurentjes 2009). Metabolome studies can be combined to transcriptomic data to identify the key factors (Mounet, Moing et al. 2009; Do, Prudent et al. 2010). Metabolomics has an important role to play in characterization of natural diversity in tomato (Schauer, Zamir et al. 2004; Fernie et al. 2011). As well, it can boost the biochemical understanding of fruit content and be an enhancer for quality breeding (Fernie and Schauer 2009; de Vos, Hall et al. 2011).

7.4 Dissection of the molecular bases of domestication and diversification

Product of human domestication and later diversification of fruit types, led to a large morphological diversity in tomato fruit (with small to large, round, blocky, elongated, pear shaped fruits, with color ranging from red to green, white, black, pink, orange or yellow). On the contrary, wild tomato species carry small, round red or green fruits, with a low intraspecific phenotypic diversity. This has drawn scientist attention on the inheritance and development of fruit size and shape in the tomato (Yeager 1937). Influence of chromosome 2 in fruit morphology was noticed (Butler 1964). Thus, using available molecular techniques, fruit traits genetic control has been widely dissected (Grandillo, Ku et al. 1999; Lippman and Tanksley 2001; Barrero and Tanksley 2004). The first QTL, fw2.2, controlling fruit weight variation was cloned (Frary, Nesbitt et al 2000). It has been suggested that diversity of fruit shape in cultivated germplasm can be explained to a large extent by four genes (Rodriguez, Muños et al. 2011). The study established a model for fruit shape evolution in tomato. This model includes four major mutations recently identified: FAS which increases locule number, fruit fasciation and size (Cong, Barrero et al. 2008), LC which increases locule number and fruit size (Muños, Ranc et al. 2011), OVATE which gives ovoid fruit shape (Liu, van Eck et al 2002) and SUN which gives an elongated fruit shape (van der Knaap, Lippman et al. 2002; Xiao, Jiang et al. 2008) or the oxheart shape when associated to LC and FAS. The allelic distribution of the four genes was associated with morphologic, geographical and historical data in a collection of diverse cultivated accessions. This study established that the selection occurred in distinct chronologic and historic periods: LC arose first, followed by OVATE, both in *S.l. cerasiforme* background but in distinct populations. FAS arose later in a LC background. Presence of those three mutations in Latin American germplasm suggests Pre-Columbian mutations. Combined with fw2.2, they must have strongly contributed to the increase in fruit size during tomato domestication. On the contrary, SUN mutation is not carried by any Latin American material tested, suggesting that SUN mutation appeared post domestication in European material (probably in Italy). This study also showed that the selection for fruit shape is strongly responsible for the underlying genetic structure in tomato cultivars. The recent discoveries of the molecular events shaping tomato fruit indicate that the germplasm is frequently more diverse phenotypically than the wild related germplasm but not necessarily showing a similar pattern at the molecular level. *"The irony of all this,"* says Steve Tanksley (geneticist at Cornell University, and precursor of all these studies) *"is all that diversity of heirlooms can be accounted for by a handful of genes. There are probably no more than 10 mutant genes that create the diversity of heirlooms you see"* (Borrell 2009). Tomato selection and spread worldwide has led to the immense diversity of varieties that characterizes many domesticated plant species (Purugganan and Fuller 2009).

8. Association genetics: New valorization of natural diversity

Recent advance in molecular genetics and computation has allowed the emergence of association mapping (Myles, Peiffer et al. 2009). Association mapping takes advantage of historical recombination events and natural genetic diversity. By using large numbers of lines and molecular markers over the whole genome, the resolution of Genome Wide Association studies (GWAS) is much higher than in conventional segregating populations. Such approach requires an accurate estimate of the genetic structure of the sample studied (Price, Zaitlen et al. 2010) and linkage disequilibrium (LD) extend among loci. Yu and

colleagues (2005) proposed a unified mixed model taking into account the genetic structure of the sample, based on single locus analysis. This model is being updated by integrating a multi-locus analysis (Ayers et al. 2010).In autogamous crops, it is expected that large extent of LD will reduce the resolution and risks to lead to false positive associations. Nevertheless, successful results have been obtained in selfing crops (Atwell, Huang et al. 2010; Ramsay, Comadran et al. 2011).

In tomato, several studies revealed contrasted results according to the samples studied. First studies of the linkage disequilibrium revealed large LD in cultivated tomatoes (Mazzucato, Papa et al, 2008; van Berloo, Zhu et al, 2008; Robbins, Sim et al, 2010). Van Berloo and colleagues performed association mapping within a collection of 94 accessions containing both old and elite (hybrids) European germplasm and about 300 markers (AFLP). Structure coinciding with fruit size was identified allowing grouping between cherry tomato and round-beef types, extensive LD was observed (15 cM average). Robbins and colleagues investigated the population structure among 70 tomato cultivars (modern and vintage, from fresh and processing market). The STRUCTURE analysis (Pritchard, Stephens et al. 2000) revealed groups predefined by market niche and age into distinct subpopulations. Furthermore, they detected two subpopulations within the processing varieties, corresponding to historical patterns of breeding conducted for specific production environments. They found no subpopulation within fresh-market varieties. High levels of admixture were shown in several varieties representing a transition in the demarcation between processing and fresh-market. Mapping and LD analysis on a genome wide level was performed (Robbins, Sim et al. 2010). Using a panel of 102 accessions including 95 cultivars (heirloom, fresh and processing cultivars) and 9 wild species), effect of selection on genome variation was studied using 340 markers (SNP, SSR, and INDEL[12]). LD value varied from 6-8 cM (all accessions) up to 3-16 cM (fresh market cultivars). Inter-chromosomal LD appeared to be population dependent, suggesting cautious approach for association mapping. Notably, a genetic divergence between fresh market and processing types was also shown. On the contrary, the use of cherry tomato allowed the construction of core collection with a reduced structure and lower LD (Ranc et al, 2008; 2010). In a pilot study on chromosome 2, using markers distant from several cM to few kb, Ranc (2010) showed that LD varied strongly from one region to the other. A few distant markers remained in strong LD, but could be removed from the analysis.

The first association study was performed by Nesbitt and Tanksley (2002) to identify the SNP responsible for FW2.2 gene they had cloned. They failed to find any association between fruit size and genomic sequence of the *fw2.2* region in a collection of 39 cherry tomato accessions. Ranc and colleagues (2010) identified significant association in the promoter region, thanks to a larger and more representative sample. From a breeding point of view, the admixture mapping between the cultivated tomato and its closest relative is a method of choice for allele mining in wild germplasm. Muños and colleagues (2011) used this approach to identify causal polymorphism of QTL controlling locule number on chromosome 2. New SNP arrays are now available thanks to Next Generation Sequencing technologies (NGS), as the genotyping array developed under the Solanaceae Coordinated Agriculture Project (SolCAP) initiative carrying 7,000 effective SNP (SolCAP 2008). These tools will be very useful to scan the whole genome for associations.

[12] Insertion-Deletion

9. Conclusion: Toward a change in the way to manage and use diversity

Crossing wild and cultivated species can reveal alleles left behind during the domestication process. Molecular markers strongly helped to reinforce the use of wild relatives (Zamir 2008). Interfacing genetic resources management and plant breeding, pre-breeding is now recognized as an important adjunct to plant breeding, as a way to introduce new traits from non-adapted populations and wild relatives, notably for abiotic stress (FAO 2010). Nevertheless, the extensive use of this genetic richness contained in seed banks and germplasm collection faces limits. The difficulty to introgress accurately the targeted allele (with favorable effect) without unfavorable ones, carried on by "linkage drag", remains.

With the emergence of bioinformatics and nanotechnologies -so called "post-genomics" era- the last decade has opened high throughput sequencing era. Now, conducting large intra-specific studies becomes a reality in tomato, allowing a better characterization of its genetic diversity. With the completion of its genome sequence (Mueller, Lankhorst et al. 2009) and rich annotation as well as a large number of tools available via SGN (SOL Genome Network; http://solgenomics.net/organism/solanum_lycopersicum/genome) platform (Bombarely, Menda et al. 2010), tomato and its relatives is the most advanced vegetable crop. A draft of the genome sequence of *S. pimpinellifolium* LA1589 is also released by D. Ware, W. R. McCombie, and Z. B. Lippman at Cold Spring Harbor Laboratory allowing a detailed comparison of both species. The genome sequences of tomato provide clues for understanding the Solanum clade evolutive history and identify genes involved in fleshy fruit development.

Progress in sequencing technologies has reached the point where genotyping by sequencing (GBS) is now possible (Davey, Hohenlohe et al. 2011; Elshire, Glaubitz et al. 2011). This opens new perspectives in terms of genetic diversity management, notably toward conservation and survey of large populations. In a near future, techniques such as GBS may allow breeders and scientists of the tomato community to determine population characteristics prior concretely establishing genome or nucleotide diversity. GBS opens ways to a global and quantitative management of diversity, and let foresee an *a priori* genetic resource management. It also opens perspectives in allele based breeding called genomic selection (Hamblin, Buckler et al. 2011).

If *ex situ* germplasm conservation is well developed and will benefit of these developments, *in situ* conservation of tomato and its wild relatives is becoming critical due to major ecological changes in its area origin. Efforts on *in situ* conservation and participatory approaches as proposed by Jarvis, Brown et al (2008) and (Thomas, Dawson et al. 2011) could be very useful to maintain the adaptive potential of tomato genetic resources. Nuez and colleagues proposed to use *S. cheesmanii* accessions now stored in germplasm banks to reinstate some extinct populations in Galapagos Islands (Nuez et al. 2004). This could help avoiding the present paradox: the more knowledge we gain on tomato diversity and its evolutive history, the less available those genetic resources are available in the wild.

10. References

Albrecht, E., M. Escobar, et al. (2010). "Genetic diversity and population structure in the tomato-like nightshades Solanum lycopersicoides and S. sitiens." *Annals of Botany* 105(4): 535-554.

Alonso-Blanco, C., M. G. M. Aarts, et al. (2009). "What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation?" *The Plant Cell Online* 21(7): 1877-1896.

Atwell, S., Y. S. Huang, et al. (2010). "Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines." *Nature* 465(7298): 627-631.

Ayers, K. L. and H. J. Cordell (2010). "SNP Selection in genome-wide and candidate gene studies via penalized logistic regression." Genetic Epidemiology 34(8): 879-891.

Bai, Y. and P. Lindhout (2007). "Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future?" *Annals of Botany* 100(5): 1085-1094.

Barrero, L. S. and S. D. Tanksley (2004). "Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars." *Theoretical and Applied Genetics* 109(3)): 669-679.

Barry, C. S. and J. J. Giovannoni (2006). "Ripening in the tomato Green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling." *Proceedings of the National Academy of Sciences* 103(20): 7923-7928.

Beaumont, M. A., W. Zhang, et al. (2002). "Approximate Bayesian Computation in Population Genetics." *Genetics* 162(4): 2025-2035.

Bedinger, P., R. Chetelat, et al. (2011). "Interspecific reproductive barriers in the tomato clade: opportunities to decipher mechanisms of reproductive isolation." *Sexual Plant Reproduction* 24(3): 171-187.

Berloo, R., A. Zhu, et al. (2008). "Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes." *Theoretical and Applied Genetics* 117(1): 89-101.

Biodiversity-International (2006) "Crop wild relatives." 25.

Birchler, J. A., H. Yao, et al. (2010). "Heterosis." *The Plant Cell Online* 22(7): 2105-2112.

Bombarely, A., N. Menda, et al. (2010). "The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl." *Nucleic Acids Research* 39(Database): D1149-D1155.

Borrell, B. (2009). "How to Grow a Better Tomato: The Case against Heirloom Tomatoes." *Scientific American*
http://www.scientificamerican.com/article.cfm?id=case-against-heirloom-tomatoes

Bovy, A., E. Schijlen, et al. (2007). "Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics." *Metabolomics* 3(3): 399-412.

Brandwagt, B. F., L. A. Mesbah, et al. (2000). "A longevity assurance gene homolog of tomato mediates resistance to *Alternaria alternata* f. sp. *lycopersici* toxins and *fumonisin* B1." *Proceedings of the National Academy of Sciences* 97(9): 4961-4966.

Bredemeijer, M., J. Cooke, et al. (2002). "Construction and testing of a microsatellite database containing more than 500 tomato varieties." *Theoretical and Applied Genetics* 105(6-7): 1019-1026.

Brommonschenkel, S. H., A. Frary, et al. (2000). "The Broad-Spectrum Tospovirus Resistance Gene Sw-5 of Tomato Is a Homolog of the Root-Knot Nematode Resistance Gene Mi." *Molecular Plant-Microbe Interactions* 13(10): 1130-1138.

Butler, L. (1952). "The linkage map of the tomato." *Journal of Heredity* 43: 25-36.

Caicedo, A. L. (2008). "Geographic diversity cline of R gene homologs in wild populations of *Solanum pimpinellifolium* (Solanaceae)." *American Journal of Botany* 95(3): 393-398.

Caicedo, A. L. and B. A. Schaal (2004). "Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene." *Molecular Ecology* 13(7): 1871-1882.

Canady, M. A., V. Meglic, et al. (2005). "A library of *Solanum lycopersicoides* introgression lines in cultivated tomato." *Genome* 48(4): 685-697.

Casals, J., L. Pascual, et al. (2011). "The risks of success in quality vegetable markets: Possible genetic erosion in Marmande tomatoes (*Solanum lycopersicum* L.) and consumer dissatisfaction." *Scientia Horticulturae* 130(1): 78-84.

Causse, M. Duffe, P. et al (2004). "A genetic map of candidate genes and QTLs involved in tomato fruit size and composition." *Journal of Experimental Botany* 55(403): 1671-1685.

Causse, M., C. Friguet, et al. (2010). "Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness." *Journal of Food Science* 75(9): S531-S541.

Causse, M., V. Saliba-Colombani, et al. (2002). "QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits." *Journal of Experimental Botany* 53(377): 2089-2098.

Chen, K. Y., B. Cong, et al. (2007). "Changes in Regulation of a Transcription Factor Lead to Autogamy in Cultivated Tomatoes." *Science* 318(5850): 643-645.

Chunwongse, J. C., L. Black, et al. (2002). Molecular mapping of the Ph-3 gene for late blight resistance in tomato. *Journal of Horticultural science and Technology* 77(3): 281-286.

Comai, L. and S. Henikoff (2006). "TILLING: practical single-nucleotide mutation discovery." *The Plant Journal* 45(4): 684-694.

Cong, B., L. S. Barrero, et al. (2008). "Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication." *Nature Genetics* 40(6): 800-804.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London, J. Murray.

Darwin, C. and A. Wallace (1858). "On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by natural means of selection." *Journal of the Proceedings of the Linnean Society of London. Zoology* 3(9): 45-62.

Darwin, S. C., S. Knapp, et al. (2003). "Taxonomy of tomatoes in the Galpagos Islands: native and introduced species of *Solanum* section *Lycopersicon* (Solanaceae) " *Systematics and Biodiversity* 1(1): 29-53.

Davey, J. W., P. A. Hohenlohe, et al. (2011). "Genome-wide genetic marker discovery and genotyping using next-generation sequencing." *Nature Reviews Genetics* 12(7): 499-510.

de Candolle, A. P. (1882). *L'origine des plantes cultivées*. Paris.

de Vos, R. C. H., R. D. Hall, et al. (2011). Metabolomics of a Model Fruit: Tomato. *Annual Plant Reviews* 109-155.

Diez M.-J. and Nuez, F. (2008). Tomato. *Vegetables II Fabaceae, Liliaceae, Solanaceae, and Umbelliferae*. Prohens J. and Nuez, F.: 249-323.

Dixon, M. S., D. A. Jones, et al. (1996). "The Tomato Cf-2 Disease Resistance Locus Comprises Two Functional Genes Encoding Leucine-Rich Repeat Proteins." *Cell* 84(3): 451-459.

Do, P. T., M. Prudent, et al. (2010). "The influence of fruit load on the tomato pericarp metabolome in a *Solanum chmielewskii* introgression line population." *Plant Physiology* 154(3): 1128-1142.

Doebley, J. F., B. S. Gaut, et al. (2006). "The molecular genetics of crop domestication." *Cell* 127(7): 1309-1321.

Doganlar, S., A. Frary, et al. (2002). "Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589)." *Genome* 45(6): 1189-1202.

Elshire, R. J., J. C. Glaubitz, et al. (2011). "A robust, simple genotyping-by-sequencing (GBS) Approach for High Diversity Species." *PLoS ONE* 6(5): e19379.

Ernst, K., A. Kumar, et al. (2002). "The broad-spectrum potato cyst nematode resistance gene (Hero) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region." *The Plant Journal* 31(2): 127-136.

Eshed, Y. and D. Zamir (1995). "An Introgression Line Population of Lycopersicon pennellii in the Cultivated Tomato Enables the Identification and Fine Mapping of Yield-Associated QTL." *Genetics* 141(3): 1147-1162.

FAO (2010). *Plant genetic ressource for food and agriculture*. Rome, Food and Agriculture Organisation of the United Nations

Fernie, A. and H. Klee (2011). "The use of natural genetic diversity in the understanding of metabolic organization and regulation." Frontiers in Plant Science 2.

Fernie, A. R. and N. Schauer (2009). "Metabolomics-assisted breeding: a viable option for crop improvement?" *Trends in Genetics* 25(1): 39-48.

Finkers, R., H. de Weerd, et al. (january 2011). *BreeDB:* A database suppporting quantitative aspects of plant breeding. Plant and Animal Genomes XIX, San Diego.

Finkers, R., A. van Heusden, et al. (2007). "The construction of a Solanum habrochaites LYC4 introgression line population and the identification of QTLs for resistance to Botrytis cinerea." *Theoretical and Applied Genetics* 114(6): 1071-1080.

Food, Royal Norvegian Ministry of Agriculture (2008) "Svalbard global seedvault - more about the constuction from the directorate of public construction and property." 27.

Foolad, M. R. (2007). "Genome Mapping and Molecular Breeding of Tomato." *International Journal of Plant Genomics* 2007: 1-52.

Frary, A., Nesbitt et al (2000). "fw2.2: A quantitative trait locus key to the evolution of tomato fruit size." *Science* 289(5476): 85-88.

Fridman, E., F. Carrari, et al. (2004). "Zooming on a quantitative trait for tomato yield using interspecific introgressions." *Science* 305(5691): 1786-1789.

Gady, A., F. Hermans, et al. (2009). "Implementation of two high through-put techniques in a novel application: detecting point mutations in large EMS mutated plant populations." *Plant Methods* 5(1): 13.

García-Martínez, S., L. Andreani, et al. (2006). "Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: utility for grouping closely related traditional cultivars." *Genome* 49(6): 648-656.

Grandillo, S., R. Chetelat, et al. (2011). *Solanum* sect. *Lycopersicon Wild Crop Relatives: Genomic and Breeding Resources*. C. Kole, Springer Berlin Heidelberg: 129-215.

Grandillo, S., H. M. Ku, et al. (1999). "Identifying the loci responsible for natural variation in fruit size and shape in tomato." *Theoretical and Applied Genetics* 99(6): 978-987.

Gur, A., S. Osorio, et al. (2010). "*hi2-1*, A QTL which improves harvest index, earliness and alters metabolite accumulation of processing tomatoes." *Theoretical and Applied Genetics* 121(8): 1587-1599.

Gur, A., Y. Semel, et al. (2011). "Yield quantitative trait loci from wild tomato are predominately expressed by the shoot." *Theoretical and Applied Genetics* 122(2): 405-420.

Hamblin, M. T., E. S. Buckler, et al. (2011). "Population genetics of genomics-based crop improvement methods." *Trends in Genetics* 27(3): 98-106.

Harlan, J. R. (1971). "Agricultural Origins: Centers and Noncenters." *Science* 174(4008): 468-474.

He, C., V. Poysa, et al. (2003). "Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars." *Theoretical and Applied Genetics* 106(2): 363-373.

Hemming, M. N., S. Basuki, et al. (2004). "Fine mapping of the tomato gene for fusarium wilt resistance and elimination of a co-segregating resistance gene analogue as a candidate for *I-3*." *Theoretical and Applied Genetics* 109(2): 409-418.

Hey, J. and R. Nielsen (2004). "Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*." *Genetics* 167(2): 747-760.

Horkheimer, H. (1973). *Alimentación y obtención de alimentos en el Perú prehispánico.* Lima.

Isaacson, T., G. Ronen, et al. (2002). "Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of β-carotene and xanthophylls in plants." *The Plant Cell Online* 14(2): 333-342.

Itkin, M., H. Seybold, et al. (2009). "TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network." *The Plant Journal* 60(6): 1081-1095.

Jarvis, D. I., A. H. D. Brown, et al. (2008). "From the Cover: A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities." *Proceedings of the National Academy of Sciences* 105(14): 5326-5331.

Jenkins, J. (1948). "The origin of the cultivated tomato." *Economic Botany* 2(4): 379-392.

Jimenez-Gomez, J. and J. Maloof (2009). "Sequence diversity in three tomato species: SNPs, markers, and molecular evolution." *BMC Plant Biology* 9(1): 85.

Kamenetzky, L., R. Asís, et al. (2010). "Genomic analysis of wild tomato introgressions determining metabolism- and yield-associated traits." *Plant Physiology* 152(4): 1772-1786.

Keurentjes, J. J. B. (2009). "Genetical metabolomics: closing in on phenotypes." *Current Opinion in Plant Biology* 12(2): 223-230.

Kramer, M. G. and K. Redenbaugh (1994). "Commercialization of a tomato with an antisense polygalacturonase gene: The FLAVR SAVR™ tomato story." *Euphytica* 79(3): 293-297.

Krieger, U., Z. B. Lippman, et al. (2010). "The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato." *Nature Genetics* 42(5): 459-463.

Ku, H. M., S. Doganlar, et al. (1999). "The genetic basis of pear-shaped tomato fruit." *Theoretical and Applied Genetics* 99(5): 844-850.

Kurlovich, B. S., S. I. Rep'ev, et al. (2000). "The significance of Vavilov's scientific expeditions and ideas for development and use of legume genetic resources." *Plant Genetic Resources* 124: 23-32.

Labate, J. A. and A. Baldo (2005). "Tomato SNP Discovery by EST Mining and Resequencing." *Molecular Breeding* 16(4): 343-349.

Labate, J. A., S. Grandillo, et al. (2007). Tomato. *Genome mapping and molecular breeding in plants*. C. Kole. NY, Springer Publishing. 5: 1-125.

Labate, J. A., L. D. Robertson, et al. (2009). "Multilocus sequence data reveal extensive departures from equilibrium in domesticated tomato (*Solanum lycopersicum* L.)." *Heredity* 103(3): 257-267.

Labate, J. A., S. M. Sheffer, et al. (2011). "Diversity and population structure in a geographic sample of tomato accessions." *Crop Science* 51(3): 1068-1079.

Lanahan, M. B., H. C. Yen, et al. (1994). "The never ripe mutation blocks ethylene perception in tomato." *The Plant Cell Online* 6(4): 521-530.

Lanfermeijer, F. C., J. Warmink, et al. (2005). "The products of the broken Tm-2 and the durable Tm-22 resistance genes from tomato differ in four amino acids." *Journal of Experimental Botany* 56(421): 2925-2933.

Langford, A. N. (1937). "The parasitism of *Cladosporium fulvum* cooke and the genetics of resistance to it." *Canadian Journal of Research*(15): 108-128.

Levin, I., N. Gilboa, et al. (2000). "*Fgr* , a major locus that modulates the fructose to glucose ratio in mature tomato fruits." *Theoretical and Applied Genetics* 100(2): 256-262.

Li, W. and R. T. Chetelat (2010). " A pollen factor linking inter- and intraspecific pollen rejection in tomato." *Science* 330(6012): 1827-1830.

Lieberman, M., O. Segev, et al. (2004). "The tomato homolog of the gene encoding UV-damaged DNA binding protein 1 (DDB1) underlined as the gene that causes the high pigment-1 mutant phenotype." *Theoretical and Applied Genetics* 108(8): 1574-1581.

Ling, H.-Q., P. Bauer, et al. (2002). "The tomato fer gene encoding a bHLH protein controls iron-uptake responses in roots." *Proceedings of the National Academy of Sciences* 99(21): 13938-13943.

Ling, H.-Q., G. Koch, et al. (1999). "Map-based cloning of chloronerva, a gene involved in iron uptake of higher plants encoding nicotianamine synthase." *Proceedings of the National Academy of Sciences* 96(12): 7098-7103.

Lippman, Z. and S. D. Tanksley (2001). "Dissecting the Genetic Pathway to Extreme Fruit Size in Tomato Using a Cross Between the Small-Fruited Wild Species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom." *Genetics* 158(1): 413-422.

Lippman, Z. B. and D. Zamir (2007). "Heterosis: revisiting the magic." *Trends in Genetics* 23(2): 60-66.

Liu, J., van Eck et al (2002). "A new class of regulatory genes underlying the cause of pear-shaped tomato fruit." *Proceedings of the National Academy of Sciences* 99(20): 13302-13306.

Lopes, J. S. and M. A. Beaumont (2010). "ABC: A useful Bayesian tool for the analysis of population data." *Infection, Genetics and Evolution* 10(6): 825-832.

Luckwill, L. C. (1943). "The genus *Lycopersicon*: an historical, biological, and taxonomical survey of the wild and cultivated tomatoes. ." *Aberdeen University Studies*(120): 1–44.

Manning, K., M. Tör, et al. (2006). "A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening." *Nature Genetics* 38(8): 948-952.

Martin, G., S. Brommonschenkel, et al. (1993). "Map-based cloning of a protein kinase gene conferring disease resistance in tomato." *Science* 262(5138): 1432-1436.

Mazzucato, A., R. Papa, et al. (2008). "Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (Solanum lycopersicum L.) landraces." *Theoretical and Applied Genetics* 116(5): 657-669.

McCue, G. A. (1952). "The history of the use of the tomato: an annotated bibliography." *Annals of the Missouri Botanical Garden* 39(4): 289-348.

McKhann, H. I., C. Camilleri, et al. (2004). "Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*." *The Plant Journal* 38(1): 193-202.

McMeekin, D. (1992). "Representations on pre-columbian spindle whorls of the floral and fruit structure of economic plants." *Economic Botany* 46(2): 171-180.

Menda, N., Y. Semel, et al. (2004). "In silico screening of a saturated mutation library of tomato." *The Plant Journal* 38(5): 861-872.

Mendel, G. (1866). Experiments in plant hybridization : original traduction from *Versuche über Plflanzenhybriden* by W. Bateson completed by R. Blumberg, Electronic Scholarly Publishing Project.

Miller, J. C. and S. D. Tanksley (1990). "RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*." *Theoretical and Applied Genetics* 80(4): 437-448.

Minoia, S., A. Petrozza, et al. (2010). "A new mutant genetic resource for tomato crop improvement by TILLING technology." *BMC Research Notes* 3(1): 69.

Monforte, A. J. and S. D. Tanksley (2000). "Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery." *Genome* 43(5): 803-813.

Moreau, P., P. Thoquet, et al. (1998). "Genetic Mapping of Ph-2, a Single Locus Controlling Partial Resistance to Phytophthora infestans in Tomato." *Molecular Plant-Microbe Interactions* 11(4): 259-269.

Mounet, F., A. Moing, et al. (2009). " Gene and metabolite regulatory network of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development." *Plant Physiology* 149(3): 1505-1528.

Mueller, L. A., R. K. Lankhorst, et al. (2009). "A Snapshot of the emerging tomato genome sequence." *Plant Gen.* 2(1): 78-92.

Müller, C. H. (1940a). "A revision of the genus *Lycopersicon*. ." *U.S.D.A. Miscellanous Publications*(382): 1–28.

Müller, C. H. (1940b). "The taxonomy and distribution of the genus *Lycopersicon*." *National Horticulture Magazine*(19): 157–160.

Muños, S., N. Ranc, et al. (2011). "Increase in tomato locule number is controlled by two SNPs located near WUSCHEL." *Plant Physiology*. 156(4): 2244-2254.

Myles, S., J. Peiffer, et al. (2009). " Association mapping: critical considerations shift from genotyping to experimental design.." *The Plant Cell Online* 21(8): 2194-2202.

Nakazato, T. and E. A. Housworth (2011). "Spatial genetics of wild tomato species reveals roles of the Andean geography on demographic history." *American Journal of Botany* 98(1): 88-98.

Navez, B. (2011). *Tomate: Qualité et préférences* Centre Technique Interprofessionnel Fruits Legumes.

Nei, M., T. Maruyama, et al. (1975). " The bottleneck effect and genetic variability in populations." *Evolution* 29(1): 1-10.

Nesbitt, T. C. and S. D. Tanksley (2002). "Comparative sequencing in the genus *Lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes." *Genetics* 162(1): 365-379.

Nuez, F., J. Prohens, et al. (2004). "Relationships, origin, and diversity of Galápagos tomatoes: implications for the conservation of natural populations." *American Journal of Botany* 91(1): 86-99.

Okabe, Y., E. Asamizu, et al. (2011). "Tomato TILLING Technology: development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries." *Plant and Cell Physiology* 52(11): 1994-2005.

Ori, N., Y. Eshed, et al. (1997). "The I2C Family from the wilt disease resistance Locus I2 belongs to the nucleotide binding, Leucine-Rich Repeat superfamily of plant resistance genes." *The Plant Cell Online* 9(4): 521-532.

Paran, I., I. Goldman, et al. (1995). "Recombinant inbred lines for genetic mapping in tomato." *Theoretical and Applied Genetics* 90(3): 542-548.

Paran, I. and E. van der Knaap (2007). "Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper." *Journal of Experimental Botany* 58(14): 3841-3852.

Park, Y.-J., M. West, et al. (2004). "Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars tomato (*Lycopersicon esculentum* L.)." *Genome* 47(3): 510-518.

Paterson, A., S. Damon, et al. (1991). "Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments." *Genetics* 127: 181-187.

Paterson, A. H., E. S. Lander, et al. (1988). "Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms." *Nature* 335: 721-726.

Peralta, I. E. and D. Spooner (2007). History, origin and early cultivation of tomato (*Solanaceae*). *Genetic improvement of Solanaceous crops*. M. K. Razdan and A. K. Mattoo. Enfield (NH), Science Publisher. 2: 1-24.

Peralta, I. E., D. Spooner, et al. (2007). "Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; *Solanaceae*)." *Systematic Botany Monographs* 84.

Peralta, I. E. and D. M. Spooner (2001). "Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*)." *American Journal of botany* 88(10): 1888-1902.

Peralta, I. E. and D. M. Spooner (2005). Morphological characterization and relationships of wild tomatoes (*Solanum L.* Section *Lycopersicon*). *A Festschrift for William G. D'Arcy*. T. B. Croat, V. C. Hollowell and R. C. Keating, Missouri Botanical Garden Press. 104: 227-257.

Pertuzé, R. A., Y. Ji, et al. (2002). "Comparative linkage map of the *Solanum lycopersicoides* and *S. sitiens* genomes and their differentiation from tomato." *Genome* 45(6): 1003-1012.

Piron, F., M. Nicolaï, et al. (2010). "An Induced Mutation in Tomato eIF4E Leads to Immunity to Two potyviruses." *PLoS ONE* 5(6): e11313.

Price, A. L., N. A. Zaitlen, et al. (2010). "New approaches to population stratification in genome-wide association studies." *Nature Reviews Genetics* 11(7): 459-463.

Pritchard, J. K., M. Stephens, et al. (2000). "Inference of population structure ssing multilocus genotype data." *Genetics* 155(2): 945-959.

Purugganan, M. D. and D. Q. Fuller (2009). "The nature of selection during plant domestication." *Nature* 457(7231): 843-848.

Quinet, M., J.-M. Kinet, et al. (2011). "Flowering response of the uniflora:blind:self-pruning and jointless:uniflora:self-pruning tomato (*Solanum lycopersicum*) triple mutants." *Physiologia Plantarum* 141(2): 166-176.

Ramsay, L., J. Comadran, et al. (2011). "INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1." *Nature Genetics* 43(2): 169-172.

Ranc, N. (2010). Analyse du polymorphisme de gènes de composantes de la qualité des fruits dans les ressources génétiques sauvages et cultivées de tomate ; recherche d'association gènes/QTL. Montpellier, SupAgro. Thèse de doctorat: 348.

Ranc, N., S. Muños, et al. (2008). "A clarified position for *Solanum lycopersicum* var. cerasiforme in the evolutionary history of tomatoes (*Solanaceae*)." *BMC Plant Biology* 8(1): 130.

Rick (1977). "Conservation of tomato species germplasm. ." *California Agriculture* 31((9)): 32-33

Rick, C. M. (1991). "Tomato Paste: A concentrated review of genetic highlights from the beginnings to the advent of molecular." *Genetics* 128(1): 1-5.

Rick, C. M. and R. T. Chetelat (1995). "Utilization of related wild species for tomato improvement." *Acta Horticulturae* 412: 21-38.

Rick , C. M. and F. Fobes (1975). "Allozyme variation in the cultivated tomato and closely related species." *Bulletin of the Torrey Botanical Club* 102(6): 376-384.

Rick, C. M., J. F. Fobes, et al. (1977). "Genetic variation in *Lycopersicon pimpinellifolium*: Evidence of evolutionary change in mating systems." *Plant Systematics and Evolution* 127(2): 139-170.

Rick, C. M., J. F. Fobes, et al. (1979). "Evolution of mating systems in *Lycopersicon hirsutum* as deduced from genetic variation in electrophoretic and morphological characters." *Plant Systematics and Evolution* 132(4): 279-298.

Rick , C. M. and M. Holle (1990). "Andean *Lycopersicum esculentum* var. *cerasiforme*: Genetic Variation and Its Evolutionary Significance." *Economic Botany* 44: 69-78.

Rigola, D., J. van Oeveren, et al. (2009). "High-throughput detection of induced mutations and natural variation using KeyPoint™ technology." *PLoS ONE* 4(3): 4761.

Robbins, M. D., S. C. Sim, et al. (2010). "Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato." *Journal of Experimental Botany* 62(6): 1831-1845.

Rodriguez, G. R., S. Munos, et al. (2011). "Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity." *Plant Physiology* 156(1): 275-285.

Ronen, G., L. Carmel-Goren, et al. (2000). "An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold color mutations in tomato." *Proceedings of the National Academy of Sciences* 97(20): 11102-11107.

Rousseaux, M., C. Jones, et al. (2005). "QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines." *Theoretical and Applied Genetics* 111(7): 1396-1408.

Sahagún, B. d. (1988). *Historia General de las Cosas de Nueva España*. Madrid.

Saliba-Colombani, V., M. Causse, et al. (2001). "Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits." *Theoretical and Applied Genetics* V102(2): 259-272.

Sanders, R. A. and W. Hiatt (2005). "Tomato transgene structure and silencing." *Nature Biotechnology* 23(3): 287-289.

Schauer, N., D. Zamir and A.R. Fernie (2004). "Metabolic profiling of leaves and fruit of wild species tomato: a survey of the Solanum lycopersicum complex." Journal of Experimental Botany 56(410): 297-307.

Schoen, D. and A. Brown (1993). " Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers." *Procceedings of the National Academy of Sciences* 90(22): 10623-10627.

Semel, Y., Nissenbaume et al (2006). "From the Cover: Overdominant quantitative trait loci for yield and fitness in tomato." *Proceedings of the National Academy of Sciences* 103(35): 12981-12986.

Shalit, A., A. Rozman, et al. (2009). "The flowering hormone florigen functions as a general systemic regulator of growth and termination." *Proceedings of the National Academy of Sciences* 106(20): 8392-8397.

Sharma, M., N. Singh, et al. (2008). "Expression of toxin co-regulated pilus subunit A (TCPA) of Vibrio cholerae; and its immunogenic epitopes fused to cholera toxin B subunit in transgenic tomato (*Solanum lycopersicum*)." *Plant Cell Reports* 27(2): 307-318.

Sicard, A. and M. Lenhard (2011). "The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants." *Annals of Botany* 107(9): 1433-1443.

Smith, A. F. (2000). *Souper tomatoes: the story of America's favorite food*, Rutgers University Press.

SolCAP (2008). USDA-CSREES awards $5.4 million to SolCAP *SolCAP News* . Michigan State University, East Lansing.

Stadler, T., A. M. Florez-Rueda, et al. (2012). "Testing for "Snowballing" hybrid incompatibilities in *Solanum*: impact of ancestral polymorphism and divergence estimates." *Molecular Biology and Evolution.* 29(1): 31-34

Stevens, M. A. and C. M. Rick. (1986). Genetics and breeding. *The Tomato Crop: A Scientific Basis for Improvement.* J. G. A. a. J. Rudich. London, Chapman and Hall: 35-109.

Stevens, R., M. Buret, et al. (2007). "Candidate genes and quantitative trait loci affecting fruit ascorbic acid content in three tomato populations." *Plant Physiology* 143(4): 1943-1953.

Szymkowiak, E. J. and E. E. Irish (2005). "JOINTLESS suppresses sympodial identity in inflorescence meristems of tomato." *Planta* 223(4): 646-658.

Tadmor, Y., E. Fridman, et al. (2002). "Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication." *Journal of Agricultural and Food Chemistry* 50(7): 2005-2009.

Tam, S. M., C. Mhiri, et al. (2005). "Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR." *Theoretical and Applied Genetics* 110(5): 819-831.

Tanksley, D., H. Medina-Filho, et al. (1981). "The effect of isozyme selection on metric characters in an interspecific backcross of tomato — basis of an early screening procedure." *Theoretical and Applied Genetics* 60(5): 291-296.

Tanksley, S. D., M. W. Ganal, et al. (1992). "High density molecular linkage maps of the tomato and potato genomes." *Genetics* 132(4): 1141-1160.

Tanksley, S. D., S. Grandillo, et al. (1996). "Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*." *Theoretical and Applied Genetics* 92(2): 213-224.

Tanksley, S. D. and G. S. Khush (2002). Charles Madera Rick. *Bibliographical Memoirs* 84: 307-319.

Tellier, A., I. Fischer, et al. (2011). "Fitness effects of derived deleterious mutations in four closely related wild tomato species with spatial structure." *Heredity* 107(3): 189-199.

Tellier, A., S. J. Y. Laurent, et al. (2011). "Inference of seed bank parameters in two wild tomato species using ecological and genetic data." *Proceedings of the National Academy of Sciences* 108(41): 17052-17057.

Thomas, M., J. Dawson, et al. (2011). "Seed exchanges, a key to analyze crop diversity dynamics in farmer-led on-farm conservation." *Genetic Resources and Crop Evolution* 58(3): 321-338.

Tieman, D., M. Taylor, et al. (2006). "Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde." *Proceedings of the National Academy of Sciences* 103(21): 8287-8292.

Tikunov, Y. A. Lommen , et al. (2005). "A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles." Plant Physiology 139(3): 1125-1137.

van Berloo, R., A. Zhu, et al. (2008). "Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes." *Theoretical and Applied Genetics* 117(1): 89-101.

van der Knaap, E., Z. B. Lippman, et al. (2002). "Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions." *Theoretical and Applied Genetics* 104(2): 241-247.

van der Knaap, E., A. Sanyal, et al. (2004). "High-Resolution Fine Mapping and Fluorescence in Situ Hybridization Analysis of sun, a Locus Controlling Tomato Fruit Shape, Reveals a Region of the Tomato Genome Prone to DNA Rearrangements." *Genetics* 168(4): 2127-2140.

Van Deynze, A., K. Stoffel, et al. (2007). "Diversity in conserved genes in tomato." *BMC Genomics* 8(1): 465.

Villand, J., P. W. Skroch, et al. (1998). "Genetic Variation among tomato accessions from primary and secondary centers of diversity." *Crop Science* 38: 1339-1347.

Vrebalov, J., D. Ruezinsky, et al. (2002). "A MADS-Box gene necessary for fruit ripening at the tomato ripening-inhibitor (Rin) locus." *Science* 296(5566): 343-346.

Xiao, H., N. Jiang, et al. (2008). "A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit." *Science* 319(5869): 1527-1530.

Yakovleff, E. a. F. L. H. (1935). "El mundo vegetal de los antiguos peruanos. ." *Revista del Museo Nacional Botánica etnológica* 3(3): 3-55.

Yang, W., X. Bai, et al. (2004). "Discovery of single nucleotide polymorphisms in Lycopersicon esculentum by computer aided analysis of expressed sequence tags." *Molecular Breeding* 14(1): 21-34.

Yeager, A. (1937). "Studies on the inheritance and development of fruit size and shape in the tomato." *Journal of Agricultural Research* 55: 141-152.

Yen, H. C., S. Lee, et al. (1995). "The Tomato Never-ripe Locus Regulates Ethylene-Inducible Gene Expression and Is Linked to a Homolog of the Arabidopsis ETR1 Gene." *Plant Physiology* 107(4): 1343-1353.

Yen, H. C., B. A. Shelton, et al. (1997). "The tomato high-pigment (hp) locus maps to chromosome 2 and influences plastome copy number and fruit quality." *Theoretical and Applied Genetics* 95(7): 1069-1079.

Yu, J., G. Pressoir, et al. (2005). "A unified mixed-model method for association mapping that accounts for multiple levels of relatedness." *Nature Genetics* 38(2): 203-208.

Zamir, D. (2001). "Improving plant breeding with exotic genetic libraries." *Nature Review Genetics* 2(12): 983 - 989.

Zamir, D. (2008). "Plant breeders go back to nature." *Nature Genetics* 40(3): 269-270.

Zamir, D., I. Ekstein-Michelson, et al. (1994). "Mapping and introgression of a tomato yellow leaf curl virus tolerance gene, *TY-1*." *Theoretical and Applied Genetics* 88(2): 141-146.

Zanor, M. I., J.-L. Rambla, et al. (2009). "Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents." *Journal of Experimental Botany* 60(7): 2139-2154.

Zhang, X., N. A. Buehner, et al. (2006). "Tomato is a highly effective vehicle for expression and oral immunization with Norwalk virus capsid protein." *Plant Biotechnology Journal* 4(4): 419-432.

Zhong, X. B., J. Bodeau, et al. (1999). "FISH to meiotic pachytene chromosomes of tomato locates the root-knot nematode resistance gene Mi-1 and the acid phosphatase gene Aps-1 near the junction of euchromatin and pericentromeric heterochromatin of chromosome arms 6S and 6L, respectively." *Theoretical and Applied Genetics* 98(3): 365-370.

Zuriaga, E., J. M. Blanca, et al. (2009). "Genetic and bioclimatic variation in *Solanum pimpinellifolium*." *Genetic Resources and Crop Evolution* 56: 39-51.

9

Genetic Diversity of Nigerian Cashew Germplasm

Olawale Mashood Aliyu
Department of Cytogenetics and Genome Analysis
Institute of Plant Genetics and Crop Research (IPK), Gatersleben
Germany

1. Introduction

Cashew (*Anacardium occidentale* L) is a tropical tree nut crop that is native to tropical South American countries around Bolivia, Brazil, Peru and West Indies (Nakasone & Paull, 1998; Samal et al., 2003), with the Amazonia forest of Brazil being the centre of origin (Mitchell & Mori, 1987). Recent archeological data from 47-million year old lake sediment in Germany has shown evidence of earlier distribution of cashew in Europe during the Tertiary period, thus, suggesting bio-geographic link between America and Africa continents in the distribution of genus *Anacardium* (Manchester et al., 2007). Since the introduction of the modern cashew into Africa and Asia continents about five centuries ago, the crop has spread widely and these new areas have become the centre of diversity of cashew today. Cashew has now become important commodity export crop in the third world countries like Benin Republic, Cote d'Ivore, Guinea Bissau, Ghana, India, Mozambique, Nigeria, Philippines, Srilanka,Tanzania and Vietnam.

Cashew is a drought resistance and evergreen perennial small tree plant with dense foliage and can grow as high as 15 meters or more. It is a member of *Anacardiaceae* family with about 75 genera and 700 species (Nakasone & Paull, 1998). Other members of the family *Anacardiaceae* include mango and pistachio. Out of the eight species identified in the genus *Anacardium*, only cashew (*occidentale*) is of economic importance because of its edible hypocarp (apple) and nutritious kernel from the drupaceous nut (Fig. 1). Cashew tree is mostly single-stemmed with umbrella-shaped canopy and the flowering is normally preceded by vegetative growth flush at the end of wet season in the southern hemisphere. The flowers comprise of male and hermaphrodite types in varying proportion and are produced at the end of new shoots in the periphery of the tree canopy. And because of the sticky nature of the cashew pollens, the plant tends more to insect pollination, with some low degree of selfing (Northwood, 1966; Masawe, 1994; Feitas & Paxton, 1996, Aliyu, 2008). The kidney shaped nut (drupe) attached to the swollen hypocarp (pseudo-fruit or apple) is the true biological fruit of the cashew tree (Fig. 1). The kernel, eaten as dessert is the most nutritional part of the cashew and important delicacy because of it high protein and low cholesterol fat. It contains about 47% fat and oil, 21% protein and 22% carbohydrates (Fetuga et al. 1975; Nayar, 1998) and 82% of this fat is unsaturated that helps in reducing body cholesterol level. Structurally, the kernel is protected by a hard shell or endocarp and a

spongy mesocarp which contains some acidic oil called cashew nut shell liquid (CNSL). The liquid has been reported for a number of potential industrial uses e.g. for brake linings, paints and vanishes etc. The juice derived from the fresh cashew apples contains high content of vitamin C, about three times higher than in citrus and pawpaw. This characteristic informed fresh consumption of cashew apples by many people, but unfortunately, this pseudo-fruit can only be kept afresh for a short time after harvesting because of it high sucrose content that aid rapid degradation (rotten).

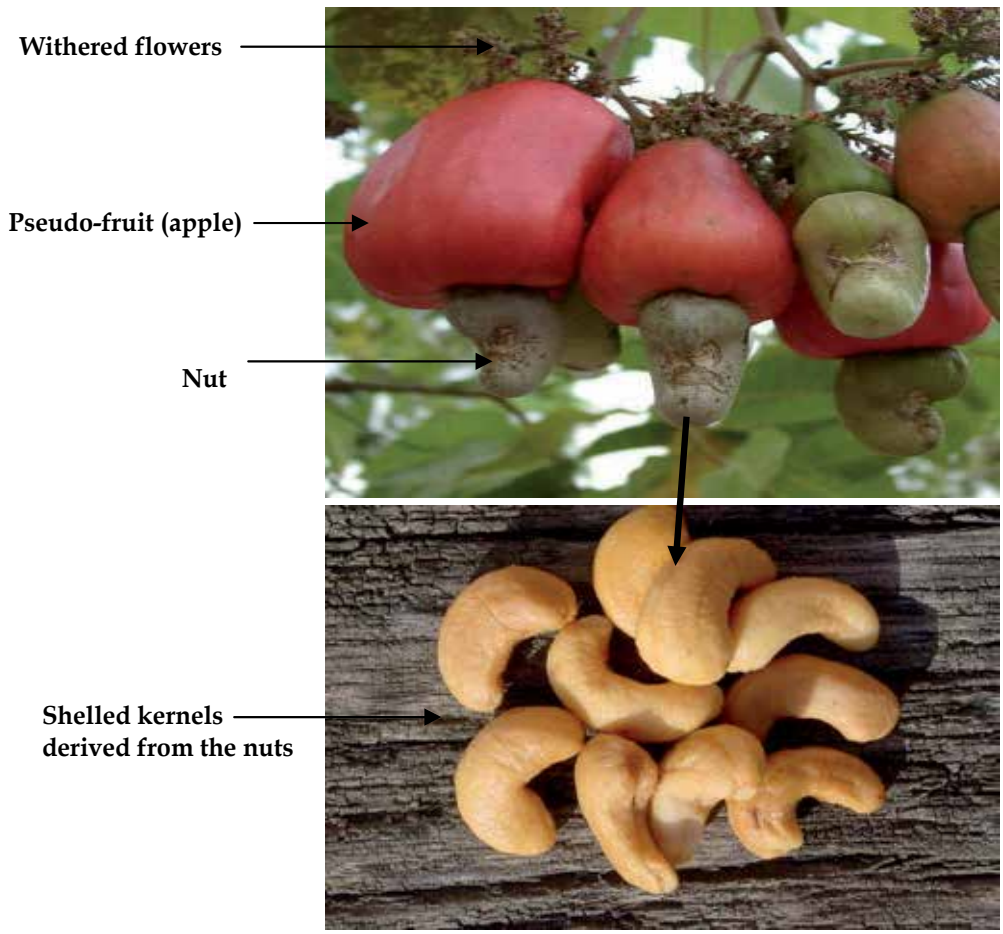


Fig. 1. The most economic products of cashew tree (i.e. hypocarps or apples, the nuts and its derivative kernels after shelling). (Fruit photo from Topper et al., 2001).

2. Cashew economy and production

The world produced about 3.4 million tons of raw cashew nuts in 2009 (FAO, 2011) and about one-third of the world cashew nut production comes from Africa with 50% of the continent exports from Nigeria (FAO, 2011) (Table 1). Cashew nut is one of the important agricultural commodities in Africa and has been contributing to Gross Domestic Product (GDP), National Income (NI) and foreign exchange earnings of many of the African cashew

producing nations. For example, cashew is not only one of the export commodity crops in Nigeria but a major source of livelihood for many smallholder farmers especially in the eastern and central parts of Nigeria (Topper et al., 2001). Cashew industry has play important role in the realization of the economic development of many of the African states, and has been one of the veritable platforms for the achievement of the United Nations Millennium Development Goals (MDGs) through economic empowerment of smallholder farmers and rural women, employment generation and small-medium scale industrialization especially in the rural areas (Fig. 2). Because it is currently a predominantly smallholder crop mostly grown as a monocrop though can be intercropped with food crops like cassava, cocoyam, cowpea, ground-nut, maize, pineapple and yam at the early stage of the crop development, cashew farming has been and would continue to provide jobs for teeming rural populace especially the women and youths.

Cashew thrives in a woodland-tall-grass savanna and dry-rainforest ecologies, and such ideal vegetations spread across about thirty (30) states in Nigeria and twenty-seven (27) of these effectively growing cashew as a commodity crop (Figure 3). The producing states are categorized into minor i.e. those with less than 10,000 hectares and major with greater than 10,000 hectares of cashew farms. The minor producing states are mostly from the southwest, south-south and north-eastern states because they combine cashew with other major commodity crops like cocoa, oil-palm, rubber and kola in the south and cereals and pulses in the north. By contrast, states with major plantations are spread across central and southeastern states where emphasis is on cashew production. As a matter of fact, survey of cashew production across the country in 2001 revealed that less than 20% of available cropable lands are under cultivation in most of these states (Topper et al., 2001), which imply prospect for future expansion.

Country/Year	2009	2006	2004	2002	2000	1998	1995	1990	1985	1980	1975	1970	1965	1961
Angola	1667	1590	1307	1139	800	1200	900	1200	1200	1200	1400	1300	1000	1000
Benin	49487	55000	45000	46771	40000	29084	15000	3000	1200	1086	345	627	50	50
Burkina Faso	3168	6141	4904	4364	3732	4015	2500	1074	645	200	N/A	N/A	N/A	N/A
Cote d'Ivoire	246383	235000	140636	104985	63380	39275	39400	6500	3500	600	450	300	400	400
Ghana	35647	34000	25000	9000	7697	8417	1208	480	N/A	N/A	N/A	N/A	N/A	N/A
Guinea-Bissau	64653	95000	96649	86000	72725	64000	29007	30000	13000	3500	2500	2500	2000	2000
Kenya	8381	11349	9332	10031	12500	14531	5000	7000	8500	15000	21600	22200	9000	3000
Madagascar	6072	6700	7289	6349	6500	6500	6000	5300	4000	3400	2900	2400	1900	1600
Mozambique	67846	62821	42988	50177	57894	51700	33423	22524	25000	71100	188000	184000	136000	107000
Nigeria	580761	636000	555000	514000	466000	152000	95000	30000	25000	25000	25000	25000	22000	7000
Senegal	4031	6332	5057	4500	7000	7000	1811	500	N/A	N/A	N/A	N/A	N/A	N/A
Tanzania	79100	77400	92810	55000	121200	93200	63400	17060	32750	41416	115840	107445	76000	50000
Togo	559	700	550	230	320	180	750	587	N/A	N/A	N/A	N/A	N/A	N/A
Total Africa	1151888	1235657	1032655	897746	862998	474736	295474	126858	116960	163218	358035	345772	248350	172050
Total World	3350929	3502184	2900969	2239194	1932142	1249827	1130730	732669	574013	464195	563785	511939	386303	287535
% of World	34.38	35.28	35.60	40.09	44.67	37.98	26.13	17.31	20.38	35.16	63.51	67.54	64.29	59.84

Source: FAO 2011 http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor N/A: Data not available.

Table 1. Cashew production (tons) in some African countries, total production for Africa and the world, and the percentage of Africa's production in the world for the period 1961 to 2009.

Fig. 2. Rural women separating nuts from pseudo-fruits (apple) i.e. postharvest processing (up) and grading of kernels after shelling before packaging (bottom). (Photos from Ghana Cashew Development Project report).

By 1995, the total land area cultivated to cashew in Nigeria was about 40,000 hectares with about 60% by small holders, 20% grow in the "wild", and 20% by the medium-large scale farmers. Currently, cashew cultivation has increased to about 330, 000 hectares (FAO, 2011) (Table 2) and consequently, annual nut production has been on the steady increase too, from 30,000 metric tons in the 1990 to 727,000 metric tons in the 2007 (FAO, 2011) (Table 2). The improvement in production has been attributed largely to increased cultivation and favourable economic policy that encourage more private sector investments.

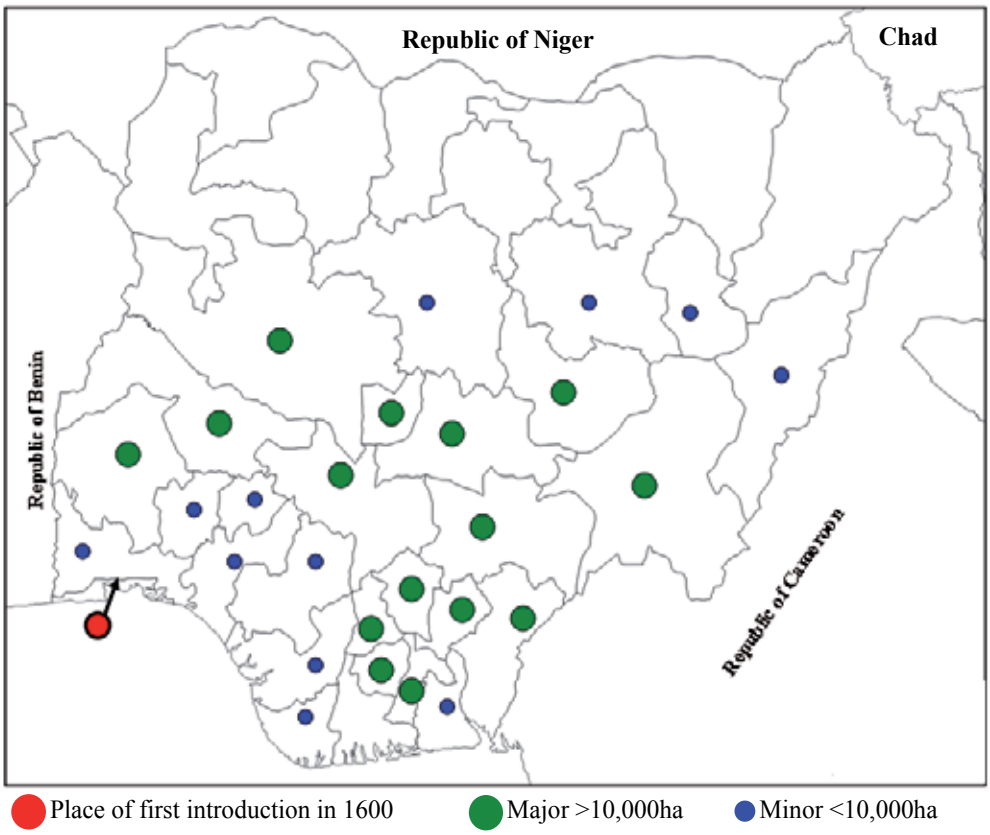


Fig. 3. Point of cashew introduction into Nigeria (Agege, Lagos) in the 16th century and the current production spread across 15 major and 12 minor producing states.

3. Historical perspective of the Nigerian cashew germplasm

Available records showed that cashew was introduced to Africa the same period with India i.e. about 16th century through trade mission by the Portuguese explorers (Johnson, 1973; Mitchell and Mori, 1987). In other word, Nigeria shared similar history with India and the first introductions around 400 years ago were planted in the coastal area around Agege, Lagos, Nigeria (Fig. 3), similar to Goa in India (Archak et al., 2009). Spontaneous planting from this coastal area facilitated its spread to other parts of Lower Niger of the country (Woodroof, 1967; Venkataramah, 1976; Togun, 1977; Ohler, 1979). After the introduction and

Year	Harvested area (ha)	tonnes/ha	Annual Production (tons)
1990	50,000	0.60	30,000
1991	75,000	0.60	45,000
1992	90,000	0.61	55,000
1993	120,000	0.63	75,000
1994	135,000	0.63	85,000
1995	155,000	0.61	95,000
1996	175,000	0.63	110,000
1997	243,000	0.51	125,000
1998	243,020	0.63	152,000
1999	248,000	1.75	417,000
2000	259,000	1.80	466,000
2001	265,000	1.83	485,000
2002	273,000	1.88	514,000
2003	277,000	1.89	524,000
2004	292,000	1.90	555,000
2005	309,000	1.92	594,000
2006	320,000	1.99	636,000
2007	330,000	2.00	660,000
2008	330,000	2.20	727,603
2009	330,000	1.76	580,761

Source: FAO (2011) http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor

Table 2. Cashew and nut production (land area (ha) yield/ha and annual total output) in Nigeria from 1990 – 2009.

spontaneous spread, cashew trees thrive in the wild for about three centuries (1600-1900s), with no commercial value and exploited mainly for the afforestation and control of gully erosion in most eastern parts of the country with subtropical savannah ecology.

3.1 First and second cashew germplasm introductions

Historically, the development of the Nigerian cashew industry shared history with the political independence of the nation with the emerging regional governments (Eastern, Northern and Western) of self-rule in the 1950s striving for economic independence and development. While the northern region with vast savannah ecology focused on the cultivation of cereals and pulses, the east and west explored tree crops (cocoa, rubber, oil palm and cashew). This period coincided with the development and advancement of cashew nut processing technology by Indians in the 1950s and the demand for raw nuts was on the increase. These historical events led to the establishment of first set of large commercial cashew plantations in Udi and Oghe by the Eastern Nigerian Development Corporation (ENDC) and Iwo, Eruwa and Oke-ogun by Western Nigeria Development Corporation (WNDC) in the period 1953-1960. In addition to sourcing planting materials from the first cashew introductions that had spread across the country (wild and few Agriculture and Forestry Departments), significant percentage of the planting materials (cashew seeds) used for the establishment of these large plantations (ENDC & WNDC) were imported directly from India by the Ministry of Agriculture of these two regional governments, and this

constituted the first major cashew genetic resources in the country. Cashew aggressively spread to the other parts of the country from these two regions thereafter and today the crop now grow effectively in twenty-seven (27) out of the thirty-six (36) federal states (Fig. 3) of Nigeria with estimated 330,000 hectares of cultivated cashew land and raw nut outputs of 640,000 metric tons (FAO, 2011). The establishment of large plantations in the country ushered in era of establishment of cashew kernel processing plants for value addition locally. And by the late 1960s to early 1970s, Nigeria was exporting both the raw nuts and processed kernels to India. However, with limited expertise locally, the emerging industry encountered challenges in all areas of the cashew value chain that later led to the enactment of a law in 1971 mandating the Cocoa Research Institute of Nigeria (CRIN) to carry out research and development into production, processing and marketing for the Nigerian cashew industry.

And as a part of the first initiative to improve production, Cocoa Research Institute of Nigeria embarked on its first exploration and germplasm collection from the existing cashew in the wild, farms and plantations (including WNDC) across the country as early as 1973 (Sanwo, 1973), and assembled the collections at the Institute's field gene banks located in Ibadan (Western Nigeria), Uhonmora (Mid-Eastern Nigeria) and Ochaja (North-central Nigeria) thereafter. These cashew collections were later found to be of narrow genetic base (see Akinwale & Esan, 1989; Aliyu & Awopetu, 2007a, 2007b) and step was taken to broadening the genetic pool by introducing more materials from India, Tanzania and Mozambique around 1978 and 1980, which constituted the 2nd germplasm introduction.

3.2 First cashew on-farm evaluation and selection programme

The introduction of germplasm was simultaneously accompanied by on-farm (farmers' fields or plantations) evaluation of some of the selected materials especially at the Western Nigerian Development Corporation (WND)'s plots. The outcome of such preliminary evaluation resulted in the selection of the 25 half-sib progenies (genotypes) with potential high yielding ($\geq$ 1000kg/ha) and were released as improved cashew cultivars tagged G-series by CRIN in the 1980s (see Akinwale & Esan, 1989). These G-varieties have been the main improved cashew materials that were distributed to farmers since 1988. This preliminary research intervention led to a dramatic improvement especially in terms of access to improved planting materials by farmers, which have hitherto sourcing materials from the wild, and laid a strong foundation for a investable cashew industry in the country and neighbouring nations.

3.3 Third cashew germplasm introduction

Introduction of the International Monetary Fund (IMF) structural adjustment programmes (SAP) and adoption of liberalization policy in late 1980s by the then government altered landscape for the cashew trade as well and ushered in increased competition within the African cashew industry with emphasis on quality. This paradigm shift brought into the fore the need for compliance with global quality standard in the competitive markets. Unfortunately, most of the African cashew exports, Nigerian inclusive fell short of the required global grading standards especially in size and quality of the nuts and kernels. The challenge prompted some affluent farmers to introduce new cashew materials (with characteristic bold nuts and high grade kernels) directly from Brazil towards the end of 1980s. In partnership with these farmers, the Cocoa Research Institute of Nigeria collected germplasm accessions from these introductions to expand the National Cashew germplasm

base. These Brazilian materials now constitute the third major introduction into Nigerian cashew germplasm. In total, Cocoa Research Institute of Nigeria currently housed about 22 hectares of cashew germplasm field across 6 locations in the country.

4. Challenges for the cashew production in Nigeria

Cashew production in Nigeria like other major producing nations is constrained by low yield and variable nut yields, nut quality and pests and diseases infestation. Incidentally, most of the existing farms were established with open pollinated seeds sourced from the wild and unimproved land races. Expectedly, large percentage of such trees exhibited significant variation in all traits, from growth form, yield to diseases and pests tolerance. For example, cashew trees have been found producing between 0kg and about 45kg nuts per tree with average yield per hectare sometimes between 0.4ton and 1.2tons/ha (Aliyu & Awopetu 2007a). Thus, broadening the genetic base through introduction of new alleles from exotic germplasm (Faenza et al., 1982) and systematic exploitation of heterosis (Masawe, 1994) of the germplasm through a recurrent selection can only be the best solution for tackling yield-related problems in cashew. The Cocoa Research Institute of Nigeria established with a national mandate to address these production challenges has initiated a comprehensive programme for the documentation and evaluation of the existing cashew germplasm in the country with the goal to identify better cultivars that combine higher yield with nut quality (size, colour, etc.). Unfortunately, such characterization and evaluation efforts were biased towards phenotypic and agronomic traits such as nut size, nut weight, sex ratio, colour of apple, size of the fruits, tree canopy, length of panicle and yield performance (Akinwale & Esan 1989; Mneney et al., 2001; Aliyu & Awopetu, 2007a) due to low capacity and limited molecular resources for tree crop research in the country. Although the traditional phenotypic method is useful, more often than not its efficiency could be masked by environmental effect, hence, the need to complement the phenotypic data with molecular method. A preliminary data from the on-going evaluation is summarized in the next section.

5. Update on the genetic diversity of the Nigerian cashew trees

In the last ten years, efforts have been concentrated on the fifty-nine (59) accessions (Tables 3) of Nigerian major cashew germplasm, comprising of three sub-populations (old land races, Indian and Brazilian). These trees were selected because of their reliable passport data and the presence of atleast three replicates for each accession. Available records showed that eleven (11) of these accessions were collected as clonal materials (CC-) from the farmers' fields along the Ochaja-Ankpa axis of the Kogi State (representing north central) in 1987/88. Furthermore, that these materials were probably from the remnants of the first introduction over three centuries ago (Sanwo, 1973; Akinwale & Esan, 1989). The second lot of twenty-three (23) accessions were mainly Indian introductions (CSI-) collected as open pollinated progenies from the Eastern- and Western- Nigeria Development Corporation (ENDC and WNDC) plantations in the 1970s, and were planted in the current locations (CRIN, Ibadan and Ochaja) between 1985 and 1986. And the third set of twenty-five (25) exotic accessions (CSO-) were from cashew materials recently introduced from Brazil by a private farm, Kosoni Ola Farms Limited, Oro, Kwara State and were planted in 1987/1988. These fifty-nine (59) accessions were planted at a spacing of 9.0m x 9.0m, with each accession represented by three entries.

Accession	Source of introduction	Pedigree	Current location
CC01	Unknown	Ochaja area	CRIN, Ibadan
CC02	″	″	″
CC03	″	″	″
CC04	″	″	″
CC05	″	″	″
CC06	″	″	″
CC07	″	″	″
CC08	″	″	″
CC09	″	″	″
CC10	″	″	″
CC11	″	″	″
CSI00	India	ENDC /WNDC	″
CSI01	″	″	″
CSI05	″	″	″
CSI06	″	″	″
CSI07	″	″	″
CSI09	″	″	″
CSI10	″	″	″
CSI11	″	″	″
CSI13	″	″	″
CSI14	″	″	″
CSI18	″	″	″
CSI23	″	″	″
CSI27	″	″	″
CSI30	″	″	″
CSI31	″	″	″
CSI36	″	″	″
CSI51	″	″	″
CSI58	″	″	″
CSI61	″	″	″
CSI62	″	″	″
CSI63	″	″	″
CSI66	″	″	″
CSI67	″	″	″
CSO01	Brazil	KFL Oro.	CRIN, Ibadan &CRIN, Ochaja
CSO02	″	″	″
CSO03	″	″	″
CSO04	″	″	″
CSO05	″	″	″
CSO06	″	″	″
CSO07	″	″	″
CSO08	″	″	″
CSO09	″	″	″
CSO10	″	″	″

CSO11	″	″	″
CSO12	″	″	″
CSO13	″	″	″
CSO14	″	″	″
CSO15	″	″	″
CSO16	″	″	″
CSO17	″	″	″
CSO18	″	″	″
CSO19	″	″	″
CSO20	″	″	″
CSO21	″	″	″
CSO22	″	″	″
CSO23	″	″	″
CSO24	″	″	″
CSO25	″	″	″

ENDC: Eastern Nigerian Development Corporation, WNDC: Western Nigerian Development Corporation, KFL: Kosoni-Ola Farm Limited.

Table 3. List of the fifty-nine cashew accessions and their pedigree analyzed for the yield and yield components variability over a period of ten years (1999-2009).

These cashew trees have attained full maturity (i.e. above 10 years old) at the commencement of the phenotypic evaluation in 1999/2000. Based on the data from previous studies (Masawe, 1994; Azevedo et al., 1998; Aliyu, 2006; Aliyu & Awopetu 2007a), only ten yield-related component characters (whole fruit weight – WWT, individual nut weight – NWT, total nut yield per tree – NYT, kernel weight – KWT, number of hermaphrodite flowers per panicle – HPP, percentage of pollen fertility - PPS, tree canopy size –TCS, days to optimum flowering- DFF, days to optimum fruit maturity - DFM and effective harvesting period (day) – HPD were selected for the phenotypic evaluation. The ten years data (1999-2009) were statistically analyzed using SYSTAT version 13.0 softwares (Systat Software, Inc. USA) and compared thereafter with three years data (Aliyu & Awopetu, 2007a).

Cluster analyses {Eucledian distance ward dendogram and principal component analysis (PCA} from the two studies showed no significant difference between the genetic groupings of the cashew accessions (see Aliyu & Awopetu, 2007a). The data grouped the accessions into five (5) major classes and eight (8) subsets that depict fusion based on source of introduction and/or breeding history (Figs. 4 & 5; Table 4). An overview of the phenotypic (genetic) variability of the accessions showed that the cluster analyses (Figs. 4 & 5) separated the trees into high yielding (clusters I, II & III) and low yielding (clusters IV & V) genotypes. And more than 50% of the accessions fell within the low yielding category. This trend does not only reflect enormous variability that could exist in a typical cashew field, but brought into the fore the level of redundancy in cashew farm in terms of yield performance. To facilitate the efficient utilization of better candidates identified from this evaluation exercise, a short qualitative description of their agronomic importance for each set derived from the cluster analyses (Figs. 4 & 5) is given below.

5.1 Some agronomic values of the Nigerian cashew trees

5.1.1 Very high yielding with moderate quality kernels

A mixture of two (2) Farmers clones and five (5) Indian accessions were the most prolific (producing >3000 nuts/tree/year) of all the fifty-nine accessions studied (Figs. 4 & 5; Table 4). These trees were characterized by heavy fruit clustering on the panicle and have potential to produce >2.5 tons of raw nuts per hectare annually. But the average weight of nuts is about 6.0g and individual nut weight ranging between 4.50 and 8.20g. Derivable kernels from these trees were mixture of W320 and W450 grades with the latter being prominent. The heavy annual production was consistent throughout the study and fruit maturity is mostly in the middle of the season i.e. March. The trees shared prolific fruiting (nut per tree) with accessions that constituted Cluster III (Figs. 4 & 5; Table 4). These cashew accessions are mostly suitable for immediate use as planting materials because of the superior yield characteristic, but little improvement work on the kernel quality is needed to select candidates with stable W320 kernels.

5.1.2 Moderate yielding with high quality kernels

Eleven (11) Brazilian accessions that combine moderate-high yielding, regular production with high quality nuts and kernels were identified from the analysis ((Figs. 4 & 5; Table 4). Weight of individual nut from these trees range between 8.50g and 14.0g, and was characterized by a mixture of W320 and W240 kernels, with the latter being in abundance. Furthermore, these trees were noted for early fruit production and short harvesting time. These materials are better than the G-series that were released and given to farmers since the late 1980s by the Cocoa Research Institute of Nigeria (CRIN) (see Akinwale & Esan, 1989). These eleven superior cashew accessions in addition to those described above (5.1.1) can be clonally propagated for release as improved cultivars to the farmers to meet the short term need and/or undergo further evaluation across different locations through a long term national and/or regional cashew improvement programme. In addition, these new materials can be used for the establishment of polyclonal seed gardens for both research and commercial uses.

5.1.3 Cashew trees prolific fruiting with compact canopy

Nine (9) of the accessions were (see list Table 4 & Fig. 4.) characterized by trees with small and compact canopy, though they are highly prolific in fruiting but the total output per tree were significantly influenced by smallness of its nuts and kernels. The genetic attributes of these nine cashew accessions include small sized trees with compact canopy. These plant materials could be useful for future breeding of cashew cultivars that would be adaptable to high density planting in an effort to improve outputs per unit area.

5.1.4 Cashew trees with low yield and poor agronomic qualities

The characterization exercise also revealed that about 40% of the experimental lots (25 accessions, Table 4; Figs. 4 & 5) were of low yield and poor in agronomic traits. However, because of the large canopy nature of these trees, they are good genetic resources for the afforestation, land reclamation and erosion controlled programmes in the arid regions and areas threatened by gully erosions. The proportion of such trees in the gene pools should be reduced significantly. Further studies are needed to understand the poor correlation between tree size and nut yield in cashew (Masawe et al. 1999; Aliyu & Awopetu, 2007a).

Cluster	Sub-cluster	Accessions in subclusters	No. of Indian accession	No. of Brazilian accession	No. of Farmers Land races	% of total accession (n=59)	% of Indian accession (n=23)	% of Brazilian accession (n=25)	% of Farmers Land races (n=11)
I	Ia.	CC06, CSI62, CSI31, CSI58, CSI66, CC05, CSI36	5	0	2	11.9	21.7	0	18.2
II	IIb.	CSO14, CSO19, CSO20, CSO12, CSO06, CSO05, CSO07, CSO01, CSO02, CSO15, CSO10	0	11	0	18.6	0	44.0	0
III	IIIc.	CSI16, CSO16, CSO14, CSI05, CSI18, CSO03, CSO13, CC11, CSI00	4	4	1	15.3	17.4	16.0	9.1
IV	IVd.	CSO18, CSO11, CC07, CC04, CC09, CSO17	0	3	3	10.2	0	12.0	27.3
	IVe.	CC10, CSI06, CSI30, CSI01, CSI09, CSI61, CSI10, CSI27,	7	0	1	13.6	30.4	0	9.1
	IVf.	CC03, CC01, CC02, CC08	0	0	4	6.8	0	0	36.4
	IVg.	CSI51, CSI67, CSI23, CSI07, CSI13, CSI14, CSI11	7	0	0	11.9	30.4	0	0
V	Vh.	CSO09, CSO08, CSO24, CSO23, CSO25, CSO24, CSO22	0	7	0	11.9	0	28	0

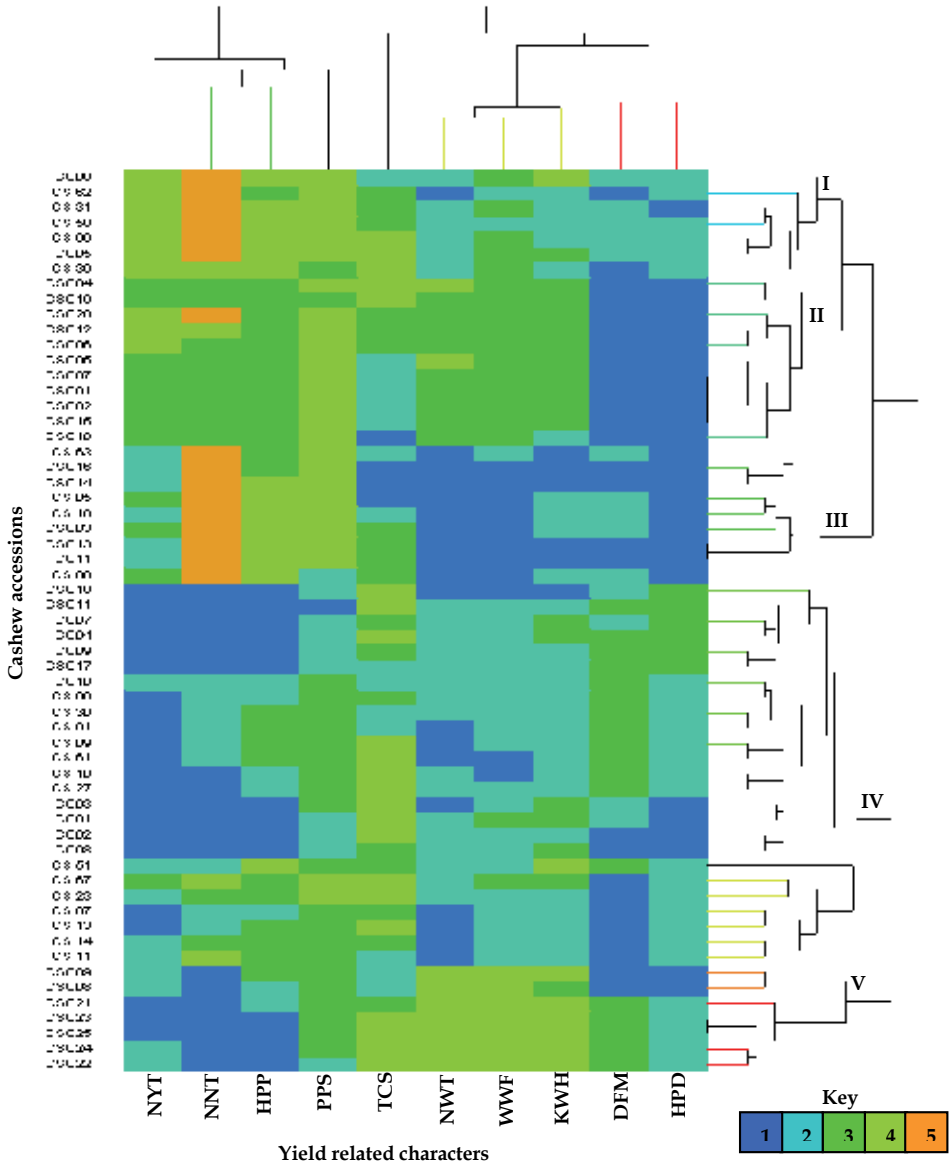
Table 4. Inter- and intra-cluster fusion of the 59 cashew accessions derived from the genetic diversity analysis showing source of origin/or breeding history effect.

5.1.5 Cashew trees with low yield and high quality kernels

Seven (7) accessions that consistently produced fewer but extra fruits (both apple and nut) were identified from the study. They are predominantly Brazilian collections (Table 4 and Figs. 4 & 5) with low yield because these trees rarely produce more than 300-400 fruits per trees. With average nut weight of 16g, these trees can be a good source for the introgression of genes for high grade kernels in cashew breeding programme. Other attributes of these five accessions include high volume of apple juice yield with low astringency. This category of cashew trees are often referred to as Jumbo varieties by farmers and are inadvertently collected as planting materials from unapproved sources. Unfortunately, recent study on nut size and number trade-off in cashew (Aliyu & Awopetu, 2011) has shown poor yield in this category of cashew trees, thereby corroborating this result. Hence, such materials are not suitable for the establishment of investable cashew farms, but could be used as a good source of genetic resources for research and developments of better varieties/cultivars.

6. Recurrent selection strategy for the development of hybrid cashew

Following the conclusion of the evaluation exercise and grouping of the entries into their respective agronomic groups, the Institute is embarking on a long term recurrent selection strategy for the development of improved cashew varieties. Apart from multilocational evaluation of most of the superior materials identified in the characterization exercise, the breeding plan included exploitation of heterosis between the highly prolific, but small fruits



Legends for the characters and key.

NYT:- Total nut yield per tree per year (kg) (1-Low, 2-Moderate, 3-High, 4-Very High); NNT:- Nuts per tree/year (1-Low, 2-Moderate, 3-High, 4-Very High, 5- Extra-super high); HPP:- Hermaphrodite flowers per panicle (1- Low, 2-Moderate, 3-High, 4-Very High); PPS:- Pollen grain fertility (%) (1- Low, 2-Moderate, 3-High, 4-Very High); TCS:- Tree canopy size (m^2) (1- Compact, 2- Moderate, 3- Large open, 4- Spread with extensive branches); NWT:- Nut weight (g) (1- Small, 2- Medium, 3- Large, 4- Extra large); WWF:- Whole fruit weight (g) (1- Small, 2- Medium, 3- Large, 4- Extra large);KWH:– kernel weight (g) (1- W450, 2-W320, 3-W240, 4-W180); DFM:- Days to optimum fruit maturity (1- Early, 2-Mid-season, 3- Late); HPD:- Harvesting period (1-Short, 2- Intermediate, 3-Prolong).

Fig. 4. Phenotypic variability of the between fifty-nine cashew germplasm accessions for the nine yield related components as obtained from the ten (10) years data (1999-2009).

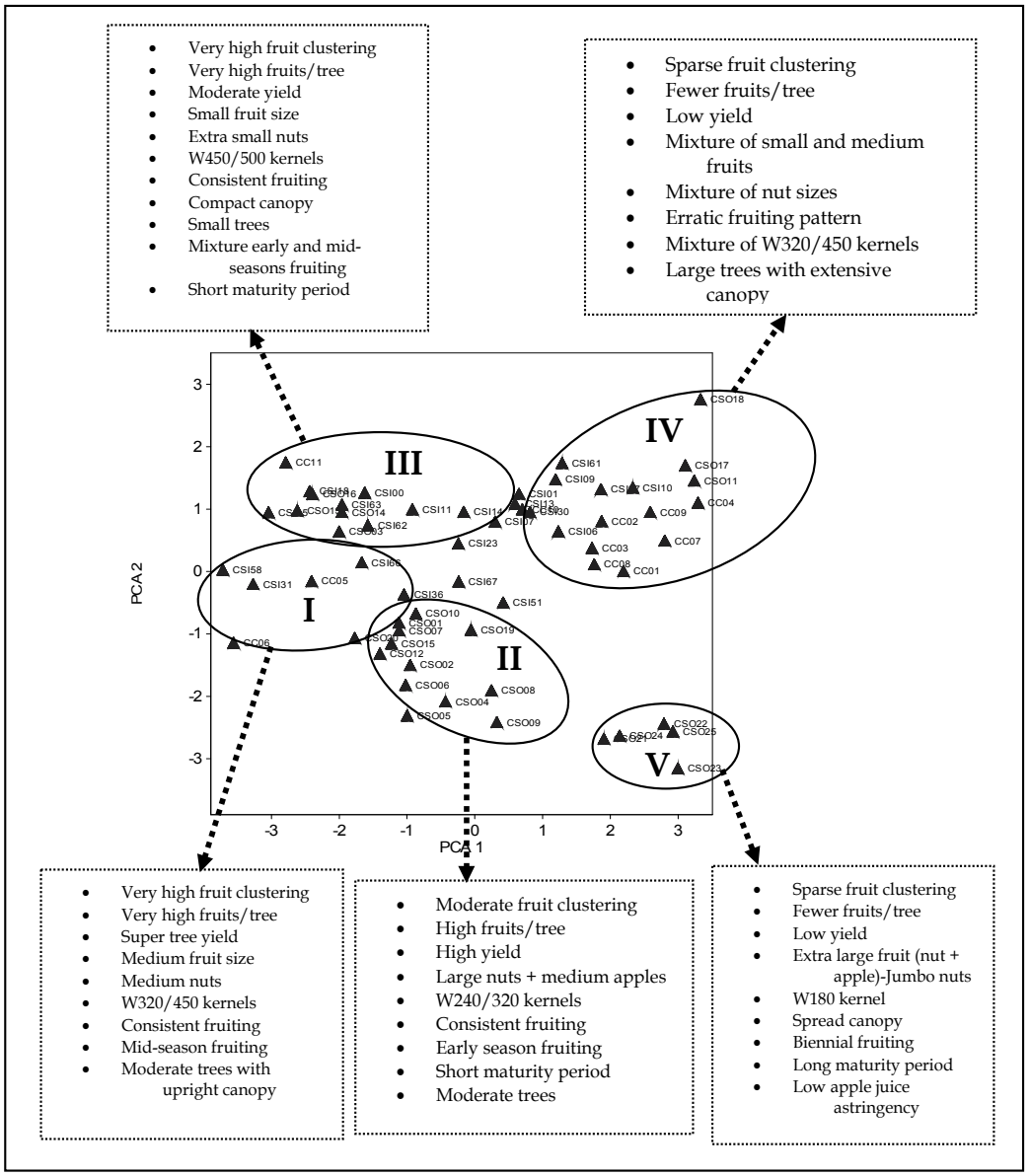


Fig. 5. Principal component analysis (PCA) of the fifty-nine cashew accessions showing five clusters and their respective agronomic characteristics.

and very low productive and extra large fruit trees through controlled hybridization and *in vitro* embryo culture, where necessary. Less impediment is expected from the hybridization exercise because of the high cross-compatibility in cashew (Aliyu, 2007, 2008), albeit, that the environmental conditions are favourable.

7. Application of protein-isoenzme electrophoresis as a useful tool for cashew characterization

In the absence of a well equipped laboratory with recent markers such as Random Amplification of Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP) and Simple Sequence Repeats (SSRs), protein-isozyme sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) has proven as a reliable alternative for the characterization of cashew germplasm. The PAGE analysis differentiated the accessions into six overlapping subclasses similar to the results of morphological characterization i.e. corroborating groupings based on phenotypic attributes of the trees (see Aliyu & Awopetu, 2007b). For example, out of fifteen (15) accessions of the Brazilian origin grouped together as a cluster I at 43.0% linkage distance by PAGE analysis (see Aliyu & Awopetu, 2007b), eleven (11) were clustered together as Cluster II in the 10 year phenotypic data, suggesting genetic congruity between phenotypic attributes and breeding history/or and source of origin of these cashew plants. Similarly, five accessions CSO24, CSO25, CSO22, CSO23 and CSO21 that are morphologically and agronomically similar for their extra-large (jumbo) fruits (nut and apple), late and irregular flowering and fruiting characteristics, and spreading and extensive branching pattern, clustered together on the PAGE analysis because they shared common isoenzymes and probably reinforcing theory of source of origin and/or increased nuclear DNA content as the underline genetic factor for the grouping. Beside the general characteristics, isoenzyme study further differentiates the accessions within a sub-cluster on the basis of apple skin pigmentation into red, yellow and orange. Similar to the results of the ongoing molecular (SSR) analysis, the protein-isozyme study equally revealed moderate genetic base for the Nigerian cashew germplasm. The result of the PAGE analysis (Aliyu & Awopetu, 2007b) has demonstrated that this biochemical method can be used as a reliable and effective alternative characterization tool in the assessment of genetic relationships and grouping into morphotypes in cashew.

8. Microsatellite analysis reveals genetic redundancy in Nigerian germplasm

In collaboration with the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, one hundred and eighty-seven (187) cashew accessions from different populations are being evaluated using 10 simple sequence repeats (SSR) markers (Croxford et al., 2006) (Table 5). An overview of the phenogram derived from the molecular data showed that eight (8) of the ten (10) microsatellite loci except, mAoR2 and mAoR47 were polymorphic (Fig. 6) and could be efficiently used for characterization and management of cashew germplasm. Preliminary data from the study showed significant level of redundancy (homogenous group) (Fig. 6) within the Nigerian cashew germplasm and the need to develop a core collection to reduce the cost of management of cashew germplasm. The data also depicts narrow genetic base of the Nigerian cashew germplasm. Apparently, future efforts at enhancing genetic base of Nigeria cashew germplasm should focus on collaborations and extensive linkages with relevant research institutions from Asia and South America.

9. *In vitro* embryo culture in cashew

Embryos from immature nuts of cashew have been successfully cultured *in vitro* at the Cocoa Research Institute of Nigeria, to establish a routine protocol for the regeneration of plantlets from crosses between distant parents (genotypes). In the trial, 2-, 4-, 6- and 8-weeks

Locus	Primer sequence (5′- 3′)	Repeat motif	T_a (°C)	Allelic size range (bp)
mAoR2	F: GGCCATGGGAAACAACAA R: GGAAGGGCATTATGGGTAAG	$(CA)_{10}(TA)_6$	58.2	366-375
mAoR3[a]	F: CAGAACCGTCACTCCACTCC R: ATCCAGACGAAGAAGCGATG	$(AC)_{12}(AAAAT)_2$	60.3	241-247
mAoR6[c]	F: CAAAACTAGCCGGAATCTAGC R: CCCCATCAAACCCTTATGAC	$(AT)_5(GT)_{12}$	58.2	143-157
mAoR7[b]	F: AACCTTCACTCCTCTGAAGC R: GTGAATCCAAAGCGTGTG	$(AT)_2(GT)_5AT(GT)_5$	58.2	178-181
mAoR11[c]	F: ATCCAACAGCCACAATCCTC R: CTTACAGCCCCAAACTCTCG	$(AT)_3(AC)_{16}$	60.3	234-236
mAoR17[b]	F: GCAATGTGCAGACATGGTTC R: GGTTTCGCATGGAAGAAGAG	$(GA)_{24}$	56.1	124-159
mAoR42[c]	F: ACTGTCACGTCAATGGCATC R: GCGAAGGTCAAAGAGCAGTC	$(CAT)_9TAT(CTT)_7$	60.3	197-206
mAoR47	F: AAGAGCTGCGACCAATGTTT R: CTTGAACTTGACACTTCATCCA	$(TAAA)_2(TA)_7(AAT)_5$	58.2	161-173
mAoR48[a]	F: CAGCGAGTGGCTTACGAAAT R: GACCATGGGCTTGATACGTC	$(GAA)_6(GA)_3$	58.2	172-178
mAoR52	F: GCTATGACCCTTGGGAACTC R: GTGACACAACCAAAACCACA	$(GT)_{16}(TA)_2$	58.2	191-203

Table 5. List of ten (10) microsatellite markers with optimal annealing temperature (T_a) and allelic size ranges (Croxford et al. 2006) used for the molecular analysis of the genetic diversity in 187 accessions of cashew.

old embryos were evaluated on different media compositions ranging from pure Murashige and Skoog (MS) agar medium (Murashige & Skoog, 1962) to modified MS medium supplemented with 1 mM each of naphthaleneacetic acid (NAA), benzyladenine (BA) and gibberellic acid (GA_3). The results showed that age of the embryo significantly influenced the rates of response and survival of the plantlets, with older embryos i.e. 6 weeks old and above performed better (Table 6) (Aliyu & Awopetu, 2005). Among the media composition tested, only modified MS medium supplemented with 1 mM of gibberellic acid (MS+GA_3)

Age of embryos (WAPo)[a]	MS[b]	MS+GA_3[c]	MS+NAA[d]	MS+BA[e]
2	0.00 (0.00%)	5.91 (9.85%)	0.00 (0.00%)	0.00 (0.00%)
4	12.99 (21.65%)	19.71 (32.85%)	14.79 (24.65%)	15.69 (26.15%)
6	37.20 (62.00%)	41.31 (68.85%)	36.09 (60.15%)	36.81 (61.35%)
8	45.69 (76.15%)	48.69 (81.15%)	43.20 (72.00%)	42.30 (70.50%)
DMRT ($\bar{x}$)	7.99b	9.63a	7.88b	7.85b

Percentage in parenthesis. LSD 0.05 = 0.332. DMRT: Duncan multiple range test.
a: WAPo: Week after pollination
b: Murashige and Skoog (MS) medium
c: MS medium + gibberellic acid
d: MS medium + naphthaleneacetic acid
e: MS medium + benzyladenine.

Table 6. Effect of age on the growth and development of cashew embryos in different *in vitro* culture media compositions.

supported growth of young embryos of 2 weeks old. In order words, factors such as medium composition, age of embryo and sometimes the genotype influence success rate of *in vitro* culture of cashew embryos. Photoperiod, temperature, source of explants, browning, and contamination are other known factors capable of affecting *in vitro* propagation of cashew (Jha, 1988; Das et al., 1996). The study deed showed that older cashew embryos seem to be autonomous of growth regulators i.e. inclusion of these synthetic plant hormones into media composition only became critical for very young explants.

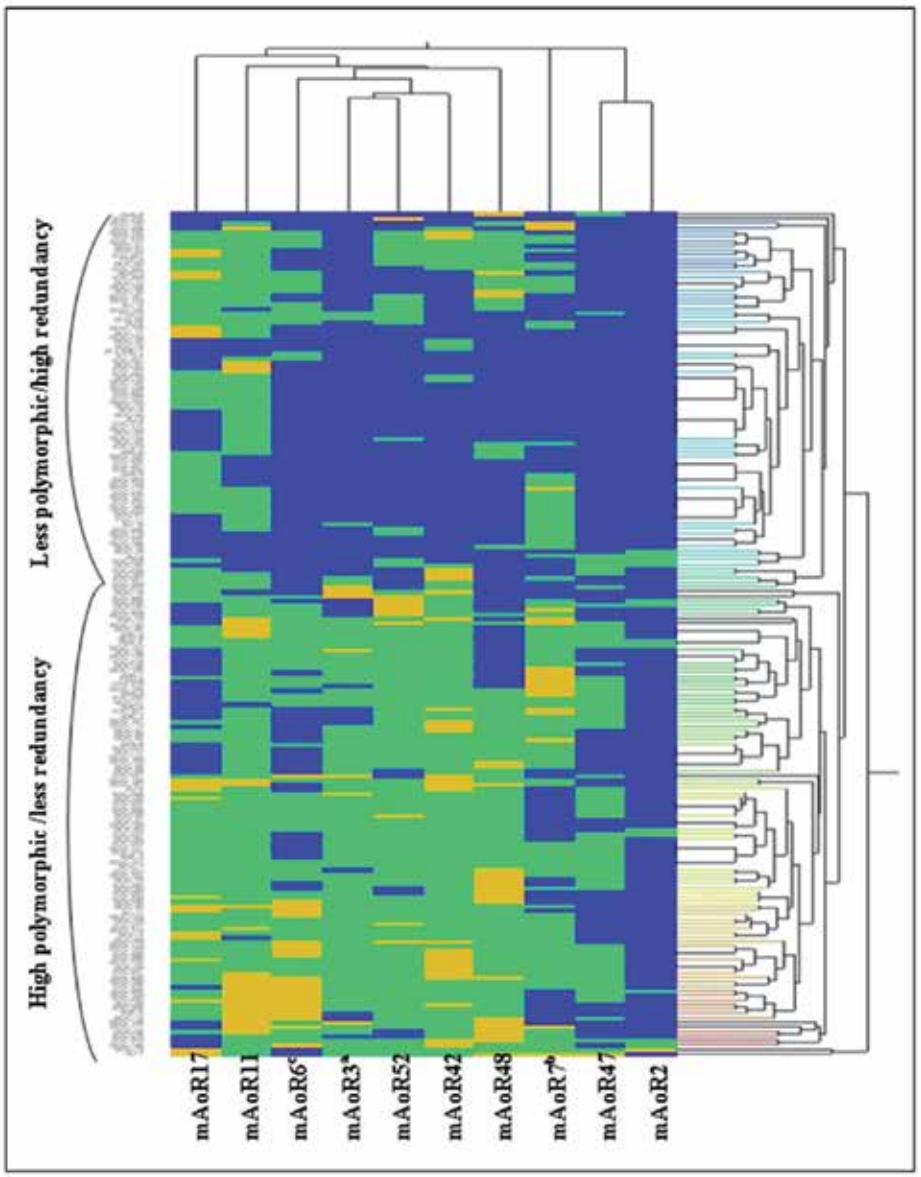


Fig. 6. An overview of the genetic diversity of 187 Nigerian cashew accessions from different populations derived from molecular analysis of 10 microsatellite marker loci.

10. Cashew cytology and cytometry

10.1 Relative nuclear DNA content and genome size in cashew

Information on genome size and ploidy needed for basic molecular breeding of this important commodity is rare. Recent flow cytometric analysis of fifty-four (54) cashew accessions from the Nigerian germplasm was carried out to determine the relative genome size, intraspecific variation and ploidy status of the species using *Solanum lycopersicum* cv. Stupicke as an internal standard reference. And because of the dearth of literatures on application of flow cytometry to cashew or its relatives, the study was preceded by the protocol optimization for the buffer system, sample size, internal reference standard and incubation time for isolated nuclei before analysis (Aliyu, 2011under review). From the analysis of the nuclear suspension in terms of fluorescence intensity, background yield (%), nuclear yield (nuclei s^{-1} mg^{-1}) and coefficient of variation (%) of *G0/G1* peak, the cashew plant showed preference for Otto's buffer, leaf sample size of about 70mg and maximum of 20 mins incubation period. *S. lycospersicum* cv. Stupicke and *Glycine max* cv. Polanka are the ideal internal reference standards for the genome size determination in cashew (Fig. 7a). The average relative nuclear DNA content in cashew is small with relative genome size of 2C = 1.01 pg i.e. about 988Mbp (1C = 490Mbp) recorded for the analyzed accessions (Aliyu, 2010). The relative nuclear DNA contents ranged from smallest (0.903 pg) to largest (1.285 pg) i.e. about 1.36 – fold range. Slight variation for relative genome size was recorded between the accessions (Fig. 7b) and it would be too hasty to conclude that such observation is an

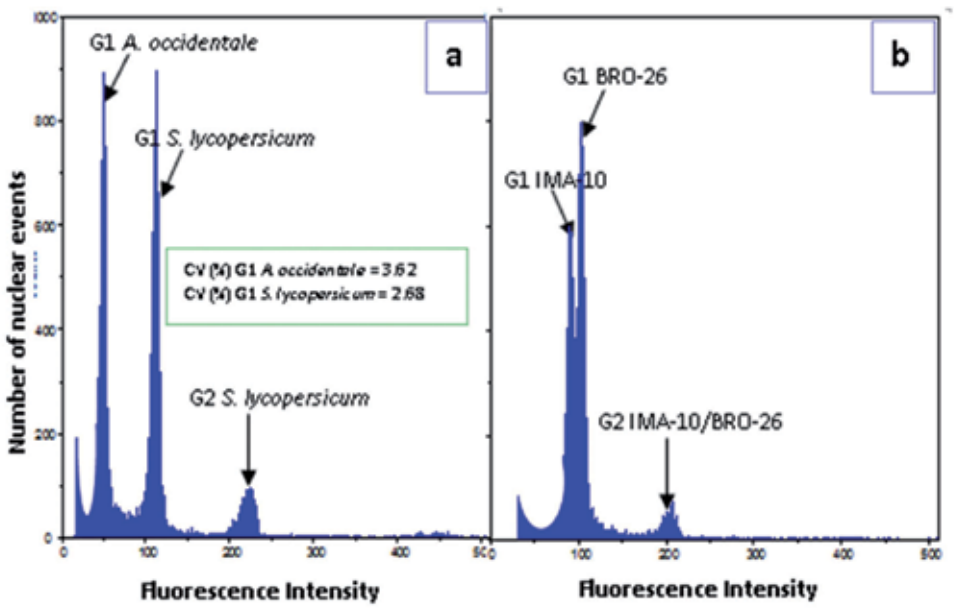


Fig. 7. (a): Flow cytometric histogram showing relative DNA content between G1 nuclei of cashew (*A. occidentale*) and tomato (*S. lycopersicum*) (tomato as internal reference standard). (b): Histogram showing suspected intraspecific variation in relative DNA content between two accessions (IMA-10 and BRO-26) of cashew (*A. occidentale*) with different phenotypic characteristics (see Aliyu, 2010).

intraspecific variation. However, all the sampled accessions showed consistent histogram peak position throughout the measurement, suggesting stability of the genome ploidy and the likelihood that the cashew, *A. occidentale* are predominantly diploids. Slight variation in genome size distribution tends to correlate with the history and/or source of origin/introduction and ecological adaptation of the accessions (Aliyu, 2010). Further studies of the cashew genome size more importantly with a more sensitive dye like Propidium Iodide (PI) may provide clearer information on the absolute genome size, likelihood of intraspecific variation and genome stability of this tropical tree species.

10.2 Chromosome counts in cashew

Aceto-orcein squashing of meristematic tissues (root tips and flower buds) of some genotypes has been analyzed for the number, structure and behavior of chromosome in cashew. Although there are divergent opinions on the ploidy status of the *occidentale* species, the study revealed that most genotypes had 42 chromosomes (Aliyu and Awopetu, 2007c) and could be tentatively described as $2x = 42$ for a diploid species. And that cashew karyotypes are usually symmetric and less divergent, with mainly metacentric and submetacentrics pairs (Aliyu and Awopetu, 2007c). There were similarity in the morphology, number and behavior of the chromosomes among genotypes from different sources of introduction, which could imply common progenitors and slow mutation rate (chromosomal) among the existing *occidentale*. Another possible explanation for the chromosomal congruity can be drawn from the outcrossing nature of the plants, thereby enhances exchange of DNA materials between close and distant sympatric relatives and overlap in phenotypic features between lineages. Such phenotypic similarity between genotypes permits cross-compatibility (free gene exchange) with improved adaptation.

11. Conclusions and future outlook

11.1 West Africa regional cashew improvement programme

With the establishment of African cashew alliance (ACA), a body responsible for the improvement in the value chain in the African cashew industry in 2006, the investments and production of cashew in West Africa region has been on the steady increase. West African countries like Benin, Burkina Fasso, Cote d'Ivore, Gambia, Ghana, Guinea-Bissau and Senegal, which have hitherto not been active in cashew production, are now actively investing in the industry. Unfortunately, as indicated in the report of Sustainable Tree Crops Programmes in 2001, collectively the region lacks expertise and improved varieties to support a sustainable cashew industry. Hence, there is the need to evolve a well funded and coordinated regional cashew improvement programme to aggressively develop improved planting materials for the farmers to boost production and improve the livelihood of the rural households. To achieve this, there is the need to carry out extensive survey of cashew growing areas, document the existing genetic resources and evaluate (characterization) these materials for useful agronomic and yield traits using a set of standardized descriptors across the region. With massive production of cashew in Nigeria (highest in Africa), extensive germplasm, and about four decades of research in cashew by the Cocoa Research Institute of Nigeria (CRIN) and existing research linkages with other sister institutes in Ghana and Cote D'Ivore, Nigeria is better positioned to serve as a platform for such regional

effort. And for smooth implementation and meaningful impact on the rural people, such initiative should be in partnership with non-governmental organizations (NGOs) like Sustainable Tree Crops Programme (STCP), Common Funds for Commodities (CFC), African Cashew Alliance (ACA) and other relevant stakeholders in the industry. Similar effort has been used to reposition cashew industry in the East Africa in the recent past.

In addition, the proposed regional programmes should include, establishment of polyconal seed gardens across cashew growing communities for easy access to improved planting materials, development of hybrid cashew that will out-perform the existing cultivars and routine training and capacity building for all the stakeholders to boost production, enhance value addition and above all, improve standard of living of cashew small holders.

12. References

Akinwale, S. A., & Esan, E. B. (1989) Advances in Cashew Breeding in Nigeria. In: Progress in Tree Crop Research. 2nd Edition. Cocoa Research Institute of Nigeria, Ibadan, Nigeria. 166-174.

Aliyu, O. M., & Awopetu, J. A. (2005). *In-vitro* regeneration of hybrid plantlets of cashew (*Anacardium occidentale* L.) through embryo culture. African Journal of Biotechnology; 4 (6) 548-553.

Aliyu, O. M., & Awopetu, J. A. (2007a). Multivariate analyses of cashew (*Anacardium occidentale* L.) germplasm in Nigeria. Silvae Genetica, 56 (3-4): 170 – 179.

Aliyu, O. M., & Awopetu, J. A. (2007b). Assessment of genetic diversity in three populations of cashew (*Anacardium occidentale* L.) using protein-isozymes-electrophoretic analysis. Genetic Resources and Crop Evolution, 54: 1489 – 1497.

Aliyu, O. M. & Awopetu, J. A. (2007c). Chromosome studies in cashew (*Anacardium occidentale* L.). African Journal of Biotechnology Vol. 5 (2): 131-136.

Aliyu, O. M. (2006). Phenotypic correlation and path coefficient analysis in cashew (*Anacardium occidentale* L.). Silvae Genetica, 55 (1): 19-25.

Aliyu, O. M. (2007). Pollen-style compatibility in cashew (*Anacardium occidentale* L.). Euphytica, 158: 249 – 260.

Aliyu, O. M. (2008). Compatibility and fruit set in cashew (*Anacardium occidentale* L.). Euphytica, 160:25–33.

Aliyu O. M. (2010). Intraspecific genome size variation and phenotypic (nut and leaf) traits polymorphism in cashew (*Anacardium occidentale* L.). In: V.A. Chikaleke, O.A. Adetula and S.A. Olakojo (eds). Genetics and sustainable livelihood in a developing economy. *Peer-reviewed* Proceedings of the 34th Annual Conference of Genetic Society of Nigeria (GSN), National Horticultural Research Institute, Ibadan, 19th -24th Spetember, 2010. pp 221 -229.

Archak, S., Gaikwad, A. B., Swamy, K. R. M., & Karihaloo, J. L. (2009). Genetic analysis and historical perspective of cashew (*Anacardium occidentale* L.) introduction into India. Genome 52 (3): 222-230.

Azevedo, D. M. P., Crisóstomo, J. R., Almeida, F. C. G., & Rossetti, A. G. (1998).
Estimates of genetic correlations and correlated responses to selection in cashew (*Anacardium occidentale* L.) Genetics and Molecular Biology (online) vol. 21 n. 3 doi: 10.1590/S1415-47571998000300019.

Croxford, A. E., Robson, M., & Wilkinson, M. J. (2006). Characterization and PCR multiplexing of polymorphic microsatellite loci in cashew (*Anacardium occidentale* L.) and their cross-species utilization. Molecular Ecology Notes 6 : 249-251.

Das, S., Jha, T. B. & Jha, S. (1996). *In vitro* propagation of cashew nut. Plant Cell Reports, 15: 615-619.

Faenza, V., Conticini, L. & Partel, L. (1982). Le collezioni nazional e internazional di Anacardio in Tanzania. Riv Agric Subtrop Trop 76:191–219.

Fetuga, B. L., Babatunde, G. M., Ekpenyong, T. B. & Oyenuga, V. A. (1975). The feeding stuff potential of cashew nut scraps kernel meal. Proceedings of the Conference of Animal Feed of Tropical and Sub-tropical origin. Tropical Products Institute, London (UK), pp 201–207.

Food and Agriculture Organization (FAO) (2011). http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor Website visited June 25-26, 2011.

Freitas, B. M. & Paxton, R. J. (1996). The role of wind and insects in cashew (*Anacardium occidentale* L.) pollination in North Eastern Brazil. Journal of Agricultural Science 126: 319–326.

Jha, T. B. (1988). *In vitro* morphogenesis in cashew nut (*Anacardiumoccidentale* L.). Indian Journal of Experimental Biology, 26: 505-507.

Johnson, D. (1973). The botany, origin and spread of cashew (*Anacardium occidentale* L.). Journal of Plantation Crops, 1: 1-7.

Manchester, S. R., Wilde, V. & Collinson, M. E. (2007). Fossil cashew nuts from the Eocene of Europe: Biogeographic links between Africa and South America. International Journal of Plant Science, 168(8):1199–1206.

Masawe P. A. L., Cundall E. P. and Caligari P. D. S. (1999). Studies on genotype-environment interaction (GxE) in half-sib progenies of cashew (*Anacardium occidentale* L.) in Tanzania. *Tanzanian Journal of Agricultural Science*, 2 (1): 53-62.

Masawe, P. A. L. (1994). Aspect of breeding and selecting improved cashew genotypes (*Anacardium occidentale* L.). Ph. D. Thesis, University of Reading, UK.

Mitchell, J. D. & Mori, S. A. (1987). The cashew and its relatives (*Anacardium, Anacardiaceae*). Memoirs of New York Botanical Gardens English 42: 1–76.

Mneney, E. E. & Mantell, S. H. (2001). *In vitro* culture of cashew (*Anacardium occidentale* L.). Tanzanian Agricultural Research and Training Newsletter, Vol. XVI, Nos. 1-4: 4-10.

Murashige, T. & Skoog, F. (1962). A revised medium for rapid growth and bioassays in tobacco tissue cultures. Physiol. Plant. 15: 473-497.

Nakasone, H. Y. & Paull, R. E. (1998). Tropical Fruits. CAB International, Oxford, UK.

Nayar, K. G. (1998). Cashew: A versatile nut for health. Proceedings of International Cashew and Coconut Conference – 1997 – Dar es Salaam. Tanzania, pp 195–199.

Northwood, P. J. (1966). Some observations of flowering and fruit-setting in cashew (*Anacardium occidentale* L.). Tropical Agriculture 43: 35–24.

Ohler, J. G. (1979). Cashew. Koninklijk instituut voor de Tropen; Teskin, Zutphen Co. Amsterdam, The Netherlands.

Samal, S., Rout, G. R. & Lenka, P. C. (2003). Analysis of genetic relationships between populations of cashew by using morphological characteriza tion and RAPD markers, Orissa, India, Plant Soil & Environment 49(4): 176-182.

Sanwo, J. O. (1973). Germplasm collection. 1972/73. Annual Research Report of the Cocoa Research Institute of Nigeria, Ibadan, Nigeria. 100-110.

Togun, A. (1977). A review of the prospect of cashew industry. Paper published the Cocoa Research Institute of Nigeria, Ibadan, Nigeria. pp. 39.

Topper, C. P., Caligari, P. D. S., Camara, M., Diaora, S., Djaha, A., Coulibay, F., Asante, A. K., Boamah, A., Ayodele, E. A., Adebola, P. O. (2001). West African regional cashew survey covering Guinea, Guinea Bissau, Cote d'Ivore, Ghana and Nigeria. Sustainable Tree Crop Programme (STCP) Report No. BHA-01109. Biohybrids Agrisystem Ltd. P. O. Box 2411, Earley, Reading RG65FY, U.K.

Venkataramah, T. M. (1976). Cashew nut production. Seminar paper published by the Cocoa Research Institute of Nigeria, Ibadan, Nigeria. p. 39.

Woodroof, J. G. (1967). Tree nuts: Production, Processing and Products. Vol. 1. Av. Publ. Co. Incorporation, U.K.

10

Defining Genetic Diversity in the Chocolate Tree, *Theobroma cacao* L. Grown in West and Central Africa

Peter Osobase Aikpokpodion
Department of Genetics & Biotechnology, Faculty of Science
University of Calabar, Cross River
Nigeria

1. Introduction

The cacao tree, *Theobroma cacao* L., a diploid fruit tree species (2n = 20) is the source of dried cocoa beans used as the main raw material in the manufacture of chocolate, confectioneries and some cosmetics product. Although native to the humid tropical regions of the northern parts of South America and the northern parts of Central America (Bartley 2005; Cheesman 1944; Cuatrecasas 1964, Motamayor 2008), the largest cultivation of cacao, an under-storey forest tree species takes place in West and Central Africa. Originally designated a member of the Sterculiaceae family (Purseglove, 1974), *Theobroma cacao* was recently re-classified into the Malvaceae plant family (Alverson et al., 1999). Since its first introduction in the early 19th century by the Portuguese and the Spaniards, the West and Central African region has become the largest producer accounting for some 70% of the world's cocoa output (Figure 1) of more than 3.632 million metric tons (ICCO, 2010). The main producing countries are Cote d'Ivoire (43 % of global production), Ghana (14 %), Nigeria (6 %) and Cameroon (5 %) followed by Togo, Gabon, Sao Tome, Equatorial Guniea, Sierra Leone, Congo and Liberia. The major market for cocoa beans include The Netherlands, United States of America, United Kingdom, France, Germany, Spain, Italy, Japan, China and India.

With the aim of securing the future of the world's cocoa economy, and avoid the situation in the 1930s when, due to limited genetic variability in cacao collections, the swollen shoot virus disease almost destroyed the industry, the genetic structure of cacao collections in West Africa was determined using microsatellite markers.

1.1 Economic importance

1.1.1 In consuming countries

The importance of cocoa beans in the running of the multi-billion dollar annual earning chocolate and confectionery industries cannot be over-emphasised. The world grinding of cocoa beans in 2010/2011 season alone was estimated at 3.698 million metric tonnes (ICCO 2010). The world's exports amount to some US$5–6 billion/year and use of cocoa and cocoa butter in chocolate manufacturing, cosmetics, and other cocoa products drive approximately

US$70 billion market and provides over 60,000 jobs in the US alone (Guiltinan 2007). In an annual list of the top 100 global confectionery companies based on net sales in 2010 alone (Table 1), the top ten chocolate confectionery companies accounted for at least US$ 67.59 billion (Candy International, 2011). According to the Association of chocolate, biscuit and confectionery industries of Europe (CAOBISCO) based in Brussels, some 1800 companies with 245,000 direct employees are involved in use of cocoa beans in manufacturing of their products. These industries account for more than 47.8 billion Euros annual turnover, a production of 14.1 million tonnes of products and some 4.1 billion Euros of exports, that is, 10% of the total value of food exports from the European Union. The European chocolate and confectioneries industry which utilizes 50% of the world production of cocoa beans also consume some 30% of the European production of sugar, 35% skimmed milk powder at full EU price as well as a large share of the glucose, butter, wheat, eggs and dried fruit produced in the European Union.

1.1.2 In West and Central African producing countries

Cocoa production is predominantly a smallholders' enterprise in Africa with several hundred-thousand families depending on this cash crop for their livelihood and significant foreign exchange earnings for producing countries (Rice and Greenberg, 2000, Motamayor et al., 2008). Revenue derived from sale and export of crops such as cocoa provides crucial support to livelihoods of farmers in developing countries in Africa and can be a sustainable means of helping millions of households live above poverty and hunger. In West and Central African countries, domestic economies revolve around subsistence agriculture, especially from the sale of products from cash crops such as cocoa. Even in countries such as Nigeria where most of budgetary revenues come from sale of crude oil, revenue from export of cocoa beans makes significant contribution to the nations' gross domestic product (GDP). The economic growth of many of the Least Developed Countries is closely

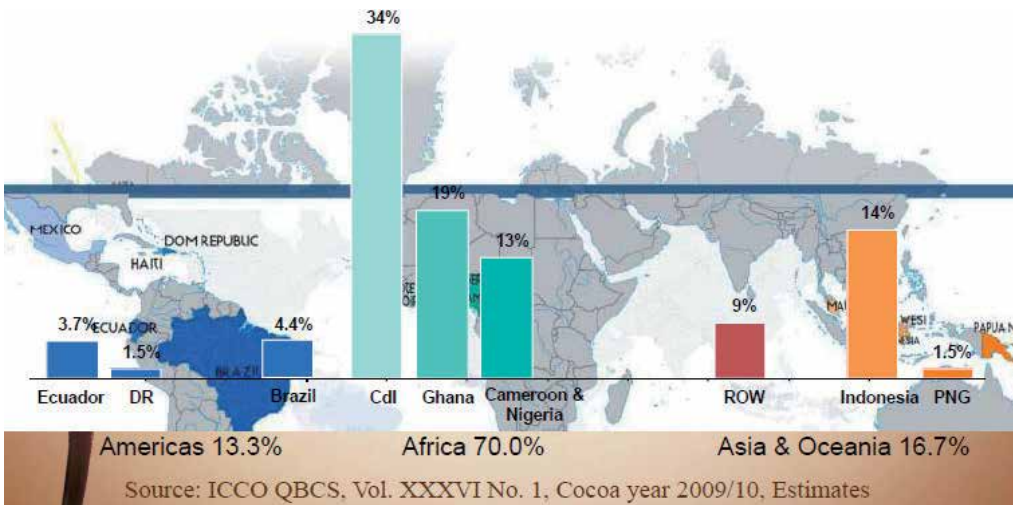


Fig. 1. World's cocoa production estimates (in percentages) according to region and countries provided by International Cocoa Organization (ICCO) for 2009/2010 crop year.

	Company	Country	Net Sales 2010 (US$ millions)
1	Kraft Foods Inc	USA	16,825
2	Mars Inc	USA	15,000
3	Nestlé SA	Switzerland	11,265
4	Ferrero Group	Italy	8,763
5	Hershey Foods Corp	USA	5,703
6	Chocoladefabriken Lindt & Sprüngli AG	Switzerland	2,602
7	Yildiz Holding	Turkey	2,180
8	August Storck KG	Germany	2,000
9	Arcor Group	Argentina	1,650
10	Meiji Holdings	Japan	1,599

Table 1. List of the top ten global confectionery companies that manufacture some form of chocolate by net confectionery sales value in 2010 (Reference: *Candy Industry*, January 2011)

linked with cocoa production, as well as other primary commodities. Many producer countries depend on cocoa exports for a large part of their foreign exchange earnings and government revenue. In Cote d'Ivoire, the largest world producer with more than 1.3 million metric tonnes (mT), cocoa contributes more than 20% of government revenue. When international cocoa prices are low, governments have difficulties meeting debt service obligations and are unable to make much needed investments in basic health, education and infrastructure. In Cote d'Ivoire, more than three of some six million people are engaged in the cocoa sector are small scale farmers. Cocoa alone makes 35 *per cent* of total export estimated at US$10.25 billion in 2010 and 15 % of the 28.2 % agriculture's contribution to the GDP estimated at US$22.82 billion in 2010 (http://www.indexmundi.com/cote_d_ivoire.html). Together with coffee, cocoa is referred to as the 'Green Gold" because of its immense contribution to the economy. It provides job for 60 % of working population and accounts for some 46 % of total export, more than a third of the nation's GDP (http://www.new-ag.info/en/country/profile.php?a=891). In Ghana, the second largest producer with more than 700,000 mT, cocoa is the primary cash crop providing about one-third of all export revenue. With higher commodity prices, gold and cocoa were the two top export revenue earning sectors for Ghana where GDP was estimated at US$38.24 billion in 2010. Cocoa remains the mainstay of Ghana's economy accounting for 40 % of agricultural exports and 12 % of country's GDP (http://www.theodora.com/wfbcurrent/ghana-economy.html). In 2007 for instance, cocoa contributed 35 % of Ghana's GDP and 60 % of employment in agriculture (Centre for the Studies of African Economies (CSAE), 2009). In spite of commencement of oil production in Ghana, agriculture, especially the cocoa sector would remain the key to rural transformation of the economy. In Nigeria, cocoa provides means of livelihood to more than five million people. Although heavily dependent on oil, agriculture contributes significantly to the economy with about 70 % of the population engaged in agriculture. The cocoa sector accounts for some 27 % of the 41.48 % of GDP attributed to agriculture. Cocoa is the single largest non-oil export earning commodity for Nigeria. In comparison with other agricultural commodities, cocoa makes the largest non-oil contribution to the nation's economic development and accounted for 65% of total agricultural export in 2004 (Aikpokpodion, 2007). In Cameroon, it is estimated that some four million people depend on cocoa and coffee for their livelihood. Most of the

cocoa is produced primarily in central southern Cameroon by millions of small scale farmers. In Cameroon, cocoa cultivation is currently one of the major sources of revenues of rural households (1 to 2 millions of people) of the forest agro-ecological zones in the country (South and South-Western parts). Cocoa is grown in more than 200 000 farms and the total cocoa growing surface is estimated to be 400 000 hectares (Efombagn et. al., 2006). In other countries such as Togo, Sierra Leone, Liberia, Equatorial Guinea, Sao Tome and Principe, Gabon and Democratic Republic of Congo, cocoa production makes significant agricultural contribution to the GDP.

2. Cacao introduction history and genetic materials

Traditionally, cacao types cultivated are subdivided into three major '*genetic*' groups: Forastero, Criollo (domesticated by the Amerindians in Central America), and Trinitario (hybrids between Forastero and Criollo, originating from Trinidad). While the Forastero trees are vigorous and more resistant to diseases, the Criollo trees are poor yielding and highly susceptible, although Criollo trees produce high premium quality beans with aromatic flavor. The Brazilian cacao of the Amelonado type (Lower Amazon Forastero) was first introduced by the Portuguese into Principe around 1822, and reached Sao Tomé in the 1850s (Bartley 2005). According to Nosti, quoted by Toxopeus (1964), it was from this collection that the Spaniards brought cacao into the Island of Fernando Po (now Bioko), Equatorial Guinea in 1854. This collection in Fernando Po became the major source of cacao introduced into mainland West Africa at several times by many persons including traders and migrant workers, agencies, missionaries among others. Available records showed that cacao was introduced from Fernando Po by workers and traders like Squiss Ibaningo into Nigeria in 1874, Tetteh Quarshie into Ghana in 1878, and Cote d'Ivoire in 1879 (Edwin and Masters 2005; N'Goran et al. 1992; Opeke 1969). Missionaries like the Basel missionaries, Royal Botanical Garden curators, colonial administrations played significant roles in the introduction of cacao types from different origins into the mainland West and Central Africa. These earlier introductions from Fernando Po formed the initial basis of cacao grown in West Africa, and was referred to as the "West African Amelonado". During the late nineteenth century, the Colonial administration also introduced some red-podded cacao materials from British West Indies into botanical gardens established in Aburi (Ghana) and Lagos (Nigeria) (Toxopeus 1964). By 1910, Ghana, followed by Nigeria, had become one of the largest producing countries, thus making the West Africa sub-region an important growing area critical to the sustainability of the world's cocoa economy, a status it still maintains today.

The introduction of cacao germplasm into island and mainland of Africa took place in response to two main waves of idea which naturally divided cacao germplasm introduction into: 1. Exploratory Colonial Period (1822 – 1909), and, 2. Expansionary Experimental Pre- and Post-Independence Period (1910 – 2010). During the first era which spanned early 19th to the end of the 19th century, cocoa seeds and plants were transported in barrels and shipments across the sea from the northern parts of Southern America and Central America to Africa. This was in response to the then imbibed and appreciated Aztec and Mayan culture of drinking '*chocolatl*', 'The Food of the Gods". The favourable similar tropical humid climatic conditions and abundant rainforest vegetation provided impetus for this transatlantic exploratory introduction which fortunately gave good results. The cultivation

of cocoa was therefore successfully established along the rainforest belt of West and Central African countries with significant economic revolution for both producing countries and chocolate lovers especially in the North. A timeline of earlier germplasm introduction efforts during the first era is presented in Table 2 below. The first materials introduced were of Amazonian origin, unlike the Criollo varieties introduced to Asiatic and Oceanic regions. The germplasm established on the island of Principe in 1822 was the main basis of the cocoa industry on the island. Original planting was said to have consisted of 30 plants which most likely were taken from a single fruit (pod). Progenies of these trees provided seeds for planting other areas of the island (Bartley 2005). Timeline of earliest cacao introduction into island and mainland West and Central Africa in the 19th and early 20th century

Date	Germplasm material & description	Origin	Remarks
Island of Sao Tome & Principe			
1822	Amazonian Bahia 'Comum' or 'Amelonado'	Brazil: States of Bahia, Espirito or Rio de Janeiro	Self-compatible and homozygous variety. Only 30 plants presumably from a single fruit (pod) was established on the island of Principe.
1840	-	-	First export of 'sizeable amount' of cocoa beans from the island of Principe.
1850	Amazonian Bahia 'Comum' or 'Amelonado' later called Sao Tome 'Creoulo'	Principe	Cocoa cultivation began in Sao Tome with seeds taken from the 'Amelonado' plantings in Principe.
1880	Non-Amazonian 'other' varieties: 1. Red fruits with smooth surfaces ('Criollo') known as Venezuela 'Caracas' 2. Narrow elongate fruits with pronounced ridges 3. 'Guayaquil' variety type 4. Very large red fruits, slightly ridged and smooth surface 5. Green fruits 6. Hybrids formed from the different types 7. 'Laranja' mutant	Ecuador, Trinidad and Venezuela	Period of active expansion of cultivation, Hybrids evolved from among the different varieties and between the non-Amazonian varieties and original 'Comum' types.

Equatorial Guinea (Fernando Po)			
1854	Sao Tome 'Creoulo' and later, the 1880's Non-Amazonian 'other'varieties.	Sao Tome	Establishment of cocoa in Fernando Po (now Bioko) with seeds from Sao Tome
Cameroon			
1876	1. Unknown materials likely from Trinidad.	Royal Botanic Gardens, England	13 plants shipment by a British missionary on Cameroon Mountains.
	2. Sao Tome 'varieties'	Sao Tome	Introduction by Preuss, Curator of Victoria Botanic Garden. Some ofthese materials became known as 'Victoria-Kakao' variety of Sao Tome
1895?	332 plants including 'Forastero'	Trinidad	Gosselin (1895), Preuss (1901)
1900	Several varieties including 'Forastero', 'Criollo', 'Puerto-Cabello', 'Venezuela', 'Maracaibo', 'Guayaquil', 'La Guira', 'Soconusco', 'Suriname' and 'Nueva Grenada' from Colombia	South and Central America (Trinidad)	Introduction by Preuss. At the beginning of the 20th century, Cameroon had established the most diverse collection of cacao varieties (Bartley, 2005)
Ghana			
1857	Seeds from Suriname	Surinam	Unsuccessful attempt by the Basel Missionaries.
1861	'Amelonado'	Sao Tome & Principe	Partial success by the Basel Missionaries.
1878? 1887	'Amelonado'	Fernando Po/Sao Tome	Tetteh Quarshie brought seeds from Fernando Po. Governor Griffiths also made introduction in 1887.
1900 - 1901	1. 'Cundeamour', 'Pentagonum' from Nicaragua	Trinidad	During this period, introduction was to obtain varieties higher in quality than the widely grown 'Amelonado' from Fernando Po.
	2. 'Red Forastero', 'Criollo', 'White variety' of 'Caracas' type from Jamaica	Royal BotanicGarden, England	
1903	'Ocumare', 'Trinidad	Trinidad	Introduced plants were

	Criollo', Nicaraguan Criollo', 'Red Criollo' and 'Yellow Criollo'		established in Aburi Botanic Garden. These materials were largely self-incompatible. These are the likely ancestors of red fruited trees in Ghana.
Nigeria	*Regional differences in genetic composition of populations likely due to different routes of introduction*		
	A. The Niger Delta Protectorate covering Midwestern to Calabar Region		
1874	'Amelonado', Sao Tome 'Creuolo'	Fernado Po	Chief Squiss Ibaningo, a migrant worker introduced cocoa into Bonny, now in River State. This variety constituted virtually all the planting material cultivated at the middle of the 20th century.
1899	Unknown varieties. Forastero?	Royal Botanic Garden, England	Up to two shipments of cocoa plants
1900	'Pentagonum' variety	Trinidad?	Variety found at Old Calabar Station
1905	Non-Amelonado 'types'	Trinidad? SaoTome?	These were brought into Old calabar Station
1909	60 fruits of 'Forastero' type	Trinidad	Barrel shipment from Trinidad
	B. The Lagos Colony		
1880?	Sao Tome 'Comum' (Amelonado)	Fernando Po	First planting of cocoa in Lagos of some 1,500 plants. The plantation near Agege was owned by JPL Davies (Webster 1964)
1877-1888	Unknown.	Fernando Po Trinidad	Most likely non-Amelonado plants sent to Ceylon
Cote D'Ívoire			
1880	Amelonado	Fernando Po	First cocoa introduction
Sierra Leone			
Pre-1900	Sao Tome 'Comum' (Amelonado); Cameroon variety, Victoria-Kakao	SaoTome, Fernando Po and Cameroon	Several Sierra Leoneans were influential farmers in Fernando Po at the end of the 19th century.
1902	-60 plants of 'Ceylon Red',	Trinidad	

	Nicaraguan Criollo', 'Forastero' and *T. pentagonum*		
	- 1000 seeds of unknown variety	Fernando Po	
Liberia 1861?	Sao Tome 'Comum' (Amelonado)	Sao Tome	Many Liberians were contract workers in Sao Tome.

Table 2.

(Bartley 2005). By 1840, some quantities of cocoa was exported from the island. The cultivation of cocoa spread to the main island of Sao Tome in the 1850s. The variety which became known as Sao Tome "Creoulo", was self-compatible and homozygous and mostly related to the 'Comum'variety in Bahia, Brazil. This variety was taken both directly and indirectly through Fernando Po to other inland West and central African countries and became the basis of cocoa grown there. However, some other varieties were also introduced into Sao Tome from Ecuador, Trinidad and Venezuela in 1880. Consequently, the bulk of cocoa grown on farmers' plantation must have consisted of a mixture of these earlier varieties, but due to differential expression of self-incompatibility systems, the self-compatible 'West African Amelonado' types must have dominated in the complex mixture of cacao of diverse origin at the beginning of the 20th century.

During the second era which began at about the end of the 19th century and beginning of the 20th century, economic considerations for higher income and premium due to greater yields and higher bean and chocolate quality were the main reasons for germplasm introduction. Previously selected individuals (clones) rather than 'types' showing potentials for high yields, resistance or tolerance to pests, diseases and abiotic stress such as drought were introduced and engaged in cultivar development processes on experimental stations. During this last decade, however, the "People, Planet and Profit'' concept of *Sustainability* has become a significant factor in cacao germplasm introduction. This has bearing with the concept of "Preventive Breeding" where clones showing resistance to regionally important diseases of cocoa growing regions could be introduced through international intermediate quarantine centers. This was to ensure that in the unlikely case of disease spread, for example, witches broom from South America to Africa, there is present in the African germplasm collections, sources of resistance to cope with the new disease in order not to paralyze the local cocoa economy as is the case during any outbreak.

Since the first successful introduction of '*Amelonado*' cacao, Lower Amazon Forastero type into West Africa in the late 19th century, there has been series of additional germplasm introductions as reviewed by (Bartley, 2005; Aikpokpodion, 2009). In Nigeria for instance, since formal selection and germplasm conservation programs around 1931 at the Nigerian Department of Agriculture in Moor Plantation, Ibadan there has been. Further germplasm introduction of Trinitario and Criollo selections from Trinidad and Ceylon (now Sri Lanka) (Jacobs et al., 1971). The British West African Colonial Administration established the West African Cocoa Research Institute (WACRI) in 1938 with headquarters in Tafo, Ghana and a mandate covering Gold Coast (Ghana), SierraLeone, Nigeria and Liberia. Several materials

belonging to Upper Amazon Forastero and Trinitario populations were introduced from Trinidad by WACRI in 1944 (Toxopeus, 1964). Efforts to increase genetic variability in the base population in response to outbreaks of disease epidemic had provided impetus for germplasm introduction into Africa. For instance, the outbreak of cocoa swollen shoot disease in the 1930s in Ghana, Togo, and Nigeria almost destroyed the cocoa industry due to insufficient genetic variability in the base population. Consequently, new introductions were made in 1944 from Upper Amazon Forastero materials collected by F. J. Pound into the West African Cocoa Research Institute headquarters in Tafo, Ghana and Ibadan in Nigeria (Aikpokpodion et al., 2009). Due to the precocity of these materials, they were widely distributed for replanting of cut out plantations and by late 1950s, some 11 selected Upper Amazon types have been used to produce second and third generations of Amazon known as "F_3 Amazon" or "Mixed Amazon" distributed to farmers (Knight and Rogers 1955). By 1961, some 60,000 ha in Ghana and an estimated 21 million seedlings had been distributed by the government of the Western Region to plant some 9,500 ha in Nigeria (Aikpokpodion 2009). Several hybrid varieties involving crosses with local Amelonado, Trinitario, and some Criollo materials were also developed from these materials in Ghana (Lockwood and Gyamfi 1979), Nigeria (Atanda and Jacobs 1974), and Cote d'Ivoire (Besse 1975; N'Goran et al. 1992).

3. Concept and relevance of genetic diversity

Although all members of a species have certain traits in common, individual members may vary significantly. While some of these may be environmental, a significant proportion is genetic. Genetic diversity is of fundamental importance in the continuity of a species as it provides the necessary adaptation to the prevailing biotic and abiotic environmental conditions, and enables change in the genetic composition to cope with changes in the environment. The first requisite study in the survival of a species is knowledge about the level of genetic diversity (Van Delden, 1992). This refers to the determination of the number of polymorphic loci, the number of alleles, genetic architecture and spatial distribution of genetic variants. Genetic diversity in a population is preserved when the population is in Hardy Weinberg equilibrium. This means that genotypes are present in expected proportions based on the allele frequencies in the population. However, one or two of the conditions for this may not hold. For instance, introduction of genetic variants through mutation or gene flow from a genetically different population will increase the genetic diversity. Assortative mating will mainly affect genotypic proportions. Directional selection against recessives will eliminate alternative alleles and make the population monomorphic at a particular locus. As a principle, balancing selection in favour of heterozygotes will preserve genetic diversity, while inbreeding and genetic drift, on the other hand, lead to loss of genetic diversity.

4. Utility of genetic markers in defining genetic diversity in cacao

Genetic markers are inherited variations that can be used to understand genetic events. These include any gene or other DNA variations that are useful for explaining observed genetic event in a population of interest. The use of genetic markers has been useful in the study of population genetics and evolution by providing methods for detecting genetic

differences among individuals. The majority of genetic markers are variations in DNA at sites that may or may not be part of a functional gene. However, their transmission to offspring follows Mendelian rules for inheritance. There are three main properties of a genetic marker: It must be locus-specific, polymorphic in the studied population and easily genotyped. The quality of a genetic marker is measured by its heterozygosity in the population of interest, and for a molecular marker, its polymorphism information content (PIC) as described by Botstein *et al*. (1980). Genetic markers that have been used in population genetics are grouped into three main classes: phenotypic or morphological markers, biochemical markers or isozymes and molecular markers utilizing variation at the nuclear DNA level.

4.1 Morphological or phenotypic markers

These are phenotypes for which variation observed in the population of interest can be explained by Mendel's law of inheritance e.g. colour variation, growth habit and fruit shape. Traditionally, morphological markers have been used to characterize varieties based on the assessment of a range of phenotypic characteristics. Several studies have been carried out using morpho-agronomic characteristics of the pods, seeds and flowers to elucidate population structure and genetic diversity of cacao populations (Aikpokpodion, 2010; Bekele and Bekele, 1996; Engels, 1992). Although morphological genetic markers proved useful in several cases, they are subject to several limitations. These include subjectivity in the analysis of character, environmental influences, limited diversity (morphological variants) among cultivars, and restriction of characterization of some useful characters to a particular stage of development, such as flowering or fruit ripening, and limitation to only one locus.

4.2 Allozymes

Allozymes are allelic variants of enzymes encoded by structural genes. Enzymes are proteins consisting of amino acids, some of which are electrically charged. Due to change in the net electric charge resulting from mutation, allelic variation can be detected by gel-electrophoresis and subsequent specific enzymatic staining. Usually, two or more loci can be distinguished per enzyme, and they are termed isoloci. Therefore, allozyme variation is also referred to as isozyme variation. The study of genetic variation in plant populations was greatly facilitated by the development of protein-based markers (i.e. allozymes). The primary contribution of allozymes to plant population biology has come from their utilization as neutral (or nearly neutral) genetic markers. Allozymes have been employed to characterize patterns of genetic variation within and among populations, and to examine the processes of dispersal and the patterns of mating that influence levels of genetic differentiation. In cacao, several workers such as Ronning & Schnell (1994) and Warren (1994) used isozyme systems to explore genetic diversity among cacao populations.

4.3 Molecular markers

Use of molecular markers has allowed the complete sampling of the genome, and helped to overcome the limitations of morphological markers and the isozyme markers. DNA markers have been successfully applied in cultivar identification, controlling seed purity of hybrids

and checking the genetic relatedness between cultivars. Some of the techniques developed for DNA manipulation in order to detect variations are:

4.3.1 Random Amplified Polymorphis DNA (RAPD)

In cacao, the first study of the use of molecular markers in cacao was reported by Wilde *et al*. (1992). These authors used Random Amplified Polymorhism DNA (RAPD) to study relationships among cocoa groups. Their work was soon followed by those of several authors who also used RAPD to determine genetic relationships among cacao populations (Russel *et al*., 1993; Laurent *et al*., 1994; Lerceteau *et al*., 1997; Whitkus *et al*., 1998). N'Goran *et al*. (1994) analyzed the genetic diversity of 106 genotypes in Cote d'Ivoire belonging to the various morphogeographic groups within Criollo, using 49 repeatable polymorphic RAPD products. They showed a clear structure among Forastero and Criollo groups with clear differentiation between Upper and Lower Amazon Forastero. Lerceteau *et al*. (1997) analysed the genetic diversity of Ecuadorian Nacional clones, Forastero, Trinitario and Criollo cacao clones using forty-three genomic probes. They found that within-group genetic diversity was almost identical between Forastero, Trinitario and Criollo. Their results showed that the populations of Amazon Forasteros and Criollo studied were highly diverse and that the Criollo and Trinitario populations showed some overlap.

4.3.2 Restriction Fragment Length Polymorphisms (RFLP)

RFLPs are generally found to be moderately polymorphic and can be applied in comparisons ranging from the individual level to closely related species. Because of their high genomic abundance and random distribution throughout the genome, RFLPs have been used by several workers to determine the genetic diversity of cacao populations (N'Goran *et al*., 2000; Lerceteau *et al*., 1997; Motamayor and Lanaud, 2002; Motamayor *et al*., 2002).

4.3.3 Microsatellites

Microsatellites, also known as Simple Sequence Repeats (SSRs) are molecular marker loci consisting of tandemly repeated DNA of short oligonucleotide sequences of two to six bases in length. They form a class of genetic markers that show variation in the number of repeats of a simple DNA sequence. They are extremely common in eukaryotic genome (Tautz and Rentz, 1984) and are highly polymorphic in length. The development and application of SSRs facilitate the acquisition of a large quantity of genetic information relevant to genotype identification, which provides opportunities to characterize germplasm collections (Mitchell *et al*., 1997). Such information generated are used by plant breeders to better understand their germplasm, guide breeding plans and better exploit genetic variation available (Lu *et al*., 2005). Microsatellites are recommended as an international standard for defining genetic identity and has been widely used in the study of genetic diversity of cacao genetic resources (Aikpokpodion *et al*., 2009, 2010; Saunders *et al*., 2000; Zhang *et al*., 2006).

5. Utility of phylogenetic tree in revealing genetic diversity

In the assessment of genetic diversity it is important to define relationships existing within or among sets of germplasm collection of a species and their evolutionary history within or

with related species. Phylogenetic trees serve as extremely powerful tools for organizing and illustrating these relationships. Phylogenetic trees have been successfully used in guiding conservation and biodiversity efforts (Sul et al., 2009) and establish relationship of cacao with its wild relatives (Figuera et al., 1994). A phylogenetic tree is a diagrammatic branching "tree" illustrating evolutionary relationships among entities within a species or various biological species based on similarities and differences in their physical and/or genetic characteristics. Organisms with similar morphologies or DNA sequences are likely to be more closely related than organisms with different structures or sequences. Entities that are joined together in the tree are implied to have evolved from a common ancestor. Each branch point represents the divergence of two species while sister taxa are groups that share an immediate common ancestor. Phylogenetic trees can either be rooted or unrooted. A rooted phylogenetic tree clearly shows relationship of each entity with the (usually imputed) most recent common ancestor of all the entities. Rooted trees are often constructed with the use of a definitive related 'outgroup' taxa. An 'outgroup' is a species or group of species that is closely related to the 'ingroup', the various species being studied. Unrooted trees, on the other hand, depict the relatedness of the entities without making assumptions about their ancestry. The principles of maximum parsimony and maximum likelihood are often used to analyze phylogenetic relationships with computer programs. The principle of maximum parsimony assumes that the tree that requires the fewest evolutionary events (appearances of shared derived characters) is the most likely. The principle of maximum likelihood states that, given certain rules about how DNA changes over time, a tree can be found that reflects the most likely sequence of evolutionary events.

6. Genetic diversity analysis results from West and Central Africa

Prior to recent studies in Cameroon, Cote d'Ivoire, Ghana and Nigeria, there was no useful information on the extent of genetic diversity in the cocoa cultivated in West and Central African countries. These studies were conducted within the framework of the Sustainable Tree Crops Program, a public-private partnership platform endowed by the United States Agency for International Development, US Department of Agriculture and chocolate industry partners such as the Mars Incorporated. These studies assessed genetic diversity in the introduced primary cacao clones and germplasm accessions used to develop improved hybrids distributed to farmers and cacao accessions on farmers' fields across the sub-region. Microsatellite markers were used to assess genetic diversity in these accessions. In addition to the microsatellite studies, some studies were also carried out to determine variation in agro-morphological and phenotypic characteristics of cacao germplasm in farmers and genebank collections.

6.1 Cameroon (Efombagn et al., 2006; Efombagn et al., 2008; Efombagn et al. 2009)

In their study of some 194 cocoa accessions collected in farms in Southern Cameroon during field surveys (Plate 1 d) and 71 Trinitario and Upper Amazon clones available in genebank collections on-station were assessed using 13 SSR markers. The gene diversity, genetic differentiation and genetic similarities were analyzed for the different populations. In total, 282 alleles were detected within all the populations studied (Plate 1c). The farm accessions were strongly differentiated based on their geographical origin, with accessions coming from the East province clustering together with local Trinitario accessions from the

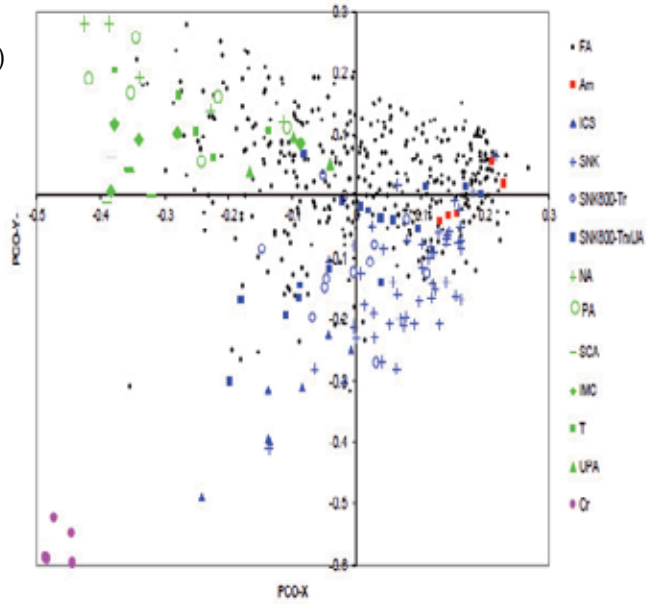
genebank while accessions from the Centre-South provinces clustered with Amazon and hybrid accessions (Plate 1b), suggesting greater uptake of seed garden materials in farms in these provinces. The genetic diversity parameters indicate that the farmers' planting material was not highly diverse, but genetically close to parental genotypes available in genebanks (Plate 1a and 1c). However, some promising Upper Amazon clones (T-clones) that have also been used as parents of released hybrid varieties were genetically distant from the accessions. Their result suggested that the progenies of the Upper Amazon parents have so far been poorly used in the cocoa farms surveyed. A large genetic diversity was observed in the farm (H_{nb} = 0.34 – 0.72) and genebank (H_{nb} = 0.64 – 0.66) materials (Plate 1c). The large variability observed in farmer plantations was attributed to the large variation of first cocoa introductions (Bartley 2005), and the introduction of UA germplasm in the 1950's with its subsequent use in the cocoa breeding program. They also observed a higher private allelic richness in farm genotypes (A_p= 2.03) than those of SNK (Selection of Nkoemvone), ICS (Imperial College Selection), T (Trinidad) and UPA (Upper Amazon) clones. This indicated that farm accessions also harbor some genes that are not present in current national field genebanks. Evidence were found of admixture in farmers' fields which must have been due to hybridization (in seed gardens) and to substantial natural recombination in farmers' fields. Also, since farmers tend to use seeds issued from open pollination in their plantations for new plantings (replacement of dead cacao trees, extension or the creation of new cacao farms), these must have resulted in the presence of admixture as observed in these materials. Surprisingly however, was the observation that there was no relationships between the 'Criollo' reference materials with cacao accessions collected from farmers plantations (Plate 1a).

In a study to determine morphological diversity existing in cacao farms in relation to genetic diversity in gene bank accessions a total of 300 farm accessions (FA) were selected in the two major cocoa producing areas (Southern and Western) of Cameroon. Seventeen quantitative and qualitative descriptors used in this study were related to leaf (flush colour), flower (ligule colour), pod (weight, length, width, apex form, shape, rugosity, colour, husk hardness, basal constriction and pod index) and seed (number, length, width, dry weight and colour) characters. For the qualitative characters evaluated, considerable morphological variation was observed using the Shannon Weaver diversity index (SWDI) within FA and gene bank accessions. Among the FA, a differentiation between southern and western regions was only possible when using quantitative pod traits. Mean quantitative traits values of FA were not too different than those of most gene bank AGs, except for a few traits of agronomical interest (seed weight and pod index). No significant variation was observed for seed traits in all FA groups (southern/western). The morphological structure (quantitative traits) showed spatial differentiation between western and southern FA and a closer relationship between gene bank and some farm accessions.

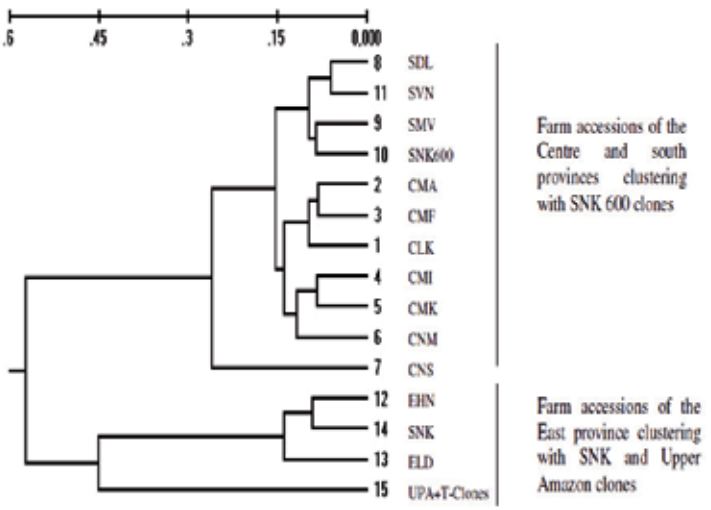
6.2 Cote D'Ivoire (Pokou et al., 2009; Tahi et al., 2008)

Since the introduction of the 'Amelonado' type in 1880 that was widely cultivated in cocoa growing regions (Plate 2a), the first step at genetic improvement of the locally available germplasm took place with mass selection of 'better' types in local farms between 1947 and 1958. Mostly due to low genetic variability available in local types as is the case in other West African countries, the Upper Amazon Forastero types were introduced in 1954. This

a)


FA
Am
ICS
SNK
SNK600-Tr
SNK600-TnUA
NA
PA
SCA
MC
T
UPA
Qr
PCO-X
PCO-Y

b)


.6
.45
.3
.15
0,000
8 SDL
11 SVN
9 SMV
10 SNK600
2 CMA
3 CMF
1 CLK
4 CMI
5 CMK
6 CNM
7 CNS
12 EHN
14 SNK
13 ELD
15 UPA+T-Clones
Farm accessions of the Centre and south provinces clustering with SNK 600 clones
Farm accessions of the East province clustering with SNK and Upper Amazon clones

c)

FAC group I							
Population	**Collection site**	**H_E**	**H n.b.**	**H_o**	**P(0.95)**	**Mean A/L**	**Effective A/L**
Farm accessions							
CLK	Lékie	0.65(0.20)[1]	0.67(0.21)	0.66(0.21)	1.00	6.53	3.92
CMA	Mefou-et-Akono	0.63(0.14)	0.66(0.15)	0.51(0.15)	1.00	5.53	3.53
CMF	Mefou-et-Afamba	0.68(0.16)	0.70(0.17)	0.65(0.20)	1.00	7.61	4.76
CMI	Mbam-et-Inoubou	0.62(0.22)	0.66(0.23)	0.53(0.32)	0.92	5.15	4.53
CMK	Mbam-et-Kim	0.60(0.21)	0.61(0.22)	0.51(0.25)	1.00	9.53	3.61
CNM	Nyong-et-Mfoumou	0.69(0.12)	0.72(0.12)	0.63(0.20)	1.00	6.00	4.46
CNS	Nyong-et-So'o	0.68(0.17)	0.74(0.18)	0.72(0.16)	1.00	5.30	5.30
SDL	Dja-et-Lobo	0.47(0.20)	0.50(0.21)	0.49(0.29)	1.00	4.23	4.07
SMV	Mvila	0.63(0.17)	0.66(0.18)	0.58(0.19)	1.00	6.76	4.30
SVN	Vallée du Ntem	0.57(0.21)	0.58(0.22)	0.54(0.27)	1.00	6.46	4.23
Genebank material							
SNK600	Nkoemvone	0.64(0.15)	0.65(0.15)	0.59(0.21)	1.00	9.00	3.53
UA (UPA)	Nkoemvone	0.66 (0.17)	0.70(0.18)	0.58(0.36)	1.00	5.23	3.43
FAC group II							
Population	**Collection site**	**H_E**	**H n.b.**	**H_o**	**P(0.95)**	**Mean A/L**	**Effective A/L**
Farmers accessions							
EHN	Haut-Nyong	0.34(0.24)	0.35(0.24)	0.16(0.16)	0.92	4.07	2.84
ELD	Lom-et-Djerem	0.61(0.21)	0.62(0.21)	0.55(0.18)	1.00	7.85	3.92
Gene bank material							
SNK	Nkoemvone	0.61(0.14)	0.61(0.14)	0.47(0.22)	1.00	7.38	3.38

H_E: Expected Heterozygosity H n.b.: non biased heterozygosity H_o:observed heterozygosity
P(0.95): proportion of polymorphic loci when the most frequent allele do not exceed 95%
Mean A/L: Mean number of alleles per locus effective A/L: Effective number of alleles per locus
[1]: standard deviation in parentheses

d)

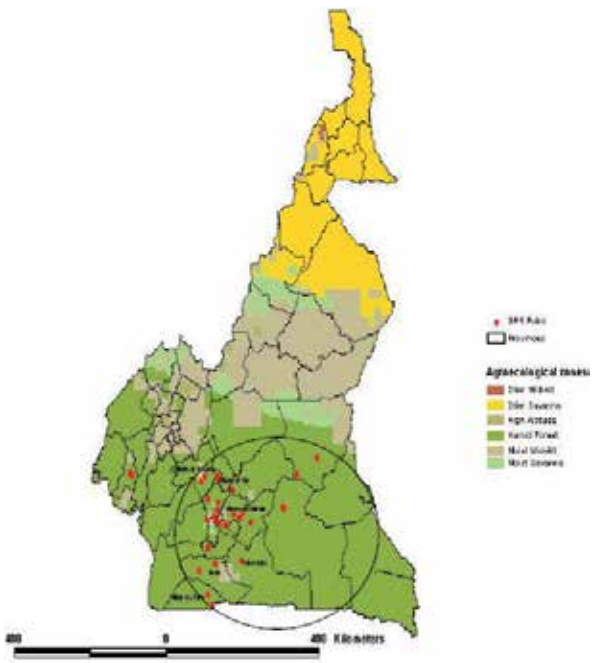
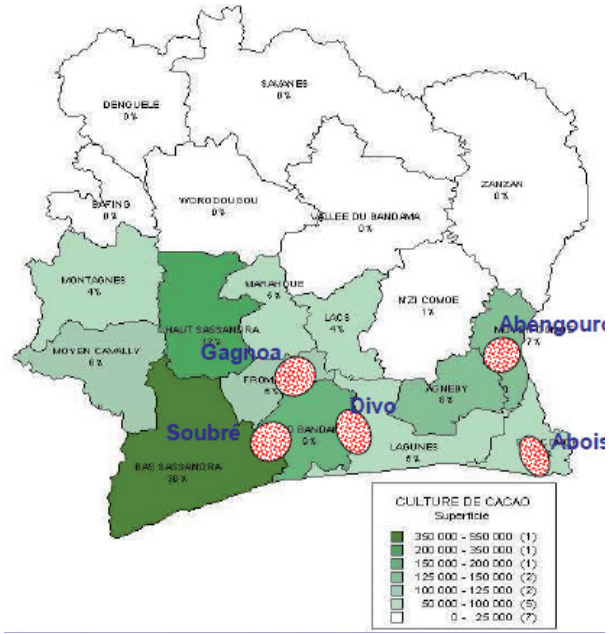


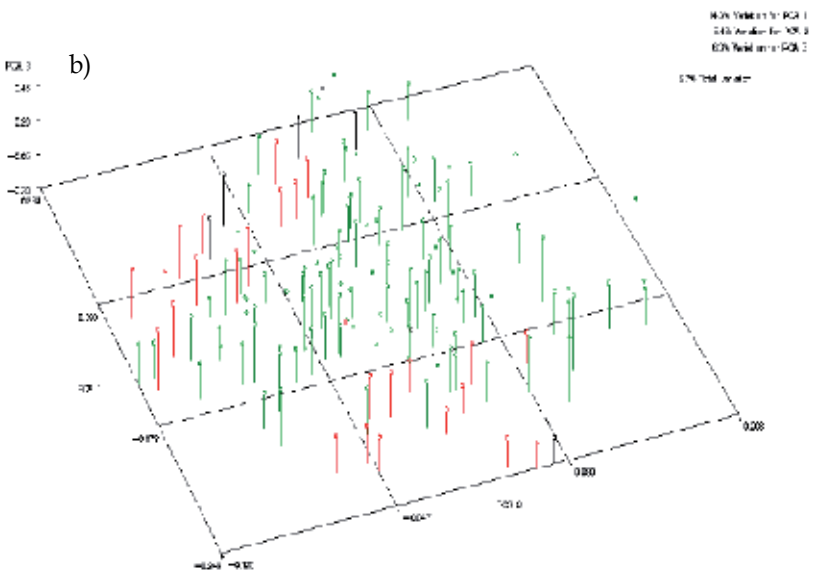
Plate 1. a. A scatter plot showing genetic structure of planting materials in farmers and genebank materials in Cameroon; b. Spatial genetic differentiation between farmers accessions in central and south provinces from accessions in eastern province of Cameroon; c. genetic diversity indices of farmers and genebank accessions of cocoa in Cameroon; d. locations of cacao accessions collected within cocoa producing area of Cameroon used for the study (Source: Efombagn et al., 2006; Efombagn et al., 2009).

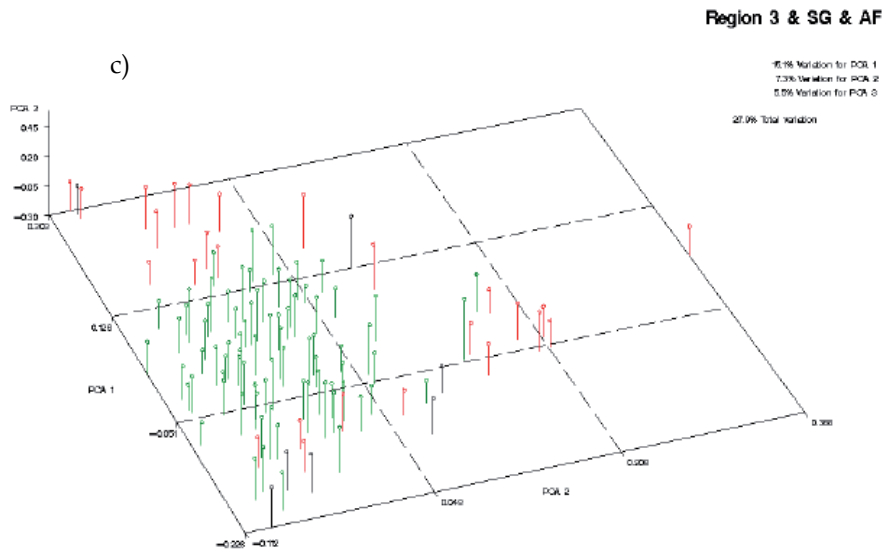
culminated in the development and distribution of selected hybrids between from the 1960s. However, much impact was made with the distribution on the hybrids developed and distributed in 1975 (Besse, 1975). In a survey conducted recently, it was shown that 71 % of materials grown on farms were locally selected 'Amelonado' type, 23 % selected improved hybrid types, while 6 % of the farms were grown with a mixture of local and improved types (Plate 2c).

a)



b)





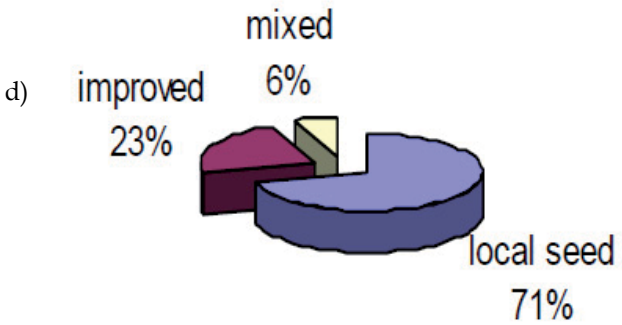


Plate 2. a. Cocoa producing regions in the humid forest of Cote d'Ivoire; b. genetic diversity in cacao accessions in farmers fields in Abengourou (green vertical lines) in relation to seed garden materials (red vertical lines); c. genetic diversity in cacao accessions in farmers fields in Divo (green vertical lines) in relation to seed garden materials (red vertical lines); d. distribution of different cacao types present in farmers fields in Cote d'Ivoire.

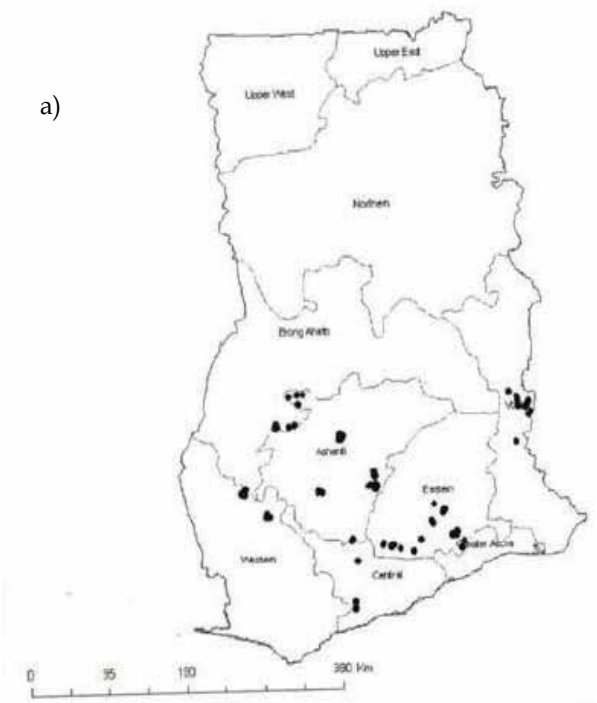
In a study conducted between 2003 and 2005 (Pokou et al., 2009), 12 microsatellites (simple sequence repeats marker) were used to asses genetic diversity of cacao types. Results showed considerable diversity in farmers accessions reflecting largely hybridization between local Amelonado types and Upper Amazon types distributed in the 1970s (Plates 2b & 2c). However, a significant proportion of diversity in seed garden materials was yet to diffuse to the farms indicating that farmers still largely used their own 'selected' materials

against those developed in research institutions. A further analysis was carried of the reciprocal recurrent selection programme set up in 1990. This involved two main genetic groups: Upper Amazon Forastero (UA) and a mixture of Lower Amazon Forastero (LA) and Trinitario (T). Based on data obtained from 12 microsatellite primers, the genetic diversity and genetic distances of the parental populations used in the first and second selection cycles are presented. The results revealed that the diversity of populations UA0 and UA1 on the one hand and (LA+T)0 and (LA+T)1 on the other is similar. The genetic distances were small between the parental populations used for the first and second cycles. Genetic diversity was greater in the UA group than in the LA+T group. The number of rare and of private alleles was reduced for both genetic groups, as well as the number of the frequent alleles in the LA+T group.

6.3 Ghana (Opoku et al., 2007)

In order to assess the genetic diversity of cacao types grown, some 377 accessions including farmers' accessions, breeders' collection and parental clones were collected from all cocoa growing regions of Ghana (Plate 3a), and analyzed using 17 microsatellite markers. Genetic diversity indices indicated that average gene diversity was high in all populations, with mean observed heterozygosity of 0.738 (Plate 3b). Although the highest was recorded in accessions from breeders' and parental collections, genetic diversity in the farmers' collection was comparatively high. Included in the study were a few extant trees among one of the earliest Tetteh Quarshie's introduction in the late 19th century.

a)



b)

Germplasm Collection	H_{ub} (Over all loci)	H_o (Over all loci)	$P_{0.95}$	A	B
Parental Collection					
Upper Amazon (81)[a]	0.764	0.760	1.000	7.1	4.7
Local Amelonado & Trinitario (23)	0.756	0.754	1.000	6.3	3.6
Breeder's Collection					
Series II hybrids & F3 Amazon Progenies (38)	0.757	0.747	1.000	6.7	4.8
Total	**0.759**	**0.754**	**1.000**	**6.7**	**4.4**
Farmers' Collection					
Tetteh Quarshie Farm (7)	0.589	0.526	0.880	3.2	2.4
Aburi Gardens (10)	0.718	0.687	1.000	5.2	4.3
Volta Region (16)	0.753	0.727	1.000	6.7	3.6
Central Region (9)	0.770	0.718	1.000	4.9	3.8
Eastern Region (28)	0.758	0.742	1.000	6.5	3.6
Ashanti Region (51)	0.771	0.764	1.000	10.8	4.4
Brong Ahafo Region (18)	0.743	0.722	1.000	5.9	4.2
Western Region (96)	0.741	0.741	1.000	6.9	4.6
Total	**0.730**	**0.703**	**0.985**	**6.3**	**3.9**
Overall Mean	**0.738**	**0.717**	**0.990**	**6.4**	**4.0**

H_{ub} = unbiased gene diversity (Nei, 1978)

H_o = observed heterozygosity; $P_{0.95}$ = proportion of polymorphic loci when the most frequent allele does not exceed 95%

A = mean number of alleles per locus

B = effective number of alleles per locus

[a] Number of individual per group are indicated in brackets

c)

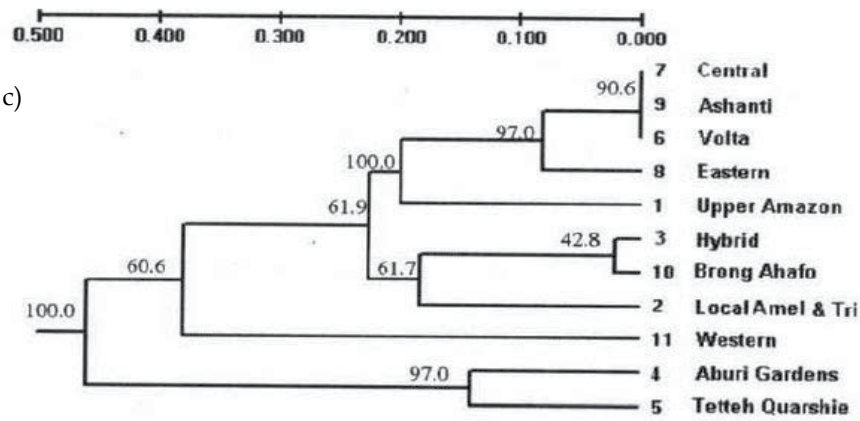


Plate 3. a. Cocoa regions of Ghana indicating sites of cacao accessions collection (dots); b. Table of genetic diversity indices in cacao form the different regions and germplasm collection; c. Relationships among cacao accessions in the different regions and field genebank accessions

The cocoa produced at the Tetteh Quarshie farm is of the Amelonado type, which originated from Brazil. Results of diversity analysis showed a clear separate clustering of accessions from Tetteh Quarshie farm and Aburi garden compared to rest of cocoa populations presently grown n Ghana (Plate 3c). These accessions were among the first introductions into the country and then spread across the country; however, the results indicated that presently these accessions have little or no influence on the current plantings in farmers'

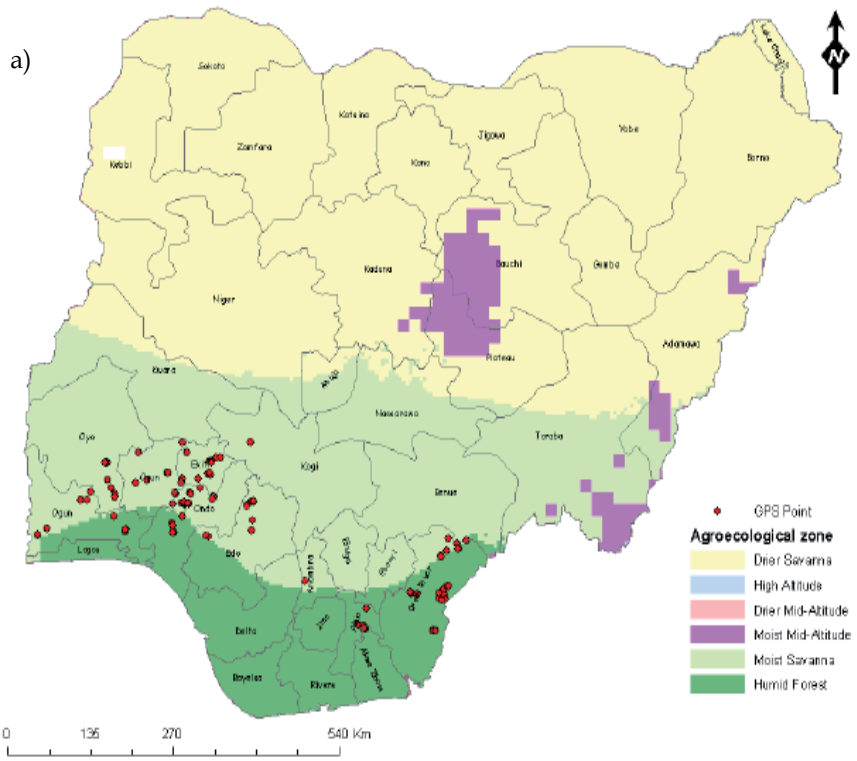
fields. Subsequent hybridization with later introductions and adoption of farmers' own or newly released improved germplasm might have been responsible for observation made. The accessions from Western region clustered separately from those of breeders' collection and populations of other regions indicating that the breeders' germplasm had less impact on planting materials in the West of the country in comparison with other regions. This proved to be in agreement with the historical records that cocoa cultivation in Ghana had spread from other adjacent regions to the West. Additionally, the seed gardens from which farmers could obtain improved planting materials developed by breeders are fewer in the region and are inaccessible due to poor road network. On the other hand, farmers from other regions mostly collected seeds from the 'Seed gardens'. This explained why the farmers' collections from these regions clustered with the Breeders' collections. Another interesting observation was that the germplasm from Central, Ashanti, Volta, and Eastern regions, which constituted the earliest cocoa-growing regions of Ghana, clustered together and separately from the Parental clones and Breeders' collection, whereas the accessions from Brong Ahafo region clustered with the Breeders' collections. This showed that most of Brong Ahafo plantings were done at the time when the "Series II hybrids" had been developed and were popular in the country. Most farmers in this region might have used those varieties as planting material. However, in the case of the other regions, a substantial number of farmers had, in addition to breeders' varieties, used materials from their own farms or other neighboring sources.

6.4 Nigeria (Aikpokpodion 2007; Aikpokpodion et al., 2009; Aikpokpodion 2010; Aikpokpodion et al., 2010)

In a recent study, 12 microsatellite markers were used to determine genetic diversity in 574 accessions representing eight groups covering parental populations in West Africa, genebank, and farmers' accessions collected from cocoa growing regions of Nigeria (Plate 4a). From this study, it was shown that appreciable genetic diversity was present in on-farm and field genebank collections. A total of 144 alleles were detected in these accessions with a mean allelic richness of 4.39 alleles/ locus. The largest genetic diversity was found in the Upper Amazon parent population (H_{nb}=0.730), followed by the 1944 Posnette's Introduction (H_{nb}=0.704), and was lowest in the Local parent population (H_{nb}=0.471). Gene diversity was appreciably high in the farmers' populations (H_{nb}=0.563– 0.624); however, the effective number of alleles was lower than that found in the genebank's Posnette population. Fixation index estimates indicated deficiency of heterozygotes in the Upper Amazon and the Local parent populations (F_{is}=0.209 and 0.160, respectively), and excess of heterozygotes in the Trinitario parent population (F_{is}=−0.341). The presence of inbreeding in the Local parent populations and substructure (Wahlund effect) in the Upper Amazon were suggested for the deficiency of heterozygotes observed. In Nigeria, restricted gene flow and spatial differentiation was evident in cacao varieties grown by farmers in Nigeria (Plate 4b). Cacao trees grown on farmers' fields in southwestern and mid-western Nigeria are mainly hybrids of the Upper Amazon and the local Amelonado varieties, while the local Amelonado variety predominates in southeastern Nigeria (Plate 4c).The non-significant genetic differentiation observed between the genebank's and farmers' populations indicated significant impact of national breeding programs on varieties grown in farmers' plantations (Plate 4c). Results also showed that a small proportion of the genetic diversity available in field gene banks at the Cocoa Research Institute of Nigeria (CRIN) had been used to develop improved varieties

in the Institute's cacao breeding programs and that 'Scavina' and 'Imperial Mixed Calabacillo' cacao varieties of the Upper Amazon Forastero have not been significantly utilized in Nigerian cacao breeding programs. Population structure analysis cacao types grown in farmers' fields showed that the Upper Amazon Forastero constitute 66 % of cacao grown, Amelonado made up 24 %, Trinitario accounted for 6 % while other types made up 4 % of cacao grown (Plate 4d). Large number of alleles were found in the farmers' populations (7.67–9.00), and compared with the number recorded for the genebank's collection, although the effective number of alleles (4.49–4.80) was lower than in Posnette's population (5.65). This situation is encouraging as it indicated that genetic diversity held in farmers' collection in commercial plantations is much greater than, and showed a major shift from, what it was since introduction in the late nineteenth century till the 1950s when the highly uniform local Amelonado cocoa was predominant on the field. We can conclude from the molecular data used in this study that there has been a significant variety replacement of the Amelonado cacao grown on fields in West Africa with Upper Amazon and Amazon × Amelonado hybrids.

In a study to determine phenotypic variation among cacao grown in Nigeria, 17 agro-morphological traits were studied in 184 accessions collected from farmers' fields (138) and field genebank collections (46). Fruit and bean traits of Upper Amazon Forasteros observed in farmers' accessions provided evidence of a shift from previously grown local 'West African Amelonado' from the Lower Amazon Forastero population. The large variation observed in this study for cacao grown by farmers indicated a high level of heterogeneity



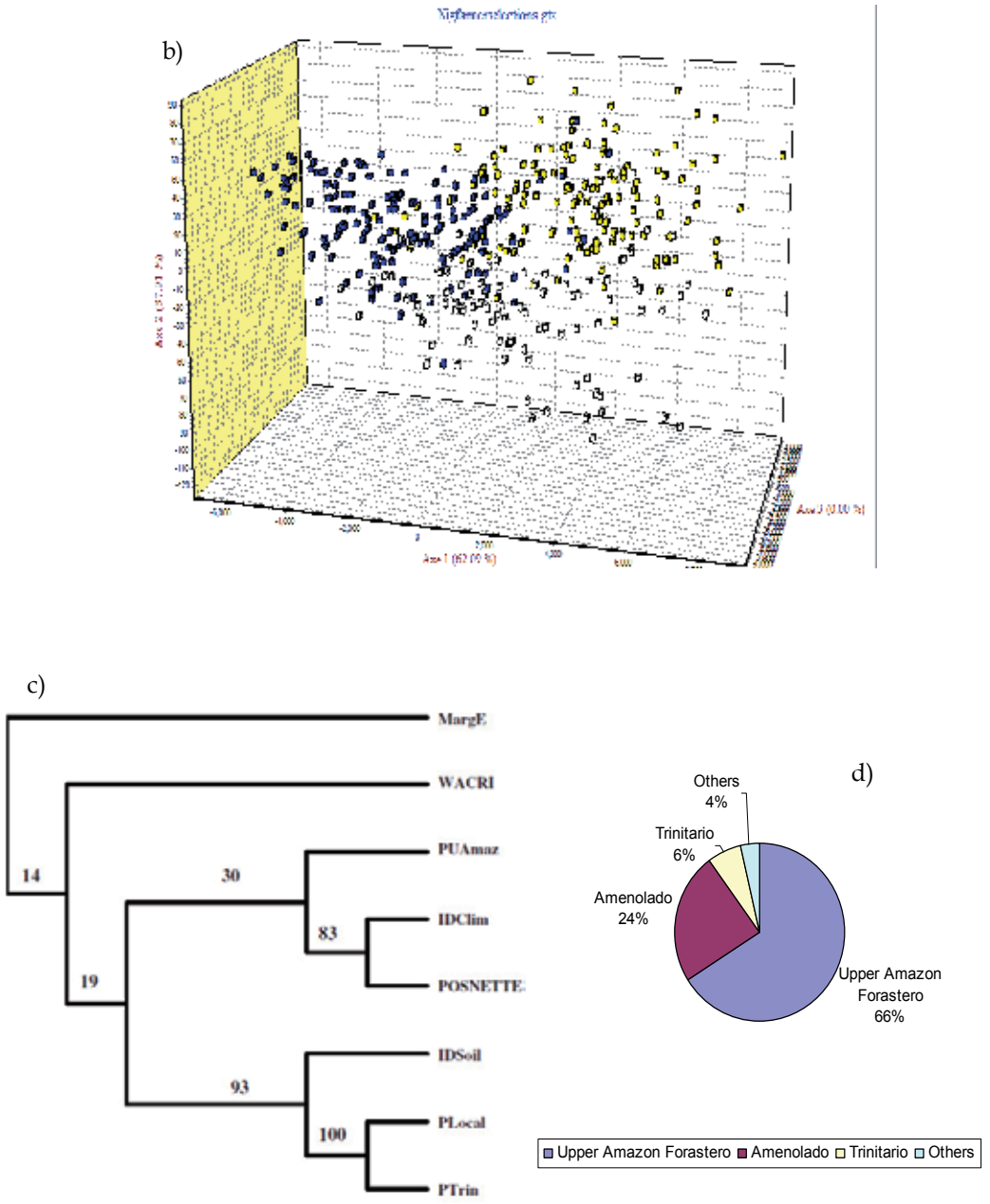


Plate 4. a. Sites of cacao germplasm collection(red dots) in cocoa producing region of Nigeria; b. Relationships showing spatial differentiation among cacao accessions collected in farmers' fields in ideal climate (yellow), ideal soil (blue) and marginal climate (white) conditions in Nigeria; c. Relationships between farmers and field genebank cacao accessions in Nigeria; d. Population structure indicating cacao types grown on farmers fields in Nigeria.

in materials maintained on-farm. The low percentage of fruit traits that are typical of 'Amelonado' and 'Trinitario' types provides some evidence of variety replacement of 'West Africa Amelonado' (WAA) cacao types in farmers' fields with Upper Amazon-derived types. This showed a radical shift from the situation preceding the 1950s, when uniform 'Amelonado' cacao types were mainly grown. This would have resulted from the use Upper Amazon-derived cacao varieties distributed to farmers through the seed gardens. Significant variation observed for bean and fruit characteristics among cacao accessions in this study also indicated the importance of on-farm collections as a valuable reservoir of genetic diversity. Some of these traits are of commercial importance and have been used as selection criteria by farmers in the choice of parent trees for raising seedlings to make new plantings and farm expansion. From this study, the complete absence or slight anthocyanin pigmentation on the ridge of the mature fruit of more than 88% of accessions indicated that most of the cacao now grown in Nigeria was apparently derived from the Amazonian Forastero origin. The preponderance of 'Cundeamour' fruit shape (76%) with slight to strong basal constriction (88%), obtuse to attenuate apex forms (95%) and intermediate to intense rugosity (86%) showed that, possibly, the Upper Amazon Forastero (UAF) 'Parinari' population, characterized by pronounced bottleneck, conspicuous apex form and the intermediate to intensely warty fruit (Bartley 2005) and to a lesser extent, the 'Nanay' population, had the most impact on cacao On the other hand, the low percentage (less than 15%) of red pigmentation in fruits, a trait associated with some 'Criollo' populations indicated that 'Criollo' and red-podded 'Trinitario' populations have, at present, only a minimal influence on field-grown cacao in Nigeria.

7. Conclusion

Cocoa beans either as export commodity or processed into cocoa products remain a significant revenue earner for the government and a veritable means for livelihood sustenance for the people of West and Central Africa. It was therefore pertinent to determine the extent of genetic diversity in this crop in order to safeguard, not only the livelihood of the people and revenue base of the government, but also the multi-billion dollars industry of processing and consuming countries. From results obtained in recent studies, the presence of appreciable genetic diversity in farmer-grown cacao indicated that there is enough resilience in the cacao types now grown in West and Central African countries to withstand any major disease outbreak, which may constitute a serious threat to the cocoa industry. This is in contrast to the situation in the 1930s and 1940s, when, due to a small range of genetic variability in materials grown in farms, the Cocoa Swollen Shoot Virus almost ruined the industry. At that time, more than 20 million trees were cut down as a control measure. The outbreak of Witches' Broom disease in the late 1980s in Brazil provides a recent example of how disease attack can affect farmers' livelihoods and the cocoa industry. The outbreak led to a drop in cocoa production from 380,000 tons per year (at that time Brazil was the world's second largest cocoa producer) to 90,000 tons in the late 1990s, when cocoa actually had to be imported. The presence of some private alleles in farmers' population also gives an opportunity to select useful recombinants on the field that have shown greater adaptation and possibly accumulated genes for resistance to prevalent diseases such as *Phytophthora* pod rot, mirid attack (*Sahlbergella singularis* and *Distantiella theobromae*), and abiotic stress (e.g., drought). Restricted gene flow and spatial differentiation as observed in Cameroon, Cote d'Ivoire, Ghana and Nigeria indicated the existence of

inefficient seed delivery systems in some cases, which are not extending the benefits of improved varieties developed by researchers to farmers. There is a need, therefore, for the establishment of a functional and efficient seed delivery system to facilitate farmers' easy access to improved planting materials.

These studies revealed that research efforts to develop improved varieties over the years have been limited to a narrow range of diversity present in germplasm collections. This, in turn, has limited the gains that would have been made from utilizing useful attributes, such as resistance genes against *Phytophthora* pod rot disease in the Scavina variety and the large bean size of the Iquitos Mixed Calabacillo variety. This information is useful for future cocoa breeding efforts and a guide for future germplasm introduction. This knowledge is also important to enable the development of appropriate breeding strategies to improve planting materials, with particular attention to the integration of available genetic diversity into future cocoa improvement programmes. However, the discontinued use of the Amelonado cocoa by farmers have significant implication for the conservation of this stock, some of which are completely homozygous at all loci used in this study. It is hereby suggested that conservation strategies should be developed to preserve the local Amelonado landraces in order to exploit their useful values in future breeding programs.

Although useful information have now been obtained on genetic diversity in the major cocoa producing (Cameroon, Cote d'Ivoire, Ghana and Nigeria) West and Central African countries, much information is still needed on the genetic diversity situation in other countries in the region and others such as Eastern and Southern African countries. For instance, information on genetic diversity in Sao Tome & Principe island and Fernando Po (Bioko) in Equatorial Guinea, the earliest places of cocoa introduction in Africa will be useful to determine the sources of many private alleles that were found in farmers accessions but absent in field genebanks. It is also very important to capture the diversity that might be present in order to exploit these for benefits of the ever increasing sophistication of the cocoa market including 'origin' and 'specialty' emphasis of the consumers. It will be very useful if funding could be made available to determine the genetic diversity of cacao types in Madagascar, Malawi, East African countries such as Tanzania and Uganda in addition to the rest West and Central African countries. This is because a thorough knowledge of genetic diversity in on-farm and field genebanks, particularly in an introduced crop species as cacao is crucial to the utilization of the genetic resources available. This will also be important for the sustainability of the global cocoa industry.

8. Acknowledgement

The author gratefully acknowledge the sponsorship provided by the United States Agency for International Development (USAID), United States Department of Agriculture (USDA), and Mars Incorporated for the project on "Assessment of Genetic Diversity in West African Cocoa Collections" conducted within the framework of the Sustainable Tree Crops Program (STCP) of the International Institute of Tropical Agriculture in association with national cocoa research centers including the Cocoa Research Institute of Nigeria (CRIN), Cocoa Research Institute of Ghana (CRIG), National Center for Agronomic Research (CNRA, Cote d'Ivoire) and Institute of Agricultural Research and Development (IRAD, Cameroon).

9. References

Aikpokpodion, P.O. (2007). Genetic diversity in Nigerian cacao, *Theobroma cacao* L. collections as revealed by phenotypic and simple sequence repeats marker. PhD Thesis submitted to the University of Ibadan, Nigeria. 147 pp

Aikpokpodion, P. O. (2010). Variation in agro-morphological characteristics of cacao, *Theobroma cacao* L. in farmers' fields in Nigeria. *New Zealand Journal of Crop and Horticultural Science* Vol. 38 (2):157-170

Aikpokpodion, P.O.; Motamayor, J.C., Adetimirin, V.O., Adu-Ampomah, Y., Ingelbrecht, I., Eskes, A.B., Schnell, R.J. & Kolesnikova-Allen, M. (2009). Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genetics and Genome* 5:699-711

Aikpokpodion, P. O.; Adetimirin, V. O., Guiltinan, M. J., Eskes, A. B., Motamayor, J-C., Schnell, R. J. & Kolesnikova-Allen, M. (2010). Population structure and molecular characterization of Nigerian field genebank collections of cacao, *Theobroma cacao* L. *Silvae Genetica* 59 (6):273-285

Alverson, W.S.; Whitlock, B. A., Nyfeller, R., Bayer, C. & Baum, D.A. (1999). Phylogeny of the core Malvales: Evidence from *ndhF* sequence data. *American Journal of Botany* 86: 1474 - 1486

Atanda, O.A. & Jacobs, V.J. (1974). Assessment of WACRI Series I and II varieties of cocoa in Nigeria. 1. Establishment and yield. *Ghana J Agric Sc* 7:215–220

Bartley, B. G. D. (2005). The genetic diversity of cacao and its utilization. CABI, Wallingford, 341 pp

Bekele, F. L. & Bekele, I. (1996). A sampling of the phenetic diversity of cacao in the international cocoa gene bank of Trinidad. *Crop Science* 36: 57-64

Besse, J. (1975). La selection generative du cacaoyer en Cote d'Ivoire: bilan et orientation des recherches en 1975. In 5th International Cocoa Research Conference, Ibadan Nigeria, 1–9 September. pp. 95–103

Botstein, D.; White, R. L., Skolnick, M. & Davis, R. W. (1980). Construction of a genetic linkage map in human usung restriction fragment length polymorphisms. *American Journal of Human Genetics* 32:314-331

Cheesman, E. E. (1944). Notes on nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture* 21:144–159

CSAE (Center for the study of African Economies) (2009): Fertile Ground: BoostingCaocoa production among Ghanaian Smallholder. Briefing Paper 01, August 2009. www.csae.ox.ac.uk

Cuatrecasas, J. (1964). Cacao and its allies: A taxonomic revision of the genus *Theobroma. Contributions to the U.S. Natural Herbarium* 35(6): 379 – 614.

Edwin, J. & Masters W. A. (2005). Genetic improvement and cocoa yields in Ghana. Expt Agric 41:491–503

Efombagn, M. I. B; Sounigo, O., Nyassé, S., Manzanares-Dauleux, M., Cilas, C., Eskes, A.B. & Kolesnikova-Allen, M. (2006). Genetic diversity in cocoa germplasm of southern Cameroon revealed by simple sequences repeat (SSRs) markers. *African Journal of Biotechnology* 5(16):1441-1449.

Efombagn, M. I. B; Motamayor, J.C., Sounigo, O., Eskes, A.B., Nyassé, S., Cilas, C., Schnell, R., Manzanares-Dauleux, M., & Kolesnikova-Allen, M. (2008). Genetic diversity

and structure of farm and genebank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. *Tree Genetics and Genome* 4:821-831.

Efombagn, M. I. B; Sounigo, O., Nyassé, S., Manzanares-Dauleux, M. & Eskes, A.B. (2009). Phenotypic variation of cacao (*Theobroma cacao* L.) on farms and in the genebank in Cameroon. *J. of Plant Breeding and Crop Science* 1(6):258-264.

Engels, J. M. M. (1992). The use of botanical descriptors for cocoa characterisation: CATIE experiences. *Proceedings of the International Workshop Conservation Characterization Utilization Cocoa Genetic Resources 21st Century*. Haines Clark& Co. (UK) Ed. Cocoa Research Unit, The University of the West Indies, Port-of-Spain, Trinidad. 69-76

Figuera, A.; Janick, J., Levy, M. & Goldsbrough, P.B. (1994): Reexamining the classification of *Theobroma cacao* L. using molecular markers. J Amer Soc Hort Sci 119, 1073-1082

Gosselin, M. (1895). Report for the year 1893-1894 0n the German Coloniesin Africa and the South Pacific. Miscellaneous Series, no. 352, 1895. Foreign Office, London, 94 pp

Guiltinan, M.J. (2007). Cacao. Biotechnology in agriculture and forestry, vol. 60. In: Pua EC, Davey MR (eds) Transgenic crops V. Springer, Berlin

ICCO (International Cocoa Organization) (2010). Quarterly Bulletin Cocoa Statistics. Vol. XXXVI. 2009/2010

Jacob, V.J.; Atanda, O.A. & Opeke, L.K. (1971). Cacao breeding in Nigeria. In: *CRIN Commemorative Publication – Progress in Tree Crops Research in Nigeria*.9-12

Knight, R. & Rogers, H. H., 1955. Incompatibility in *Theobroma cacao. Heredity* 9: 69 – 77

Laurent, V.; Risterucci, A.M. & Lanaud. C. (1994). Genetic diversity in cacao revealed by cDNA probes. *Theor Appl Genet* 88:193–198

Lerceteau, E.; Robert, T., Petiard, V. & Crouzilat, D. (1997). Evaluation of the extent of genetic variability among *Theobroma* accessions using RAPD and RFLP markers. *Theor Appl Genet* 95:10–19

Lockwood, G. & Gyamfi, M.M.O. (1979). The CRIG cocoa germplasm collection with notes on codes used in the breeding programme at Tafo and elsewhere. Cocoa Research Institute, Ghana Technical Bulletin No. 10 Tafo

Lu, H.; Redus, M.A., Coburn, J.R., Rutger, J.N., McCouch, S.R. & Tai, T.H. (2005). Population structure and breeding patterns of 145 U.S. Rice cultivars based on SSR marker analysis. *Crop Science* 45:66–76

Mitchell, S.E.; Kresovich, S., Jester, C. A., Hernandez, C. J. & Szewc-McFadden, A. K. (1997). Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Science* 37:617-624

Motamayor, J.C. & Lanaud, C. (2002). Molecular analysis of the origin and domestication of *Theobroma cacao* L. *Managing Plant Genetic Diversity* Engels, J.M.M., Ramanatha Rao, V., Brown, A.H.D. & Jackson, M.T. Eds. IPGRI. 77–87

Motamayor, J.C.; Risterucci, A.M., Lopez, P.A., Ortiz, C.F., Moreno, A. & Lanaud, C. (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89:380–386

Motamayor, J.C.; Lachenaud, P., Wallace, J., Loor, G., Kuhn, D. N., Brown, J. S. & Schnell II, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree. *PLoS One* 3: e3311.doi:10.1007/s12042-008-9011-4

N'Goran, J.A.K.; Laurent, V., Risterucci, A.M. & Lanaud, C. (2000). Comparative genetic diversity of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity* 73:589-597

N'Goran, J.A.K.; Lachenaud, Ph. & Bastide, Ph. (1992). Cocoa germplasm conservation initiatives in Cote d'Ivoire. In: Proceedings of the International Workshop on Conservation, Characterization and Utilization of Cacao Genetic Resources in the 21st Century. Cocoa Research Unit, Port of Spain, Trinidad 13–17 September 1992, pp 221–227

N'Goran, J.A.K.; Laurent, V., Risterucci, A.M. & Lanaud, C. (2000). The genetic structure of cocoa populations (*Theobroma cacao* L.) revealed by RFLP analysis. *Euphytica* 115:83–90

Opeke, L.K. (1969). Development of cacao industry in Nigeria. In: Proceedings of the 6th Conference of the Agricultural Society of Nigeria, pp. 1–5.

Opoku, S.Y.; Bhattacharjee, R. , Kolesnikova-Allen, M., Motamayor, J.C., Schnell, R., Ingelbrecht, I., Enu-Kwesi, L. & Adu-Ampomah, Y. (2007). Genetic diversity in cocoa (*Theobroma cacao*, L.) germplasm collection from Ghana. *Journal of Crop Improvement* 20(1/2):73–87.

Pokou, N. D.; N'Goran, J. A. K., Lachenaud, P., Eskes, A. B., Montamayor, J. C., Schnell, R., Kolesnikova-Allen, M., Clément, D. & Sangaré, A. (2009). Recurrent selection of cocoa populations in Côte d'Ivoire: comparative genetic diversity between the first and second cycles. *Plant Breeding* 128: 514–520. doi: 10.1111/j.1439-0523.2008.01582.x

Preuss, P. (1901). Expedition nach Central- und Sudamerika 1899/1900. Kolonial-Wirtschaftlichen Komitees, Berlin, 374 pp

Purseglove, J. W. (1974). *Tropical crops, dicotyledons*. John Wiley & Sons, New York

Rice, R. A. & Greenberg, R. (2000). Cacao cultivation and the conservation of biological diversity. *Ambio* 29(3):167-173.

Ronning, C. M. & Schnell, R. J. (1994). Allozyme diversity in a germplasm collection of *Theobroma cacao* L. *Journal of Heredity* 85: 291-295.

Russell, J. R.; Hosein, F., Johnson, E., Waugh, R. & Powell, W. (1993). Genetic differentiation of cocoa, *Theobroma cacao* L. populations revealed by RAPD analysis. *Molecular Ecology* 2: 89-97.

Saunders, J.A.; Hemeida, A.A. & Mischke, S. (2000). USDA DNA Fingerprinting programme for identification of *Theobroma cacao* accessions. *Proceedings of the International Workshop on New Technologies and Cocoa Breeding 16-17 October*. Kinabalu, Sabah, Malaysia: INGENIC, p. 108-114.

Sul, S-J.; Matthews, S., & Williams, T.L. (2009). Using tree diversity to compare phylogenetic heuristics. BMC Bioinformatics 10(Suppl 4)S3

Tahi, G. M.; N' Goran, J.A.K., Pokou, N.D., Lachenaud, P. & Eskes, A.B. (2008). Current status of planting materials in Côte d'Ivoire. In: Proceeding of CFC/ICCO/Bioversity Project Workshop on Collaborative and Farmer Participatory Approaches in Cocoa Breeding in Africa held 18 – 22 February 2008, Abidjan, Cote d' Ivoire

Tautz, D. & Rentz, M. (1984). Simple sequences are ubiquitous components of eukaryotic genomes. *Nucleic Acid Research* 12: 4127-4138

Toxopeus, H. (1964). F_3 Amazon cacao in Nigeria. *Report of West African Cacao Res. Inst. Nigeria*, 1963/64, pp 13–22

Van Delden, W. (1992). Genetic diversity and its role in survival of species. *Biodiversity and Global Change*. O.T. Solbrig, H.M. Van Emden and P.G.W.J. Van Oordt. Eds. CAB international and IUBS. 1992. Chapter 6: 41-56

Warren, J.M. (1994). Isozyme variation in a number of populations of *Theobroma cacao* L. obtained through various sampling regimes. *Euphytica* 72: 121-126

Webster, J.B. (1964). The Bible and the Plough. *Journal of Historical Society of Nigeria* II, 4:418-434

Whitkus, R.; de la Cruz, M., Mota-Bravo, L. and Gomez-Pompa, A., 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. *Theoretical and Applied Genetics* 96:621–627.

Wilde, J.; Waugh, R. & Powell, W. (1992). Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theoretical and Appliled Genetics* 83: 871–877.

Zhang, D.; Arevalo, E., Mischke, B., Zuniga, L. & Barreto, A. (2006). Genetic diversity and population structure of cocoa, *Theobroma cacao* L. in the Hullaga and Ucayali valleys of Peru. *Annals of Botany* 98:647–655

11

Genetic Diversity in Citrus

Aydin Uzun[1*] and Turgut Yesiloglu[2]
[1]*Erciyes University Department of Horticulture, Kayseri*
[2]*Cukurova University Department of Horticulture, Adana*
Turkey

1. Introduction

Citrus is the most produced fruit in the world with over the 116 million tons of production (FAO, 2009). Besides, citrus is an extremely important crop on a world-wide basis, and is grown wherever the climate is suitable. It is widely grown in most areas with suitable climates tropical, subtropical, and borderline subtropical/temperate (Kahn et al., 2001).

The genus *Citrus* L. belongs to the subtribe Citrineae, the tribe Citreae within the subfamily Aurantioideae of the Rutaceae family (Webber, 1967). The Aurantioideae is one of seven subfamilies of Rutaceae which consists of two tribes and 33 genera. Each of tribes Clauseneae and Citreae is composed of three subtribes. Clauseneae includes Micromelinae, Clauseninae and Merrillinae, and Citreae has Triphasiinae, Citrinae and Balsamocitrinae. The Citrinae is distinct from all the other subtribes in the subfamily by having pulp vesicles in the fruit. This subtribe contains three groups; primitive citrus fruit, near citrus fruit, and true citrus fruit trees. True citrus fruits have six genera: Clymenia, Eremocitrus, Microcitrus, Poncirus, Fortunella and Citrus (Swingle & Reece, 1967).

Most of genus including Citrus belongs to subfamily Aurantioideae originated from Monsoon regions and expand from West Pakistan to China, India islands, Northwest Australia, New Guinea. In this subfamily, four of 33 genus (Afraegle, Aeglopsis, Balsamocitrus and Citropsis) native to tropical Africa an one genus (Clausena) native to Monsoon and tropical Africa. Besides, Microcitrus and Eremocitrus originated from Australia (Ulubelde, 1985).

Spreading citrus to other parts of the world goes back to B.C. The first species reached to Europe was citron. This species brought to Iran by Persian and then spreaded to Europe. Sour orange and lemon were brought to Europe by the Romans through trades in first and second century A.D. A mosaic tile floor found in a Roman villa at Carthage, probably of the second century A.D., shows recognizable branches of citron and fruit-bearing lemon trees. Although there is still no consensus on the definitive homeland of lemon, this species have been described and it's name pronounced as 'li-mung' in resources books writtened by Fan Ch'eng-Ta ve Chou K'ü-Fei in 1175 and 1178 A.D. By 1150 A.D. the Arabs had brought citron, sour orange, lemon and pummelo into North Africa and Spain.

* Corresponding Author

In the written sources, sweet orange did not seem to had been widely cultivated until toward the middle of the fifteenth century. Portuguese succeeded in rounding Southern Africa soon brought better sweet oranges from India or Far East to Europe. By the beginning of the sixteenth century, there was abundant evidence showing that it had become well established and becomecommercially important in southern Europe. The mandarin, native to China was brought to Europe in 1805. Firstly it came to England and then spread to Malta, Scily and Italy. The pummelo or shaddock in its journey to Europe apparently followed about the same path as the sweet and sour oranges. It is notified that the seed of shaddock was first brought to Barbados by Captain Shaddock. The grapefruit, which probably originated as a mutation or sport from the shaddock, was first described under the name 'forbidden fruit' by Griffith Hughes in 1750 from Barbados. From Europe citrus spread to the New World firstly by Columbus in 1493 and then continued (Webber, 1967; Scora, 1975).

Citrus taxonomy and phylogeny are very complicated, controversial and confusing, mainly due to sexual compatibility between Citrus and related genera, the high frequency of bud mutations and the long history of cultivation and wide dispersion (Nicolosi et al., 2000). In addition, the level of difference in relation to species status in Citrus is uncertain. Citrus taxonomy was based on mainly morphological and geographical data in the past and many classification systems have been formulated. Two of these systems suggested by Swingle & Reece (1967) and Tanaka (1977) have been the most widely accepted. The number of recognized species is the major difference between two systems. Swingle recognized 16 species in the genus Citrus, whereas Tanaka (1977) recognized 162 species. Scora (1975) and Barrett & Rhodes (1976) suggested that there are only three 'basic' true species of Citrus within the subgenus Citrus as follow: citron (*C. medica* L.), mandarin (*C. reticulata* Blanco), and pummelo (*C. maxima* L. Osbeck). Later, Scora (1988) added *C. halimi* as another true species. Other cultivated species within Citrus were derived from hybridization between these true species or closely related genera followed, mainly, by natural mutations. Recently, this thesis has gained support from various biochemical and molecular studies (Federici et al., 1998; Nicolosi et al., 2000; Barkley et al., 2006; Uzun et al., 2009a). Elucidating relationships, taxonomy, and diversity is important for developing breeding strategies, conserving biodiversity, and improving breeding efficiency. Also understanding genetic variability in citrus is critical for characterizing germplasm, controlling genetic erosion and the registration of new cultivars (Herrero et al., 1996; Barkley et al., 2006).

Use of molecular markers has more advantages than that of morphologically based phenotypic characterization, because molecular markers are generally unaffected by external impact. It is possible to compare accessions of a collection at any time of year using molecular markers, while phenotypic characteristics can be influenced by environmental or cultural affects (The Citrus and Date Crop Germplasm Committee, USA,CDCGC, 2004). Regarding to germplasm management molecular characterization has a number of applications such as relationships between accessions, characterizing newly acquired germplasm, monitoring shifts in population genetic structure in heterogeneous germplasm, exploiting associations among traits of interest and genetic markers and genetic enhancement (Bretting and Widrlechner, 1995, as cited in The Citrus and Date Crop Germplasm Committee, USA, CDCGC, 2004).

In the present study, we summarized genetic variations and relatioships among citrus species and cultivars mostly cultivated. While preparing this, many studies were overrewieved to better explain citrus diversity. Better understanding genetic relation in citrus offer more opportunities to conservation and evaluation of genetic resources. It is also important for citrus researcher and breeders to arrange their future studies.

2. A general view of genetic relationships among cultivated citrus species

It is suggested that the cultivated citrus derived from the three true species, citron, pummelo, and mandarin (Barrett & Rhodes, 1976). These three species reproduce sexually and if different cultivars within the species are intermated, the progeny are similar to their parents. The other important types (orange, grapefruit, lemon, and lime) are believed to have originated from one or more generations of hybridization between these ancestral genera. Most of the cultivars of orange, grapefruit, and lemon are believed to have originated from nucellar seedlings or budsports. Currently citrus fruits have high level of morphologic variations and various fruit characteristic because of inter and intraspecific interaction (Fig 1). Consequently, the amount of genetic diversity within these groups is relatively low, in spite of there being many named varieties. Conversely, mandarins, pummelos, and citrons have higher levels of genetic diversity since many of the cultivars have arisen through sexual hybridization (The Citrus and Date Crop Germplasm Committee, USA, CDCGC, 2004).

Fig. 1. Citrus fruits have distinct fruit characteristics.

Relationships among citrus species was invastigated previously. In a recent study, citrus accessions were divided into two large groups. The first group included citron, lemon, lime, rough lemon, with a similarity value of 0.60 from other Citrus species (Uzun et al., 2009a). The similar results were reported in previous studies (Federici et al., 1998; Nicolosi et al., 2000; Barkley et al., 2006). Citron was one of the progenitor of lemons (Nicolosi et al., 2000; Gulsen & Roose, 2001). It was reported that lemons were thought to be natural hybrids of a citron and a lime (Scora, 1975; Barrett & Rhodes, 1976) or a hybrid of citron and sour orange (Nicolosi et al., 2000; Gulsen & Roose, 2001). Limes are apparently hybrids of citrons and papedas (Scora, 1975) or a tri-hybrid cross of citron, pummelo, and Microcitrus, and had the highest observed heterozygosity of all the taxonomic groups (Barrett & Rhodes, 1976; Nicolosi et al., 2000; Barkley et al., 2006). Rough lemon was reported as hybrid of mandarin and citron (Scora, 1975). *C. volkameriana* was clustered with rough lemon as in the RAPD and SCAR study, and reported as a hybrid between citron and sour orange (Nicolosi et al., 2000).This finding supports the citron as major progenitor of some commercial citrus cultivars such as all 'true' lemons, limes and rough lemon.

Sweet orange, mandarin, sour orange, pummelo and grapefruit nested in same large group in previous study (Uzun et al., 2009a). This group separated two subgroup at similarity level of 0.64. The first subgroup included sweet oranges, mandarins and sweet oranges were separated from mandarins at 0.78. Parental sweet orange tree was a hybrid of pummelo and mandarin (Scora, 1975; Barrett & Rhodes, 1976), which was later supported by Nicolosi et al. (2000). Barkley et al. (2006) suggested that sweet orange has a majority of its genetic makeup from mandarin and only a small proportion from pummelo. The second subcluster included pummelo, grapefruit and sour orange. In this subcluster, pummelos and grapefruits were separated from sour oranges with a similarity value of 0.68. Pummelos and grapefruits showed a similarity level of 0.83. Grapefruit was reported as a hybrid of pummelo and sweet orange (Barrett & Rhodes, 1976; Nicolosi et al., 2000), and all grapefruit cultivars originated from single parent through mutations (Corazza-Nunes et al., 2002). Pummelo was indicated as one of the 'true basic species' in cultivated Citrus (Barrett & Rhodes, 1976). On the other hand, sour orange was reported as a hybrid of mandarin and pummelo in previous studies (Barrett and Rhodes, 1976; Barkley et al., 2006; Abkenar et al., 2007).

'Rangpur' lime (*C. limonia*) and bergamot (*C. bergamia*) were nested in the same branch and closely related to sour orange (Uzun et al., 2009a). Sour orange was reported as a hybrid of mandarin and pummelo in previous studies (Barrett & Rhodes, 1976; Barkley et al., 2006; Abkenar et al., 2007). Low level of genetic variation was found among sour oranges (Uzun, 2009). On the other hand, there was no polymorphism in sour oranges based on leaf isozymes (Torres et al., 1978) and SSR markers (Luro et al., 2000). Torres et al. (1978) reported that 'Rangpur' lime is quite different morphologically and genotypically from limes and was listed under *C. reticulata*. Nicolosi et al. (2000) indicated that 'Rangpur' was a hybrid of citron and mandarin and clustered with the citrons. According to Barkley et al. (2006), Webber (1943) believed that rangpurs were more similar to mandarins therefore, the origin and parentage of the rangpurs has been unclear, but they have generally been classified with mandarins in most previous studies. Hodgson (1967) suggested origin of bergamot was obscure, but probably related to sour orange. This accession was identified as a hybrid of citron and sour orange (Nicolosi et al., 2000) and clustered with sour orange (Federici et al., 1998).

3. Genetic diversity in orange

In the cultivated citrus, sweet orange (*C. sinensis* L. Osbeck) originated as a natural hybrid between mandarin and pummelo (Barrett & Rhodes, 1976), showed low level of genetic diversity according to lots of previous studies (Luro et al., 1995; Novelli et al., 2000; Novelli et al., 2006; Uzun, 2009). It is notified that most of sweet oranges obtained by mutation from one ancestor tree. So despite of differences in morphological characters, genetic variation of sweet orange was low (Fang & Roose, 1997).

In recent study carried out using large amount of orange showed that there was high level of genetic similarity in oranges (Uzun, 2009). Similarity level of 250 orange accessions varied between 0.86 and 1.00. 'Chironja' was the most distant cultivar with 0.86 of similarity because of it derived from zygotic origin. This cultivar considered as a hybrid of orange and grapefruit and originated to Puerto Rico. Also it has large fruit and light yellow rind color (Fig. 1) (Hodgson 1967). Ambersweet had zygotic origin also separated clearly from other oranges. It is notified that this cultivar was a hybrid between a genotype obtained 'Orlando' tangelo X Clementine mandarin and unknown oranges (Jackson & Futch, 2003). Genetic similarity of all of other oranges was over 0.98 and some of them were identical. In this group there were many common orange cultivars and clones such as, many 'Washington Navel', 'Valencia' , 'Moro', 'Shamouti', 'Pineapple', 'Parson Brown', 'Salustiana', 'Sanguinello', 'Tarocco' and 'Yafa' clones introduced from other countries or selected in Turkey. On the other hand, lots of Turkish orange cultivars and clones for example, 'Agma', 'Alanya Dilimli', 'Dortyol Yerli', 'Kozan Yerli', 'Sultanhisar Yerli', also were existed in that group. Same results also reported in other studies. Barrett & Rhodes (1976) notified variations in orange, lemon, grapefruit and lime based on mutations occurred on one ancestor tree. Roose (1988), reported it was difficult to distinguish cultivars originated mutations using isozyme markers. Low level of polymorphism in orange also found with ISSR (Fang & Roose, 1997), SSR (Luro et al., 2000; Novelli et al., 2006), SRAP (Uzun et al., 2009a). On the other hand, no variation found in studied oranges in some researchs (Orford et al., 1995; Qing-Qin et al., 2007).

Orange cultivars are classified into four groups: common, low acidity, pigmented and navel oranges (Hodgson, 1967, as cited in Novelli et al., 2006). It is indicated despite the existence of substantial diversity among cultivated genotypes in respect of morphological, physiological and agronomic traits, very little DNA variation has been detected using DNA-markers (Novelli et al., 2006). Same researcher found low level of genetic polymorphism among 41 orange cultivars. Similarity level of oranges varied between 0.96 and 1.00 and most of them were identical. They notified that sweet oranges have a narrow genetic basis and that most morphological characters originated through mutations, and clonal propagation of sweet oranges is the case for the majority of citrus species ((Herrero et al., 1996; as cited in Bretó et al., 2001). Fang & Roose (1997) used ISSR markers to differentiate 41 samples of orange belongs to three groups, Valencia, blood and navel based on fruit traits. All of these cultivars found almost the same ISSR fingerprints. This notified as majority of sweet orange cultivars derived from a single ancestor by mutation. However, some cultivar distinguished from others. Among the seven Valencia orange cultivars, only 'Midknight' differed from the other Valencias. Among the blood oranges, four of the five cultivars showed unique fingerprints for 1-3 fragments which distinguished them from all other sweet orange cultivars. Also 9 of 21 navel oranges had unique fingerprint patterns. Two Parent 'Washington' and 'Navel' samples obtained from different locations differed. It is

explained as only case in which replicate samples of the same cultivar from different locations had different ISSR fingerprint patterns. It is indicated this result suggests that mutation occurred in at least one of them although horticultural traits are not known to between them (Fang &Roose, 1997). In other study, it was found identical microsatellite profiles at 9 out of 10 SSR loci among analyzed orange cultivars and clones (Hvarleva et al., 2008). For one locus 'Frost Valencia' and 'Shekeriko' acidless local Cyprus orange were discriminated from 'Shamouti', 'Jaffa', 'Valencia long' and 'Aematousiki' oranges. Researchers also investigated high level of similarity of genotypes, cultivars and clones is in contrast with the observed phenotypic variability among them (Fig. 2), indicating that the local cultivars were possibly derived through mutations which are not detectable by the used SSRs or they are clones of the same original cultivars. This is in accordance with the view that most of the orange cultivars were derived through mutations which affect mostly fruit traits.

Fig. 2. Fruit image of 'Chironja' (top) and 'Moro' (bottom) orange cultivars showed phenotypic variation among oranges(from Uzun, 2009).

4. Genetic diversity in mandarin

Mandarin was considered as one of the true citrus species (Barrett & Rhodes, 1976) and this idea supported by following researches (Nicolosi et al., 2000; Barkley et al., 2006; Uzun et al., 2009a). Mandarin group has great amount of cultivars and some of them originated from hybridization and the others derived from mutation. So, in the mandarins obtained from hybrid origin there was clear genetic variation. On the other hand, low level of diversity observed in the cultivars occurred by mutation such as Satsuma and Clementine groups (Breto et al., 2001; Barkley et al., 2006; Uzun et al., 2011a).

In the recent study carried out with SRAP markers, mandarins separated two large groups at similarity level of 0.79 (Uzun, 2009). Satsuma and Clementine mandarins was nested in different groups. 'King' mandarin (*Citrus nobilis* Loureiro) late maturing and has large fruit cultivar found closely to Satsuma group. It was also reported by Coletto Filho et al. (1998) that genetic similarity among mandarins was over 0.77 and Satsuma and 'King' nested in the same gruop. Nicolosi et al. (2000) found that Satsuma and King closely related according to their RAPD and SCAR data. 'Kara' reported as a hybrid from 'Owari' Satsuma X 'King' (Hodgson, 1967), grouped closely to Satsuma mandarins according to SRAP data (Uzun, 2009). Low level of genetic variation found in most Satsumas and similarity level of nearly 50 accessions was over 0.98. They separated several group including 2-15 accessions and there was no genetic differences into the each groups. In the another study, there was no variation among the 16 Satsuma mandarin and it was notified these genotypes obtained from mutations. In the same way, Fang and Roose (1997), found no differences in five Satsuma cultivars and Barkley et al. (2006), reported nearly all cultivars in the Satsuma group originated from mutations and they had same genetic construction. Hodgson (1967), classified Satsuma as seperate group in the mandarins and notified that Satsuma naming as *Citrus unshiu* Markovitch. Also same researcher reported Satsuma was a nonstable group and lots of cultivars and genotypes had been occurred by variations in this group.

In citrus many economically important genotypes are obtained from hybridization. In the mandarin group there are lots of hybrid accessions derived from mandarin x mandarin, mandarin x pummelo (as tangelo), mandarin x orange (as tangor) or mandarin x tangelo. According to SRAP data, all tangelos and tangors closely related to mandarin instead of orange and pummelo (Uzun, 2009). On the other hand, 'Ellendale', 'Ortanique', 'Mandora', 'Lake' tangelo, 'Orlando' tangelo (Fig. 3.), 'Thornton', tangelo, 'Seminello' tangelo, 'Sampson' tangelo and 'Robinson' and 'Nova' mandarins (both cultivars are result of Clementine X Orlando) grouped closely. 'Ortanique' and 'Mandora' were nearly identical and these cultivars were showed as synonym (Cottin, 2002). On the other hand, 'Ortanique' was reported as a natural hybrid between orange and mandarin (Hodgson, 1967). Also 'Robinson' and 'Nova' shared same parents had the high level of similarity. 'Dancy' mandarin separated clearly from other mandarins and nested alone in the dendrogram obtained by SRAP data (Uzun, 2009). According to Hodgson (1967), this cultivar classified as a species by Tanaka (1954) and originated from India. 'Tankan', 'Ponkan', 'Minneola' tangelo (Fig. 3.), 'Batangas', 'Swatow' nucellar and 'Fuzhu' grouped closely. 'Ponkan' and 'Batangas' were notified as a synonym (Cottin, 2002). Besides, 'Tankan' was reported as hybrid of mandarin and orange (Coletta Filho et al., 1998). Other cultivars have hybrid origin that 'Fremont', 'Kinnow' and 'Murcott' were nested together. 'Fremont' obtained from crossing of 'Clementine' X 'Ponkan' and 'Murcot' known as hybrid of mandarin and orange (Hodgson, 1967). 'Sunburst', 'Fairchild', 'Encore'

and 'Bower'clustered in same small group (Uzun, 2009). Also 'Sunburst' and 'Fairchild'found as closely related and clustered same group in other study (Barkley et al., 2006). 'Fairchild' was a hybrid of 'Clementine' X 'Orlando' and 'Encore' was a hybrid of 'King' X 'Willowleaf' (Hodgson, 1967). 'Fortune' as a hybrid between 'Clementine' and 'Dancy' (Hodgson, 1967), was found more related to 'Clementine' instead of 'Dancy' (Barkley et al., 2006; Uzun, 2009). 'Lee' as 'Clementine' hybrid clustered closely to 'Clementine' than other 'Clementine' hybrids such as 'Nova' and 'Robinson' (Uzun, 2009).

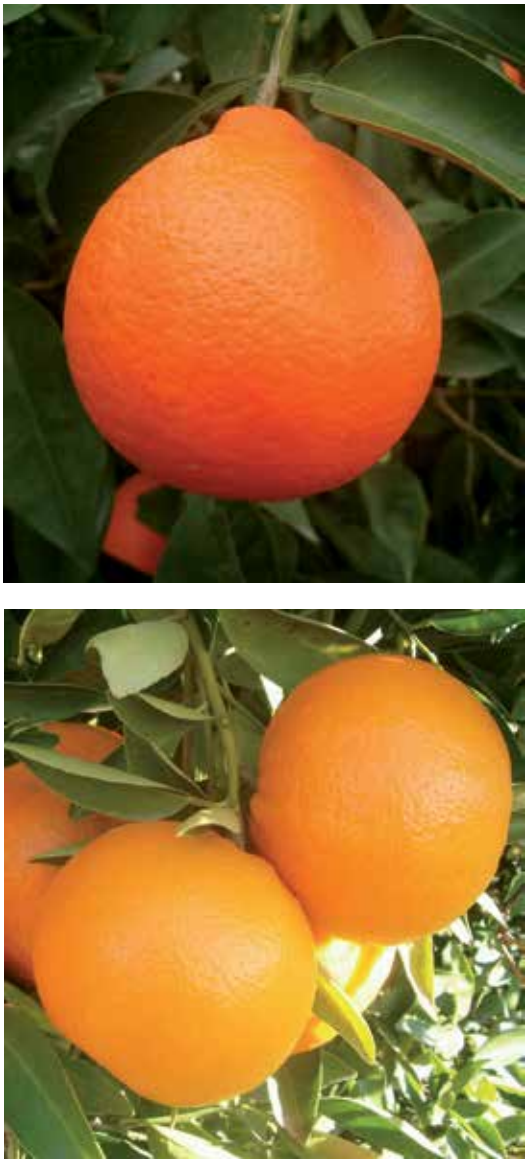


Fig. 3. 'Minneola' (top) and 'Orlando' (bottom) tangelos have same parentage (hybrid of Duncan grapefruit and Dancy mandarin) and contributed to the parentage of the such mandarin.

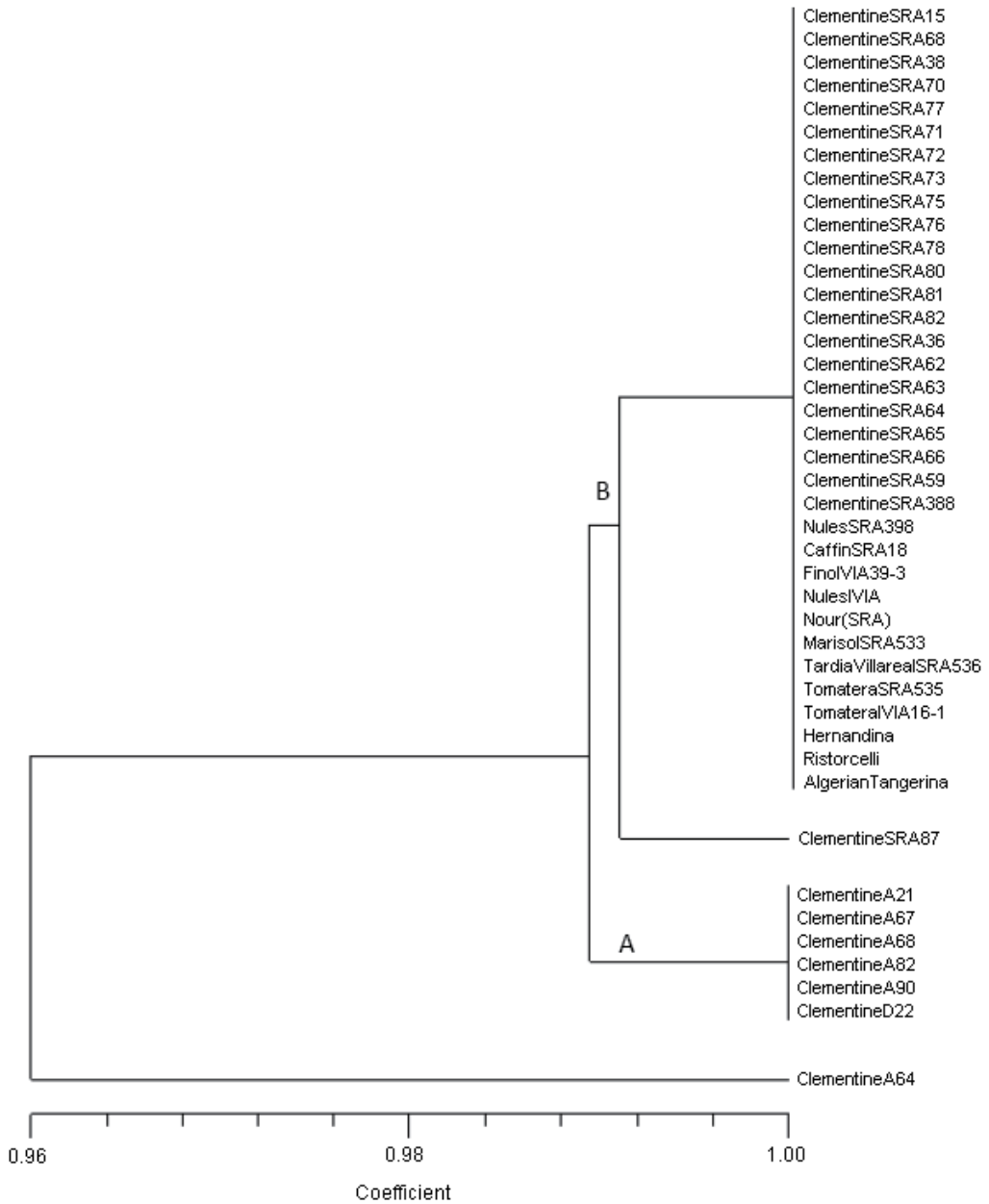


Fig. 4. Dendrogram showed relationships among 42 Clementine accessions based on SRAP markers (from Uzun et al., 2011a).

Clementine (*Citrus clementina* Hort. ex Tan.) was classified as a *Citrus* species (Tanaka, 1977). Currently, this species is one of the most important mandarin hybrid especially in the Mediterranean countries due to its good fruit quality and flavour, high yield, easy peeling. A lot of Clementine clones with high quality and different maturity time were obtained from clonal selection and most of them registered as new cultivars. Bud mutations often

arise in Clementine, as it is the case also for orange and Satsuma mandarin, which are generally detected by the growers in branches of trees showing altered horticultural traits, such as maturity and flowering time, or fruit characteristics (Breto et al., 2001). Contrasting with this diversity for agronomic traits, very low genetic variability has been found in cultivated citrus (Fang & Roose, 1997; Federici et al., 1998; Coletta-Filho et al., 1998; Luro et al., 2000; Breto et al., 2001; Corazza-Nunes et al., 2002, Uzun et al., 2011a). According to recent study carried out with 42 Clementine accession using SRAP markers genetic similarity of Clementine mandarins over the 0.96 (Uzun et al., 2011a). As a Turkish selection 'Clementine A 64' seperated from other accessions. The rest of 41 accessions divided two groups at 0.99 (A and B). Group A consisted of six Clementine accessions' selected in Turkey and all of them were identical (Fig. 4). Group B included 35 foreign Clementine accessions and 34 of them were identical. Only 'Clementine SRA 87' was distinguished from others in this group.

In a study (Uzun et al., 2011a) genetic diversity of Clementine accessions was very low and most of them were indistinguished. Seven accessions originated in Turkey were separated from foreign accessions. Six of seven Turkish selection were identical. Turkish and other accessions grouped based on their geographic origin whereas accessions originated from other countries such as SRA series from France except 'Clementine SRA 87', 'Fino' from Spain and 'Algerian Tangerina' from Algeria, were indistinguished. It can be explained that Turkish accessions have low level of polymorphism due to long period of cultivation in Turkey without influence of foreign cultivars. It is reported previously that there was limited number of polymorphism in Clementine based on RAPD data and it is suggested Clementines were genetically similar (Russo et al., 2000). Breto et al. (2001), found low level of polymorphism, distinguished only two accessions and notified Clementines are vegetatively propagated and the new cultivars are obtained after careful selection of spontaneous somatic mutations. Luro et al. (2000) speculated that the microsatellites could not distinguish mutation-derived species such as sweet and sour orange, whereas polymorphism was detected among lemon and citron cultivars.

5. Genetic diversity in lemon and relatives (citron, rough lemon, *C. volkameriana*)

Citron that major progenitor of some commercial Citrus cultivars such as all true lemons and rough lemon was reported as one of the "basic" true Citrus species and (Barrett & Rhodes 1976; Gulsen & Roose 2001). Lemon (*C. limon* (L.) Burm. f.) was accepted as a species by two important taxonomic systems (Swingle & Reece 1967; Tanaka, 1977), but it has been reported as a hybrid by other studies (Barrett & Rhodes 1976; Torres et al., 1978; Herrero et al., 1996). Besides, lemon was notified as a hybrid of citron and sour orange (*C. aurantium* L.) in recent studies (Nicolosi et al. 2000; Gulsen & Roose 2001). Most lemons have highly similar morphological and biochemical characters, and some are reported to have originated by mutation from a single parental lemon tree. Rough lemon (*Citrus jambhiri* Lush) was reported to be closely related with the citrons in previous studies (Federici et al., 1998; Nicolosi et al., 2000; Barkley et al., 2006; Pang et al., 2007) and was also reported as a hybrid of mandarin and citron (Scora 1975; Nicolosi et al., 2000; Barkley et al., 2006). *Citrus volkameriana* was reported as a hybrid between citron and sour orange (Nicolosi et al., 2000).

In a recent study genetic diversity in citron, lemon, rough lemon and *C. volkameriana* group was carried out using SRAP and SSR markers. They evalutaed 56 accessions were evaluated (Uzun et al., 2011b). Similarity level of citrons to other accessions were ~0.70. Four citron accessions that 'Buddhas Hand' (fingered citron, Fig. 5) , 'Etrog' and two Turkish selections were distinguished clearly (Fig. 6). Similarly, Gulsen and Roose (2001), reported that similarity level of citron and lemon-rough lemon group was 0.65 based on their ISSR data. On the other hand, according to Uzun et al. (2011b), genetic similarity among lemons and rough lemon-*C. volkameriana* group was 0.80. Rough lemons and *C. volkameriana* were closely related. At the same way, *C. volkameriana* was clustered with rough lemon as in the RAPD (Luro et al., 1992) and SCAR (Nicolosi et al., 2000) based studies.

Fig. 5. Fruit shape of 'Buddhas Hand' (fingered citron)

It is reported there was low level of polymorphism among most of lemons derived via clonal selection whereas higher genetic diversity was found in lemons which had hybrid origin (Uzun et al., 2011b). Genetic similarity of 45 lemons included both mutation and hybrid origin was notified between 0.80-1.00 (Fig. 6). The most distinct cultivars were 'Ponderosa' and 'Song Panache'. 'Ponderosa' was suggested as citron like fruits and notified as monoembryonic with large fruits and hybrid of citron and lemon (Hodgson 1967; Kahn et al., 2001). At the same way, 'Ponderosa' was the most distant cultivar from the other lemons based on leaf isozyme data (Torres et al., 1978). On the other hand, 'Song Panache' lemon have similar-shaped fruit as 'Ponderosa'. Another hybrid origin lemon 'Meyer' also separated from other lemons based on SRAP and SSR data (Uzun et al., 2011b). This cultivar was classified as lemon-resembling fruit by Hodgson (1967). Some Turkish lemons had hybrid origin such as 'Tuzcu 09 Aklimon' and its selections 'Tuzcu 896' and 'Tuzcu 897' was

distinguished. These lemons were also found as distinct from others in previous studies (Aka-Kacar et al., 2005; Uzun et al., 2009b).

Interdonato' lemons was clearly separated from others and genetic similarity among 'Interdonato' lemons was very high caused by their mutation origin (Uzun et al., 2011b). This cultivar was reported as a hybrid between lemon and citron (Hodgson 1967; Gulsen and Roose 2000). "Interdonato" was found as apart from other lemons in previous studies carried out with different marker systems (Deng et al. 1995; Gulsen and Roose 2001).

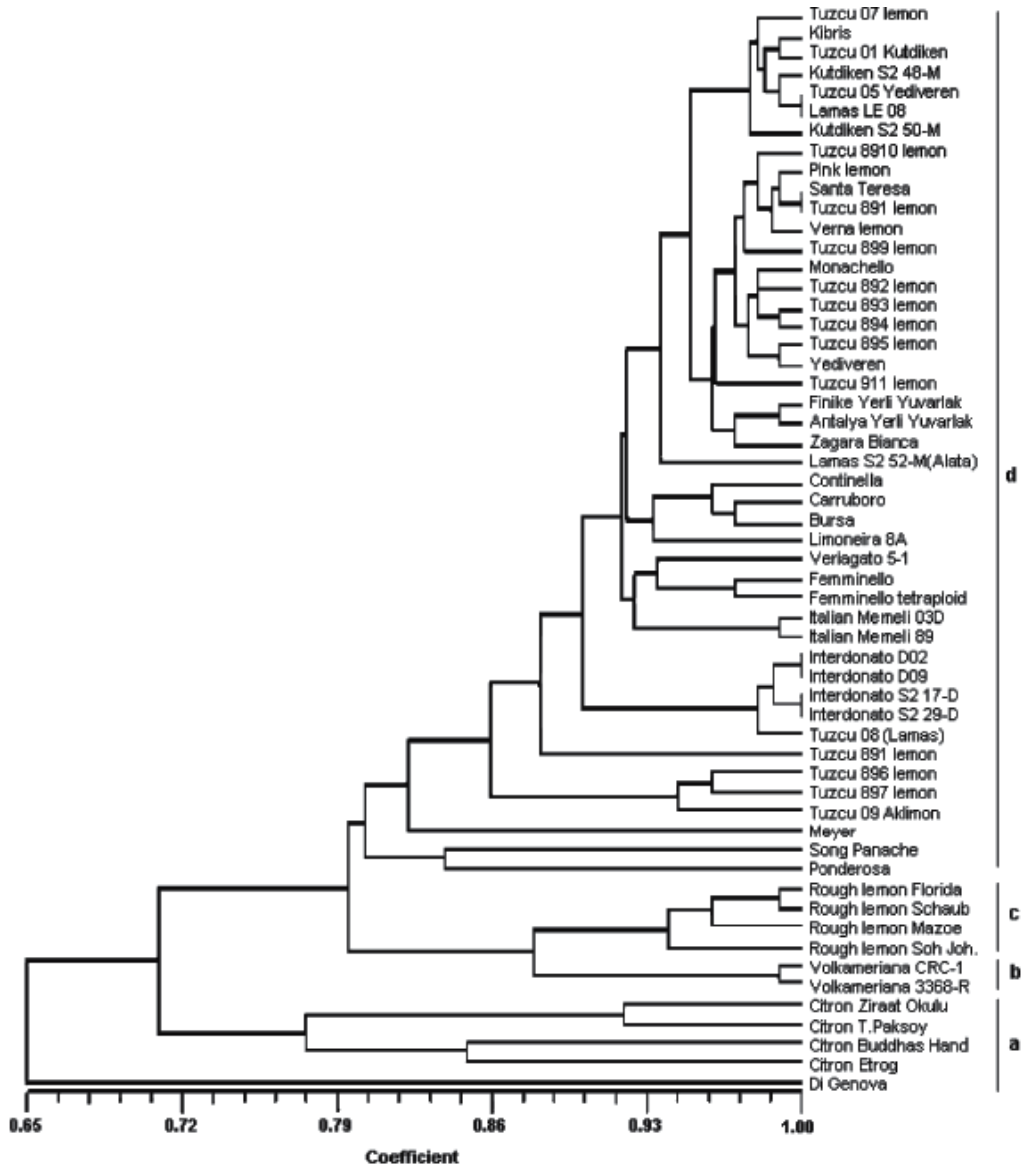


Fig. 6. Dendrogram showed relationships among citron, rough lemon, *C. volkameriana* and lemons (a; citrons, b: C volkameriana, c: rough lemons, d: lemons; from Uzun et al., 2011b)

It is notified that although they were distinguished genetic similarity of most of lemon was very high (~0.92-1.00) (Uzun et al., 2011b). Similar knowledge also was allowed by Gulsen & Roose (2001) and they indicated most lemons originated from mutation. This group included some lemons from different countries such as 'Femminello', Zagara Bianca', Carruboro', 'Continella', 'Limoneira 8A', 'Santa Teresa', 'Verna', 'Monachello' and several from Turkey that 'Kutdiken', 'Yediveren', Italian Memeli', 'Lamas', 'Kibris'. There was high level of similarities between Turkish and other lemon accessions and there was no clear clustering according to their origin. Although most lemons studied were distinguished, diversity level among these lemons was low (Uzun et al., 2011b).

As an important citrus rootstock sour orange was investigated partly in terms of genetic diversity in some studies. Recently, Lombardo et al. (2011) studied genetic variability of eight sour oranges using ISSR markers. They found a very low level of genetic variability among the cultivars; 'Canaliculata' formed a separated cluster with orange, suggesting a probable hybrid origin derived from crossing between sour and sweet orange. On the other hand, six sour oranges shared the same ISSR fingerprinting pattern not allowing to genetically distinguish anyone of them, while morphologically, they are notably different for the peculiar traits of the fruit and/or leaves. 'Crispifolia' was closely related to the previous six cultivars, suggesting a common origin of the group. Authors notified that this very low or absent genetic variability could be explained on the basis that these particular characteristics depend from mutations that do not varythe DNA length between the simple sequence repeats.

High level of genetic similarity in sour orange also reported by other researchers. Barrett and Rhodes suggested variations in orange, lemon, grapefruit, sour orange and lime based on one ancestor tree. Torres at al. (1978) found any differences in 15 sour oranges accordingto their leaf isozyme data. Luro et al. (2000) found no polymorphism among 10 accession of sour orange with SSR markers. Recently, Uzun (2009) also found very narrow genetic diversity in sour oranges. Besides sour oranges had hybrid origin such as 'Australian' or 'Smooth Seville' were distinguished from other accessions.

6. Genetic diversity in grapefruit and pummelo

The grapefruit (*C. paradisi* Macf.) was notified as a natural hybrid between pummelo (*Citrus maxima* (Burm.) Merr.) and sweet orange (*C. sinensis* L. Osb). It originates from Barbados in the Caribbean islands and was first named as *Citrus paradisi* Macf. by James Macfedyan in 1837 (Scora et al., 1982; Scora, 1988). Grapefruits are highly polyembryonic, therefore they are of nucellar and mutation origin. Genetic variation among common grapefruit cultivars was reported to be very low due to their mutation origin (Fang & Roose 1997; Corazza-Nunes et al., 2002).

The pummelo is native to tropical and subtropical regions in Asia and has been cultivated China for over 2000 years (Corazza-Nunes et al., 2002; Yong et al., 2006). Pummelo was reported as one of the three true citrus species by Barrett and Rhodes (1976) and most of subsequent studies were in agreement with this statement (Federici et al. 1998; Nicolosi et al., 2000; Barkley et al., 2006; Uzun et al., 2009a). Pummelo has played an important role as a parent of many citrus fruits, such as lemons, oranges and grapefruits.

In recent studies genetic variation in pummelo and grapefruit were investigated. Uzun et al., (2010), was determined genetic diversity among 35 accessions of grapefruits and pummelos

using ISSR markers. In that study, grapefruits and pummelos were seperated clearly and similarity value of this two species was 0.79. Besides, all pummelos were distinguished and it might be of thier zygotic origin. Same results obtained from SRAP data (Uzun et al., 2011c). At the same way, Yong et al., (2006) also seperated pummelos using SSR markers. It was concluded that pummelos were monoembryonic and there was a high level of polymorphism in the pummelos (Yong et al., 2006).

In the grapefruit group some accessions such as 'Wheeny', '*Citrus hassaku*', 'Cocktail' and 'Oroblanco' were clearly separated from pummelos and grapefruits and nested between this two species according to various marker systems (Uzun et al., 2010; Uzun et al., 2011c). 'Wheeny' originated as a chance seedling in Australia and under heat-deficient climatic conditions in Australia and New Zealand it is a summer-maturing variety. While the fruit is grapefruit-like in most respects, the monoembryonic nature of seeds and some ofthe other characters suggest that it is probably a pummelo hybrid (Hodgson 1967). *C. hassaku* was reported as an independent species (*Citrus hassaku* Hort. Ex Tanaka) and originated as a chance seedling in Japan and its characteristics strongly suggest the pummelo-mandarin parentage with pummelo predominant (Hodgson 1967). On the other hand, *C. hassaku* was notified as a pummelo hybrid (Kahn et al., 2001). 'Cocktail' was indicated as a hybrid between 'Frua' mandarin and low acidity pummelo (Kahn et al.,. 2001). Another accession 'Oroblanco' was reported as a hybrid between acidless pummelo and grapefruit (Kahn et al., 2001).

High level of similarity was found in grapefruit cultivars in various studies. Low level of polymorphism was detected in grapefruits and some of them were identical (Uzun et al., 2010; Uzun et al., 2011c). Fang and Roose (1997) found very low polymorphism in grapefruits based on ISSR data and notified that all grapefruits were derived from the same ancestral tree by mutation. There was no variation in grapefruits in other previous studies based on izozyme (Roose, 1988) and SSR (Luro et al., 2000) data. At the same way, Corazza-Nunes et al. (2002) detected high level of similarity in grapefruits. Most grapefruits, despite considerable variation in agronomical traits such as, rind and flesh color, fruit size, were nearly identical (Fig. 7). Cultivars with distinct morphological characters (pigmented or yellow flesh colour, seedy and seedless fruits) such as Henderson, Ruby, Duncan showed complete genetic similarity (Uzun et al., 2010).

Fig. 7. Although grapefruits have distinct fruit characters, low level of genetic variation found among them (from left to right; Davis Seedless, Shambar, Red Blush).

7. Conclusion

It is reported preservation of the genetic diversity represented in all the plant ecosystems throughout the world has become a major issue of international concern. The loss of increasingly large numbers of plant species through habitat destruction threatens the availability of a diverse plant germplasm base which will be needed to feed future generations. *Ex-situ* conservation of genetic resources of citrus was considered as imperative for this situation. (Bretting and Widrlechner, 1995, as cited in The Citrus and Date Crop Germplasm Committee, USA, CDCGC, 2004). Approaches to *ex-situ* conservation include methods like seed storage, field genebanks and botanical gardens. DNA and pollen storage also contribute indirectly to *ex-situ* conservation of genetic resorces. Advances in biotechnology, especially in the area of *in vitro* culture techniques and molecular biology provide some important tools for improved conservation and management of plant genetic resources (Rao, 2004).

It is suggested better understanding of genetic diversity and its distribution is essential for its conservation and use. It will help us in determining what to conserve as well as where to conserve, and will improve our understanding of the taxonomy and origin and evolution of plant species of interest. Information of these subjects is essential for collecting and use of any plant species and its wild relatives. Understanding genetic diversity that is present in collections is required to better management of conserved germplasm. Through improved characterization and development of core collections based on genetic diversity information, it will be possible to exploit the available resources in more valuable ways (Rao & Hodgkin, 2002).

Genetic variability in citrus is considered to be the result of many factors, such as hybridization, mutation and type of reproduction (mostly apomictic). The low intraspecific diversity found in cultivated species such as sweet orange contrasts with the high variability of agriculturally important traits such as ripening period and color and size of fruits (Herrero et al., 1996, as cited in Novelli et al., 2006). Understanding of genetic diversity in Citrus is essential for planning and application of breeding programs, establishing germplasm collection and carrying out molecular studies. It is also important for citrus researcher and breeders to arrange their future studies.

8. Acknowledgements

The authors wish to thank Kahraman GURCAN from Agricultural Biotechnology Department of Erciyes University for his valuable assistance.

9. References

Abkenar, A.A.; Isshiki, S. &Matsumoto, R. (2007). Comparative analysis of organelle DNAs acid citrus grown in Japan using PCR-RFLP method. *Genet. Res. Crop. Evol*. 55: 487–492.

Barkley, N.A.; Roose, M.L.; Krueger, R.R. & Federici, C.T. (2006). Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor. Appl. Genet.* 112, 1519–1531.

Barrett, H.C. & Rhodes, A.M. (1976). A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. *Syst. Bot.* 1: 105–136.

Breto, M.P.; Ruiz, C.; Pina, J.A. & Asins, M.J. (2001). The diversification of Citrus clementina Hort. Ex Tan., a vegetatively propagated crop species. *Mol. Phylo. Evol.*21:285-293.

Coletta-Filho, H.D.; Machado, M.A.; Targon, M.L.P.N.; Moreita, M.C.P.Q.D.G. &Pompea Jr., J. (1998). Analysis of the genetic diversity among mandarins (*Citrus* spp.) using RAPD markers. *Euphytica* 102: 133–139.

Corazza-Nunes, M.J.; Machado, M.A.; Nunes, W.M.C.; Cristofani, M. &Targon,M.L.P.N. (2002). Assessment of genetic variability in grapefruits (*Citrus paradisi* Macf.) and pummelos (*C. maxima* Burm. Merr.) using RAPD and SSR markers. Euphytica 126: 169–176.

Cottin, R. (2002) Citrus of the World, A Citrus Directory Version 2. SRA INRA-CIRAD, France

Deng Z.N.; Gentile, A.; Domina, F.; Nicolosi, E. &Tribulato, E. (1995) Selecting lemon protoplasts for insensitivity to *Phoma tracheiphila* toxin and regenerating tolerant plants. *J. Am. Soc. Hortic. Sci.* 120:902–905

FAO. (2009). Food and Agricultural Organization of the United Nations, http://faostat.fao.org

Fang, D.Q. & Roose, M.L. (1997). Identification of closely related Citrus cultivars with inter-simple sequence repeat markers. *Theor. Appl. Genet.* 95: 408–417.

Federici, C.T.; Fang, D.Q.; Scora, R.W. & Roose, M.L. (1998). Phylogenetic relationships within the genus Citrus (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. *Theor. Appl. Genet*. 96: 812–822.

Gulsen, O.& Roose, M.L. (2000). The origin of Interdonato lemon inferred from cpRFLP, SSR, isozyme and ISSR markers. *Proc. Int. Soc. Citricult. IX Congr*. pp. 158–159.

Gulsen, O. & Roose, M.L. (2001). Lemons: diversity and relationships with selected citrus genotypes as measured with nuclear genome markers. *J. Am. Soc. Hort. Sci.* 126: 309–317.

Herrero, R.; Asins, M.J.; Carbonell, A.E. & Navarro, L. (1996). Genetic diversity in the orange subfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. *Theor. Appl. Genet.* 92: 599–609.

Hodgson, R.W. (1967). Horticultural varieties of citrus. In: Reuther, W., Webber, H.J., Batchelor, L.D. (Eds.), *The Citrus Industry*, vol. 1. University of California Press, Berkeley, pp. 431–591.

Hvarleva, T.; Kapari-Isaia, T.; Papayiannis, L.; Atanassov, A.; Hadjinicoli, A. & Kyriakou, A. (2008). Characterization of citrus cultivars and clones in Cyprus through microsatellite and RAPD analysis. *Biotechnol. & Biotechnol. Eq.* 22: 787-794.

Jackson, L.K. & Futch, S.H. (2003). Ambersweet Orange. Fact Sheet HS-176, Horticultural Sciences Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida.

Kahn T.L.; Krueger, R.R.; Gumpf, D.J.; Roose M.L.; Arpaia, M.L.; Batkin, T.A.; Bash, J.A.; Bier, O.J.; Clegg, M.T.; Cockerham, S.T.; et al (2001). Citrus genetic resources in California: Analysis and recommendations for long-term conservation. Report No. 22. University of California Division of Agriculture and Natural Resources, Genetic Resources Conservation Program, Davis, CA, USA.

Luro F.; Laigret, F.; Bove, J.M. & Ollitrault, P. (1992). Application of random amplified polymorphic DNA (RAPD) to Citrus genetic and taxonomy. *Proc. Int. Soc. Citriculture* pp: 225–228

Luro, F.; Laigrent, F.; Bove, J.M. & Ollitrault, P. (1995). DNA amplified fingerprinting, a useful tool for determination of genetic origin and diversity analysis in *Citrus*. *HortScience* 30: 1063-1067.

Luro, F.; Rist, D. & Ollitrault, P. (2000). Sequence tagged microsatellites polymorphism: an alternative tool for cultivar identification and evaluation of genetic relationships in Citrus. *Proc. Int. Soc. Citricult. IX. Congr*. Florida, USA. pp: 170–171.

Nicolosi, E.; Deng, Z.N.; Gentile, A.; La Malfa, S.; Continella, G. & Tribulato, E. (2000). Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theor. Appl. Genet*. 100: 1155–1166.

Novelli, V.M.; Machado, M.A. & Lopes, C.R. (2000). Isoenzymatic polymorphismin *Citrus* spp. and *Poncirus trifoliata* (L.)) Raf. (Rutaceae). *Genet. Mol. Biol*. 23: 1163–1168.

Novelli, V. M.; Cristofani, M.; Souza, A. A. & Machado, M. A. (2006). Development and characterization of polymorphic microsatellite markers for the sweet orange (*Citrus sinensis* L. Osbeck). *Genetics and Molecular Biology*. 29: 90-96.

Orford S.J.; Steele Scott, N. &Timmis, J.N. (1995). A hypervariable middle repetitive DNA sequence from citrus. *Theor. Appl. Genet*. 91:1248- 52.

Pang, X.M.; Hu, C.G. & Deng, X.X. (2007). Phylogenetic relationship within Citrus and related genera as inferred fromAFLPmarkers. *Genet. Res. Crop. Evol.* 54: 429–436.

Qing-Qin, C.; Hai-Jun, M.; Xiao-Peng, W.; Hua-Lin, Y. & Xiu-Xin, D. (2007). Genetic diversity of male sterile and low fertility germplasm of Citrus revealed using SSR markers. *Chinese Journal of Agricultural Biotechnology* 4: 99–104.

Rao, V.R. & Hodgkin, T. (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture*, 68: 1-19.

Rao, N.K. (2004). Plant genetic resources: Advancing conservation and use through biotechnology. *African Journal of Biotechnology*, 3: 136-145.

Roose M.L. (1988). Isozymes and DNA restriction fragment length polymorphisms in *Citrus* breeding and systematics. *Proc. Int. Soc. Citricult. VI. Congr*. Margraf Scientific Books, Weikersheim, pp:155–165.

Russo, G.; Corona, M.G. & Vendramin, G.G. (2000). Identification of Clementine cultivars-ecotypes using molecular markers. *Proc. Int. Soc. Citricult. IX. Congr*. Florida, USA. pp:196-197.

Scora, R.W. (1975). On the history and origin of Citrus. *Bull. Torr. Bot. Club*. 102: 369–375.

Scora, R.W.; Kumamoto, J.; Soost, R.K. & Nauer, E.M. (1982). Contribution to the origin of the Grapefruit, *Citrus paradisi* (*Rutaceae*). *Syst. Bot.* 7:170-177.

Scora, R.W. (1988). Biochemistry, taxonomy and evolution of modern cultivated Citrus. *Proc. Int. Soc. Citricult. VI. Congr*. vol. 1. Margraf Publishers, Weikersheim, Germany, pp. 277–289.

Swingle, W.T. & Reece, P.C. (1967). The botany of citrus and its wild relatives. In: Reuther, W., Webber, H.J., Batchelor, L.D. (Eds.), *The Citrus Industry*, vol. 1.University of California Press, Berkeley, CA, USA, pp. 389–390.

Tanaka, T. (1954). Species problems in citrus. *Japanese Society for the Promotion of Science*, Ueno, Tokyo, 152 pp.

Tanaka, T. (1977). Fundamental discussion of Citrus classification. *Stud. Citrol.* 14: 1–6.

The Citrus and Date Crop Germplasm Committee, USA (CDCGC). (2004). Citrus and Date Germplasm: Crop Vulnerability, Germplasm Activities, Germplasm Needs. Citrus and Date Crop Germplasm Committee, USA. pp. 1-30.

Torres, A.M.; Soost, R.K. & Diedenhofen, U. (1978). Leaf isosymes as genetic markers in Citrus. *Am. J. Bot.* 65: 869–881.

Ulubelde, M. (1985). Turunçgillerin taksonomisi. Ege Bölgesi Zirai Arastırma Enstitüsü Yayınları No:55, p: 43 (in Turkish).

Uzun, A. (2009). Turunçgillerde genetik çeşitliliğin SRAP markırları ile karakterizasyonu (Characterization of genetic diversity of Citrus by SRAP markers) Ph.D. Thesis. University of Cukurova, Institute of Natural and Applied Sciencies. p. 369. (in Turkish)

Uzun, A.; Yesiloglu, T.; Aka-Kacar, Y.; Tuzcu, O. & Gulsen, O. (2009a). Genetic diversity and relationships within Citrus and related genera based on sequence related amplified polymorphism markers (SRAPs). *Scientia Horticulturae*, 121:306–312.

Uzun, A.; Gulsen, O.; Kafa, G.; Seday, U.; Tuzcu, O. & Yesiloglu, T. (2009b). Characterization for yield, fruit quality, and molecular profiles of lemon genotypes tolerant to 'mal secco' disease. *Scientia Horticulturae*, 122: 556-561.

Uzun, A.; Gulsen, O.; Yesiloglu, T.; Aka-Kacar, Y. & Tuzcu, O. (2010). Distinguishing grapefruit and pummelo accessions using ISSR markers. *Czech J. Genet. Plant Breed.* 46: 170–177.

Uzun, A.; Yesiloglu, T.; Aka-Kacar, Y.; Tuzcu, O.; Gulsen, O.; Seday, U. & Canan, I. (2011a). Genetic diversity and relationships of Clementine mandarin accessions as assessed using SRAP markers. *Acta Hort.* 892: 65-72.

Uzun, A.; Yesiloglu, T.; Polat, I.; Aka-Kacar, Y.; Gulsen, O.; Yildirim, B.; Tuzcu, O.; Tepe, S.; Canan, I. & Anil, S. (2011b). Evaluation of genetic diversity in lemons and some of their relatives based on SRAP and SSR markers. *Plant Mol. Biol. Rep.* 29: 693–701.

Uzun, A.; Yesiloglu, T.; Aka-Kacar, Y.; Tuzcu, O.; Gulsen, O.; Kafa, G.; Yildirim, B. & Anil, S. (2011c). Genetic diversity of grapefruit and pummelo accessions based on SRAP markers. *Acta Hort.* 892: 49-56.

Yong, L.; De-Chun, L.; Bo, W. & Zhong-Hai, S. (2006). Genetic diversity of pummelo (*Citrus grandis* Osbeck) and its relatives based on simple sequence repeat markers. *Chinese J. of Agr. Biotechnology,* 3:119-126.

Webber, H.J. (1943). Cultivated varieties of citrus. In: Webber HJ, Batchelor LD (eds) *The Citrus Industry,* vol 1. University of California Press, Berkeley. pp: 475–668.

Webber, H.J. 1967. History and development of the Citrus industry. In: Reuther, W., Webber, H.J., Batchelor, L.D. (Eds.), *The Citrus Industry*, vol. 1. University of California Press, Berkeley, pp. 1–39.

12

Determination of Genetic Variation Between Populations of *Abies nordmanniana* subsp. *bornmulleriana* Mattf According to some Seed Characteristics

Hakan Sevik[1], Zeki Yahyaoglu[2] and Ibrahim Turna[2]
[1]*Kastamonu University, Forestry Faculty, Landscape Architecture Dep., Kastamonu*
[2]*Karadeniz Technical University, Forestry Faculty, Silviculture Dep., Trabzon*
Turkey

1. Introduction

The success of any sustainable reforestation program, among other things, hinges on a continuous supply of high quality seeds for the production of the desired quantity of seedlings in nurseries or for successful stand establishment by direct sowing out in the field. What is seed quality then? Seed quality is defined as "a measure of characters or attributes that will determine the performance of seeds when sown or stored" (Hampton 2002). It is a multiple concept encompassing the physical, physiological, genetic, pathological and entomological attributes that affect seed lot performance (Basu, 1995).

Several factors affect the production of high quality seeds, such as insect infestation (El Atta 1993, Dajoz 2000, Bates et al. 2000, 2001), pollination failure and post-zygotic degeneration (Owens et al. 1990, El-Kassaby et al. 1993), infection by seed borne pathogens (Pritam and Singh 1997), environmental conditions during seed development (Gutterman 2000) as well as the genetic constitution (Bazzaz et al. 2000).

Genetic diversity is the richness of the hereditary information in the gene pole of one species. High level of inter-species genetic diversity is an assurance for adaptation to changing environmental conditions, an indication for adaptation potential of the species and an important part of the ecosystem stability. Also genetic diversity is a raw material for tree improvement studies. As such, most of the researches about the genetic diversity are in high priority in forest trees improvement programs (Şevik et al., 2010a, 2010b).

Genetic variation is the fundamental component, which ensures survival and thus the stability of forest ecosystems as its quantity and quality determines the potential of population to adapt the changing in environmental condition. This is particularly important with changing population and climatic condition and when the long-term stability of forest ecosystems is increasingly threatened by environmental stress. Thus, a genetic characterization of natural forest resources is the first step necessary for a better understanding of genetic resources for implementation of insitu and exsitu conservation activities (Şevik, 2010; Turna et al., 2006; Şevik et al. 2010a).

Up to now, in Turkey, studies about genetic diversity of the main forest trees have been concentrated on pine species, neglecting other main forest tree species. Turkish fir is among the one of the neglected species.

Turkish fir (*Abies nordmannianan* subsp. *bornmülleriana*) has a special importance for Turkey because of its increasing economic value in marketplace and decorative characteristic in landscape architecture. Furthermore, being an endemic species for Turkey, very decorative species, for this reason the species is the most widely preferred Noel tree in the world (Şevik et al, 2011). Turkish fir is distributed from Kızılırmak River to Mount Uludağ in Western Blacksea region, particularly in Ayancık, Ilgaz Mountains, Bolu Seben Mountains, Boyabat-Göktepe forests, Abant and Mount Uludağ. Stands of fir species occupy roughly 600.000 ha at Turkey (Anonymous, 2006).

The objectives of this study were to investigate the Genetic diversity among Turkish fir populations in Turkey, and determine the extent of between population variation, using 13 different morphological characters.

2. Materials and methods

Seed collection and sowing: Open pollinated seed materials from seventeen different natural populations of Turkish fir collected from Western Black Sea Region. Locations and description of the studied population are indicated in Fig. 1 and Table 1 in this study.

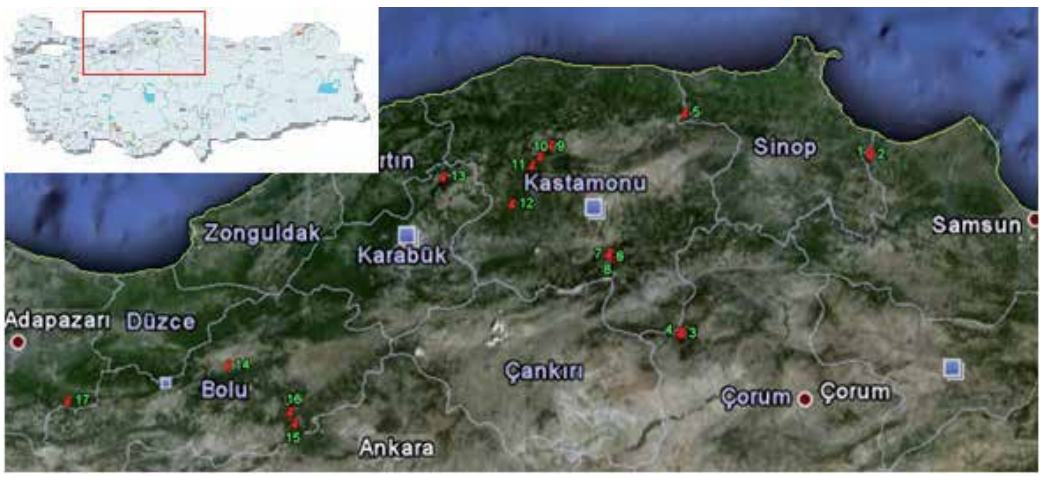


Fig. 1. Locations of the populations

Seed and seedling morphological variables studied and data collection: In this study, width, thickness, length and weight of seeds, carpel width, length and weight, carpel scape width and length, wings width and length were determined from total 303 sample trees. All length and width were measured with digital micro-compass (0,01 mm) from 10 samples for each sample tree. All weight was measured with digital weighing machine (0,001 gr).

Statistical Analyses: Data were subjected to multi-way analysis of variance, Duncan test and Hierarchical Cluster analysis with SPSS statistical package program. Relationships between 13 related characters were tested using correlation analyses.

Pop. No	Population Name	City	Number of Sample Trees	Altitude (m)	Longitude (E)	Latitude (N)
1	Bafra1	Samsun	10	828	35°21'18"	41°34'01"
2	Bafra2	Samsun	10	1012	35°21'33"	41°33'28"
3	İskilip1	Amasya	20	1673	33°46'11"	41°22'36"
4	İskilip2	Amasya	20	1852	34°13'34"	40°49'01"
5	Türkeli	Sinop	13	1348	34°16'15"	41°44'58"
6	Ilgaz1	Kastamonu	20	1430	33°49'17"	41°09'27"
7	Ilgaz2	Kastamonu	20	1624	33°49'11"	41°08'60"
8	Ilgaz3	Kastamonu	20	1995	33°50'58"	41°07'47"
9	Ballıdağ1	Kastamonu	20	1056	33°29'02"	41°37'11"
10	Ballıdağ2	Kastamonu	20	1374	33°25'29"	41°34'12"
11	Ballıdağ3	Kastamonu	20	1640	33°22'37"	41°31'58"
12	Samatlar	Kastamonu	20	1497	33°15'32"	41°22'06"
13	Eflani	Karabük	20	1102	32°51'45"	41°29'02"
14	Aladağ	Bolu	10	968	31°37'15"	40°40'21"
15	Kıbrıscık2	Bolu	20	1499	32°00'42"	40°25'46"
16	Kıbrıscık1	Bolu	20	1791	32°02'22"	41°28'43"
17	Göynük	Bolu	20	1270	30°41'27"	40°30'08"

Table 1. Description of the studied populations in Turkey

Moreover, collected data was determined with Penrose formule. Data were standardized before the calculations and the morphological distance among populations were estimated as;

$$Z_{i,k} = \frac{(X_{i,k} - \bar{x})^2}{S_k}$$

Where $Z_{i,k}$ is standardized values of the k^{th} characteristics of the i^{th} population, $X_{i,k}$ is original average of the k^{th} characteristics of the i^{th} populations for the k^{th} characteristics and S_k is the standard deviation of the studied populations for the k^{th} characteristics (Şevik, 2005, 2010).

$$D_{i,j} = \sum_{k=1}^{p} \frac{(\mu_{ki} - \mu_{kj})^2}{p.V_k}$$

Where, D_{tj} is the morphological distance between the i^{th}, population and the j^{th} populations, n is the number of studied characteristics, μ_{kj} is the standardized values of the k^{th} of the I^{th} population, μ_{kj} is the standardized values of the k^{th} characteristics of the j^{th} population, V_k is the variance of standardized averages of the k^{th} characteristics (Yahyaoğlu et al, 2001) was applied by standardized values in SPSS statistical package program (Şevik, 2010).

3. Results

The analysis of variance showed that there were significant differences among populations at 0.01 for seed width and 0.001 for other characters. Mean values and multiple comparisons of studied morphological characters shown in Tables 2.

Population Name	Carpel Length (mm)		Scale Length (mm)		Carpel width (mm)		Scale width (mm)	
Bafra1	33.31±0.81	a	25.21±0.82	ab	28.70±1.01	abcd	**4.87±0.14**	**a**
Bafra2	33.32±0.98	a	**24.70±0.92**	**a**	**27.63±1.12**	**a**	5.12±0.21	abc
İskilip1	36.41±0.68	cdef	27.36±0.61	bcde	31.01±0.71	cde	5.08±0.11	abc
İskilip2	36.18±0.85	cdef	26.50±0.63	abcde	31.07±0.90	de	5.00±0.12	ab
Türkeli	**33.54±0.87**	**a**	25.69±0.5	abc	27.86±0.85	a	5.29±0.14	abcd
Ilgaz1	33.71±0.78	ab	25.33±0.66	ab	28.39±0.60	ab	5.18±0.13	abc
Ilgaz2	36.39±0.95	cdef	27.83±0.67	cde	29.87±0.82	abcde	5.41±0.12	bcd
Ilgaz3	34.86±0.64	abcd	26.95±0.65	bcde	28.47±0.82	abc	5.10±0.12	abc
Ballıdağ1	36.99±0.73	ef	**28.69±0.62**	**e**	**32.17±0.69**	**e**	**5.71±0.16**	**d**
Ballıdağ2	36.56±0.55	cdef	28.06±0.66	de	31.53±0.69	e	5.50±0.11	cd
Ballıdağ3	36.30±0.56	cdef	27.73±0.53	cde	30.87±0.61	bcde	5.65±0.12	d
Samatlar	36.06±0.52	bcdef	27.77±0.52	cde	31.56±0.68	e	5.67±0.12	d
Eflani	34.04±0.53	abc	26.34±0.58	abcd	29.97±0.63	abcde	5.18±0.12	abc
Aladağ	**37.43±0.84**	**f**	28.21±1.12	de	32.07±0.66	e	5.53±0.15	cd
Kıbrıscık2	34.12±0.76	abc	25.79±0.69	abc	30.86±0.64	bcde	5.29±0.17	abcd
Kıbrıscık1	36.65±0.70	def	27.93±0.58	cde	31.93±0.72	e	5.45±0.15	bcd
Göynük	36.11±0.70	bcdef	28.15±0.68	de	30.67±0.72	bcde	5.45±0.11	bcd
Av.	35.41±0.72		26.96±0.65		30.27±0.74		5.32±0.13	
F	3,194***		3,007***		3,420***		3.42***	

Table 2. Mean values of studied morphological characters and results of Duncan test.

According to Table 2, Population of Kıbrıscık2 is in the first homogeny group according to all characters and population of Ilgaz1 is too except wing length. These populations showed lowest performance for thirteen characters. Populations of Aladağ, Ballıdağ1 and Ballıdağ2 showed highest performance. Aladağ population (except wing width), Ballıdağ1 and Ballıdağ2 populations (except carpel scape width) are in the last homogeny group according to Duncan test. These populations showed the highest performance almost for all characters. The mean values and standard deviation of morphological characters by populations are shown in Table 2 (Şevik, 2010).

Average carpel length is 35,41 mm, carpel width is 30,27 mm, scale length is 26,96 mm and scale width is 5,32 mm. According to results of variance; carpel length 12,4%, carpel width

16,4%, scale length 16,2%, scale width 17,2% change of minimum values to maximum values. Maximum values of scale length (28,69 mm), carpel width (32,17 mm) and scale with (5,71 mm) determined to population of Ballıdağ1. Minimum values of scale length (24,7 mm) and carpel width (27,63 mm) determined to Population of Bafra2. Minimum scale width is 4,87 mm (Bafra1), minimum carpel length is 33,54 mm (Türkeli). Maximum carpel length is 37,43 mm determined to population of Aladağ.

Popul. Name	Carpel Scape Length (mm)		Carpel Scape Width (mm)		Carpel Weight (mg)		Wing Length (mm)		Wing Width (mm)	
Bafra1	4.79±0.12	bcdef	1.61±0.07	cdef	275.64±20.03	ab	15.31±0.47	bc	**14.04±0.51**	**a**
Bafra2	**4.17±0.12**	**a**	1.41±0.09	ab	**245.60±22.69**	**a**	14.74±0.91	ab	14.05±0.66	a
İskilip1	4.86±0.11	def	**1.75±0.06**	**f**	353.41±14.83	de	16.97±0.50	cde	15.85±0.42	c
İskilip2	4.96±0.13	def	1.70±0.05	ef	**358.38±20.35**	**e**	17.37±0.59	def	**16.40±0.39**	**c**
Türkeli	4.72±0.12	bcdef	1.62±0.05	cdef	273.34±18.87	ab	**13.27±0.58**	**a**	14.45±0.49	ab
Ilgaz1	4.41±0.11	ab	1.41±0.04	ab	252.23±11.26	a	16.25±0.43	bcde	14.47±0.30	ab
Ilgaz2	4.89±0.14	def	1.58±0.04	bcdef	299.35±16.62	abcd	16.97±0.64	cde	15.98±0.46	c
Ilgaz3	4.43±0.07	abc	1.51±0.06	abcd	251.07±12.66	a	16.31±0.55	bcde	15.64±0.30	bc
Ballıdağ1	4.74±0.12	bcdef	1.54±0.05	abcde	330.65±14.27	cde	17.60±0.46	def	15.75±0.32	c
Ballıdağ2	**5.13±0.12**	**f**	1.54±0.04	abcde	339.47±11.48	cde	17.65±0.52	def	15.81±0.26	c
Ballıdağ3	4.58±0.12	bcd	1.62±0.05	cdef	299.12±12.54	abcd	16.56±0.41	cde	15.20±0.31	abc
Samatlar	4.84±0.15	cdef	1.51±0.04	abcd	325.96±17.56	bcde	17.09±0.46	cdef	15.38±0.32	bc
Eflani	4.64±0.10	bcde	**1.40±0.05**	**a**	286.68±12.46	abc	16.00±0.46	bcd	15.10±0.36	abc
Aladağ	4.94±0.24	def	1.68±0.07	def	315.50±19.19	bcde	18.06±0.66	ef	15.63±0.49	bc
Kıbrıscık2	5.07±0.10	ef	1.49±0.05	abc	272.18±14.67	ab	17.60±0.54	def	14.39±0.40	ab
Kıbrıscık1	4.99±0.09	def	1.59±0.04	cdef	352.06±16.43	de	**18.90±0.52**	**f**	15.82±0.34	c
Göynük	4.90±0.12	def	1.54±0.06	abcde	347.25±19.36	de	17.26±0.61	def	16.17±0.36	c
Av.	4.77±0.12		1.56±0.05		304.58±15.79		16.7±0.54		15.3±0.38	
F	3,808***		3,604***		5,675***		4,703***		3,376***	

Table 2. (Continue). Mean values of studied morphological characters and results of Duncan test.

Average carpel scape length is 4,77 mm, carpel width is 1,56 mm, carpel weight is 304,58 mg, wing length is 16,7 mm, wing width is 15,3 mm. According to results of variance; scale scape length 14,4%, scale scape width 25%, carpel weight 45,9%, wing length 42,4%, wing width 15,2% change of minimum values to maximum values.

According to the table, population of Bafra1 had shown minimum values of carpel scape length (4,17 mm) and carpel weight (245,60 mg). Iskilip1 had shown Maximum values of carpel weight (358,38 mg) and wing width (16,4 mm). Minimum values; wing width is 14,04 mm (Bafra1), carpel scape width is 1,4 mm (Eflani) and wing length is 13,27 mm (Türkeli).

Maximum values; wing length is 18,9 mm (Kıbrıscık1), carpel scape width is 1,75 mm (Iskilip1) and carpel scape length is 5,13 mm (Ballıdağ2).

The mean values and standard deviation of morphological characters by populations are shown in Table 2 (Şevik, 2010).

Population Name	Seed Length (mm)		Seed Width (mm)		Seed Thickness (mm)		Seed Weight (mg)	
Bafra1	11.00±0.34	ab	5.74±0.11	abc	3.97±0.09	ab	81.58±5.41	abcde
Bafra2	10.84±0.39	a	5.78±0.12	abcd	3.82±0.14	a	85.84±7.66	abcde
İskilip1	11.63±0.28	abcd	6.12±0.13	de	3.92±0.08	ab	84.55±4.38	abcde
İskilip2	11.46±0.22	abc	5.93±0.12	abcd	3.90±0.07	ab	72.94±4.29	ab
Türkeli	11.44±0.18	abc	5.95±0.14	abcd	3.94±0.10	ab	79.97±4.85	abc
Ilgaz1	**10.83±0.23**	**a**	5.65±0.12	ab	**3.73±0.08**	**a**	74.71±3.73	ab
Ilgaz2	12.01±0.28	cdef	5.86±0.12	abcd	4.03±0.08	ab	95.57±4.60	defg
Ilgaz3	11.48±0.16	abc	5.65±0.07	ab	3.77±0.06	a	80.41±3.05	abcd
Ballıdağ1	12.46±0.21	ef	6.07±0.09	cde	4.22±0.09	ab	103.29±4.53	fg
Ballıdağ2	**12.57±0.17**	**f**	6.05±0.08	cde	4.33±0.08	b	**109.11±3.82**	**g**
Ballıdağ3	12.36±0.22	def	5.85±0.09	abcd	4.05±0.07	ab	94.71±3.90	cdefg
Samatlar	11.77±0.25	bcde	5.85±0.08	abcd	3.83±0.09	a	91.68±4.66	cdef
Eflani	11.16±0.24	ab	**6.36±0.12**	**e**	3.90±0.07	ab	87.96±4.29	bcde
Aladağ	11.99±0.26	cdef	5.99±0.12	bcd	4.07±0.14	ab	95.42±7.44	defg
Kıbrıscık2	11.06±0.19	ab	**5.61±0.08**	**a**	3.77±0.05	a	**71.72±2.78**	**a**
Kıbrıscık1	11.67±0.25	bcde	5.76±0.13	abcd	**4.34±0.43**	**b**	83.27±4.59	abcde
Göynük	12.15±0.24	cdef	5.96±0.08	abcd	3.93±0.1	ab	96.80±4.65	efg
Av.	11.64±0.23		5.89±0.1		3.97±0.1		81.58±4.05	
F	5,123***		3,451***		1,789**		5,942***	

Table 2. (Continue). Mean values of studied morphological characters and results of Duncan test.

Average values of seed length (11,64 mm), seed width (5,89 mm), seed thickness (3,97) and seed weight (81,58 mg) are shown in the table. According to the table minimum seed length (10,83 mm) and seed thickness (3,73 mm) values are determined to population of Ilgaz1, minimum seed width (5,61 mm) and seed weight (71,72 mg) values are determined to population of Kıbrıscık2. Maximum values of seed length is 12,57 mm (Ballıdağ2), seed width is 6,36 mm (Eflani), seed thickness is 4,34 mm (Kıbrıscık1) and seed weight is 109,11 mg (Ballıdağ2).

On the cluster dendrogram constructed on the basis of Euclidean distances with the use of the nearest neighbourhood method for 13 quantitative morphological traits, two distinct

groups can be noticed: the first is İskilip1, İskilip2, Ballıdağ2, Göynük, Samatlar, Aladağ, Kıbrıscık1 and the others. The second group cans distinct two groups, Ilgaz1 and others. According to these results, it can be said that there are three main groups (Figure 2 and 3).

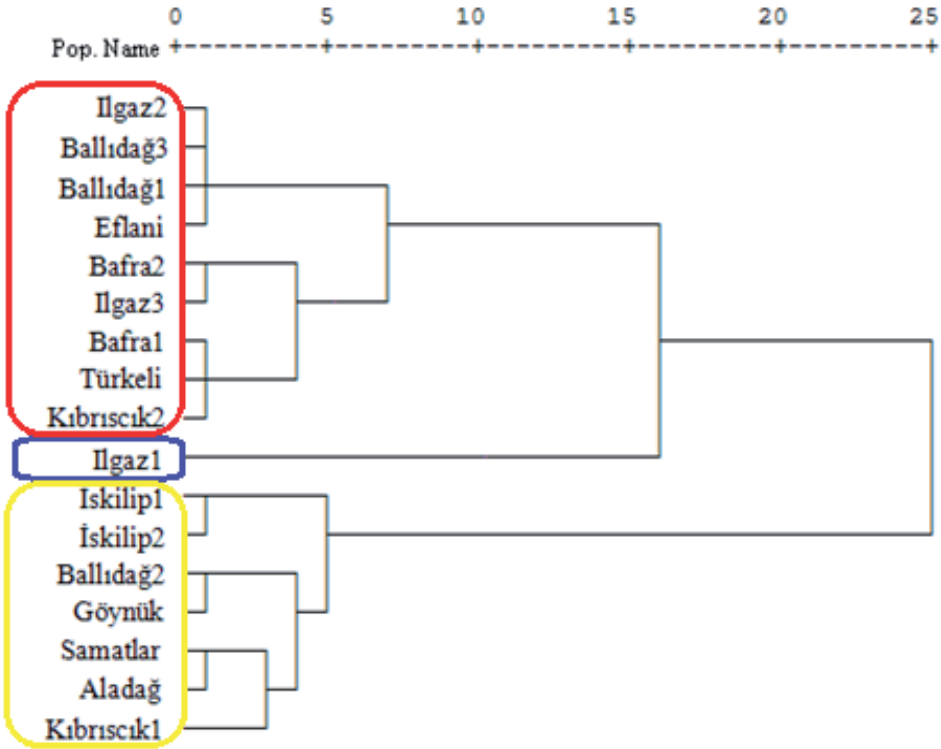


Fig. 2. Dendrogram of 17 population of Turkish fir based on 13 morphological traits.

According to results of cluster analysis, red color is the first group, blue color is the second group and the yellow color is the third group in the figure 3 (Şevik, 2010).

Some populations are geographically and genetically close to each other like, Bafra1 and Bafra2, Ilgaz2 and Ilgaz3, İskilip1 and İskilip2 populations. Some of them are geographically close to each other even though genetically different from each other. For example Ballıdağ1 and Ballıdağ2, İskilip1 and İskilip2 populations. On the contrary, some populations are genetically close to each other even though geographically different from each other. For example Bafra1 and Kıbrıscık2, İskilip1 and Göynük populations (Figure 2 and 3).

The highest 16 values calculated between Ilgaz1 and the other populations. Maximum 5 values are between the populations of Ilgaz1 and Eflani (10,3635), Kıbrıscık1 (9,9517), Ilgaz3 (9,2148), Türkeli and Ilgaz2 (8,4679). Minimum 5 values are 0,3029 (Bafra1 and Samatlar), 0,4078 (Bafra1 and Göynük), 0,4107 (Ballıdağ3 and Samatlar), 0,4673 (Bafra1 and Ilgaz2) and 0,5038 (Kıbrıscık2 and Göynük).

Results of correlation analyses are shown in Table 4. According to Table 4 there were positive significant correlation was found between all characters except carpel scape width and seed length, seed thickness, seed weight. The correlation between carpel scape length

and carpel scape width is significant at the 0.05 level, other all correlations are significant at the 0.01 level (Table 4).

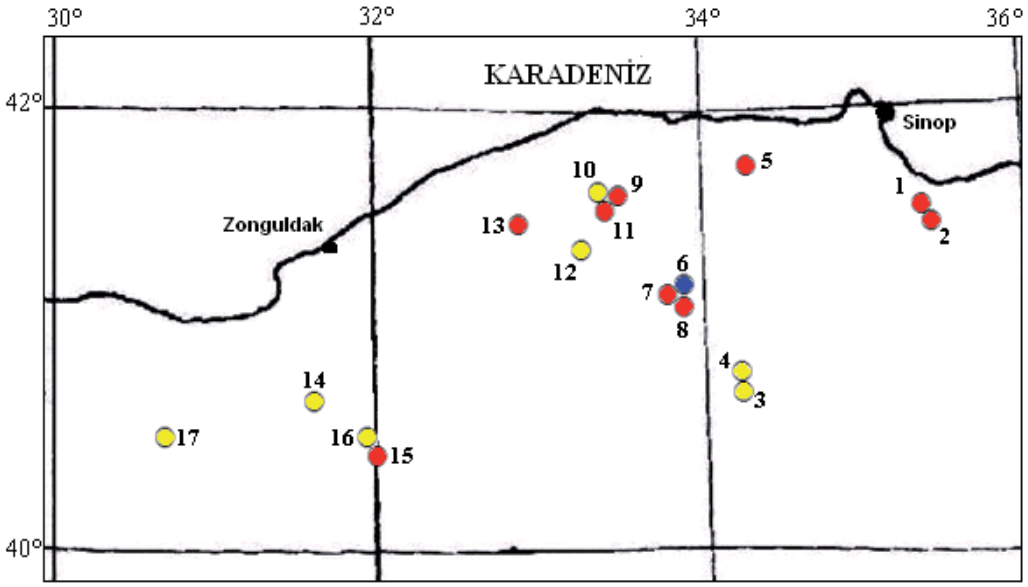


Fig. 3. Geographic positions of populations and results of cluster analysis

Morphological distance and grouping according to Penrose formula are shown that Table 3.

Pop. No	1	2	3	4	5	6	7	8
2	1.0171							
3	0.8408	1.0587						
4	0.8715	1.2715	0.6238					
5	1.1440	1.4252	1.5351	1.5285				
6	**7.1725**	**5.7758**	**7.4618**	**7.2740**	**8.8849**			
7	**0.4673**	1.5966	1.3288	1.3045	1.6894	**8.4679**		
8	0.9812	1.4185	0.8796	0.8711	1.7620	**9.2148**	1.0019	
9	0.9410	1.2547	0.7509	1.0335	1.6805	**7.2567**	0.6187	1.0099
10	1.8281	2.0630	2.8355	2.1007	2.7878	**4.6134**	1.6283	1.9982
11	0.6217	2.0264	1.5577	1.8446	2.3340	**8.0759**	0.5578	1.3256
12	**0.3029**	1.3819	1.1421	1.2455	1.5132	**8.0839**	0.7395	0.6935
13	1.5400	2.5653	0.8524	1.9061	1.9414	**10.3635**	1.3006	1.2208
14	0.5383	0.8897	1.0649	1.2452	1.0398	**5.2708**	0.9439	1.5747
15	0.9053	1.1575	1.1795	0.8243	1.1764	**7.5639**	0.5617	0.5652
16	1.3048	2.1365	1.3576	1.6116	1.1450	**9.9517**	1.1685	0.7097
17	**0.4078**	1.0211	1.0387	0.8375	1.4796	**7.0460**	0.5381	0.7332

Table 3. Morphological distance among populations according to Penrose Formula

Pop. No	9	10	11	12	13	14	15	16
10	1.9756							
11	1.1810	1.7470						
12	1.2355	1.8239	**0.4107**					
13	0.8475	3.3882	1.4892	1.5608				
14	0.8184	1.3315	0.8540	0.7276	1.7854			
15	0.5766	0.9610	1.2016	0.9795	1.3299	0.9168		
16	1.3074	2.0367	1.4842	1.1814	0.7464	1.6195	0.6667	
17	1.0942	1.2490	1.0541	0.6694	1.6872	1.0009	**0.5038**	1.1633

Table 3. (Continue). Morphological distance among populations according to Penrose Formula

	CL	SL	CW	SW	CSL	CSW	CWe	WL	WW	SeL	SeW	SeT
SL	**0.81****											
CW	0.69**	0.55**										
SW	0.51**	0.51**	0.60**									
CSL	0.38**	0.30**	0.49**	0.29**								
CSW	0.27**	0.29**	**0.19****	0.20**	**0.14***							
CWe	**0.78****	0.70**	0.71**	0.45**	0.48**	0.38**						
WL	0.60**	0.47**	**0.80****	0.43**	0.38**	**0.18****	0.63**					
WW	0.77**	0.69**	0.58**	0.35**	0.30**	0.28**	0.71**	0.63**				
SeL	0.62**	0.57**	0.57**	0.41**	0.39**	0.09^{ns}	0.54**	0.41**	0.55**			
SeW	0.39**	0.41**	0.40**	0.21**	0.22**	0.20**	0.43**	0.27**	0.51**	0.46**		
SeT	0.32**	0.31**	0.34**	0.30**	0.25**	0.04^{ns}	0.30**	0.28**	0.28**	0.44**	0.28**	
SeWe	0.64**	0.62**	0.56**	0.42**	0.38**	0.06^{ns}	0.58**	0.44**	0.60**	0.77**	0.55**	0.45**

ns: Non significant, **: significant at the 0.01 level, *: significant at the 0.05 level.

Table 4. Pearson correlation coefficients among 13 morphological characters

According to results of correlation analysis the highest relations are between the carpel length and seed length (0,81), carpel width and wing length (0,80) and carpel length and carpel weight (0,78). The minimum values are determined between carpel scape width and carpel scape length (0,14), wing length (0,18) and carpel width (0,19).

4. Discussion

According to results of the cluster analysis and variance analysis, Ilgaz1 population is very different to other populations. It could be because of its longitude and different ecological and genetical material condition. Results of the cluster analysis (Figure 2) were well accordance with morphological distances. For instance, morphological distances of Ilgaz1 were the highest than the others. Similarly population of Ilgaz1 is very different to other populations according to cluster analysis. These results could be used in preparation of

gene map, seed transfer zones, determination of breeding populations, gene conservation areas, geographic variation and resulting of provenance trials of the species in short period. Preparation of forest gene maps and determination of seed transfer zones and geographical variation by morphological distance were also suggested by Yahyaoğlu et al 2001.

Genetic variation can be determined with morphological characters (Güney, 2009; Kulaç et. al., 2010; Şevik, 2010), isosymes analysis (Bilgen and Kaya, 2007; Turna, 2003) and DNA markers (Clark et. al., 2000; Goldstein, 1995). Many researchers use these methods for determination to genetic variation on *Abies* species; Messaoud et all. (2007) *Abies balsamea,* Okada et all., (1973) *Abies sachalinensis,* Parker et all. (1981) *Abies balsamea* and *Abies lasiocarpa,* Kolotelo (1998) *Abies amabilis*, *Abies grandis* and *Abies lasiocarpa* e.c.

Shea (1990) reported that the variation among the populations is small (1,3%) but significant in *Abies Lasiocarpa*. Sorensen and Franklin (1977) reported that, year effect including interactions with places and trees in places made up an estimated 45 % of the variance in seed weight and 25% of the variance in cotyledon number. Among population genetic variance was much lower than within population variance, ranging from 6.6 to 6.8% for drought resistance traits to 7.8–14.0% for bud-break dates and a maximum of 10.0–17.9% for height growth traits to *Abies alba*. Therefore, genetic variance was predominantly within population (Sagnard et al, 2002).

The average genetic distance for all pair-wise comparisons between the ten populations of *Abies alba* in Italy was 0.014 (Parducci and Szmidt, 1997). 7,3% of the total genetic variation was due to differences among populations for gymnosperms (Hamrick et. al., 1992) and 10% for eigth Abies species (Shea and Furnier 2002). 13.3% of the total diversity is distributed among populations in *Abies alba* (Vendramin et al. 1999). Great variation was observed in the heterozygosity among the population studied and ranged from 0.010 *(A. pinsapo)* to 0.328 *(A. cephalonica).* The inter population genetic diversity was about 26% of the total genetic diversity. The average coefficient of gene differentiation (Gst) was 0.255, which means that approximately 26% of the total diversity of the Mediterranean firs exist among the populations. In particular, the geographical Area III (Turkey) has scored the highest value of Gst (25.8%), (Scaltsoyianne, 1999).

The proportion of genetic diversity among the populations of *Abies sachalinensis* is 1,5 % (El-Kassaby, 1992), populations of *Abies mariesii* is 2,6 % (Suyama et al. 1992) and populations of *Abies cephalonica* is 4.8% (Fady and Conkle 1993).

Conte (2004) reported that ANOVA analysis of *Abies nebrodensis* indicated that most of genetic variation resides within subsets (84%). More than 10% of the total genetic diversity was due to differences among populations of *Abies nebrodensis* (Vicario et, al., 1995).

Total percentage of genetic variation present in the population explained by interplot or among subpopulation differences is 0,35% of *Abies fraseri*. Thus, more than 99% of the genetic variation is due to within plot (i.e. tree to tree) variation (Diebel and Feret 1991). Most of the genetic diversity lies within populations to *Abies cephalonica* (Fady and Conkle 1993). Less than 10% of the total observed variation appeared among populations of *Abies cephalonica* (Hamrick, 1989) and the variation among the populations is 11% in *Abies alba* (Vicario et al. 1995). Vendramin et al. (1999) reported that 13.3% of the total diversity is distributed among populations in *Abies alba.* On average, the genetic diversity among populations of *Abies* species has been found to be 6,3 % (Hamrick et al. 1992).

The high within-population genetic diversity and low among-population differentiation observed in conifers have been attributed to common lifehistory traits, such as longevity and extensive gene flow (Hamrick et al., 1992; Streiff et al., 1998). The biogeographic history of a species should also contribute significantly to current patterns of genetic variation (Planter et al. 2000).

Despite the comparatively low levels of allozyme variation and the small genetic distances between populations, geographical differentiation among silver fir populations at different spatial scales could be demonstrated with markers (Konnert and Bergmann 1995). A large difference in cone legth, seed germination and seed weight was observed among the sites and among mother trees to *Abies sachalinensis* in Japan (Okada, 1973).

Contrary other Abies species, there are not enough study for Turkish fir. For this, it can be suggested that all populations, especially Ilgaz1 population, be considered for a gene conservation program. Also, future studies are necessary to provide deeper insights in to the subject. It may be concluded from the present study that studied characteristic were the important factors on morphological distance among populations. Ecological and geographical differentiation is important factors which influence the breeding and sampling strategies of tree crops. It is also essential to consider the relationship between population structure in natural and domesticated populations (Chalmers et al. 1992; Şevik, 2010; Şevik et al. 2011b). Results of this study could be taken into consideration in silvicultural purpose (afforestation, artificial regeneration) and breeding strategies (i.e. determination of breeding populations, gene conservation areas, seed transfer zones, seed sources and geographic variation, resulting of provenance trial; establishment of seed orchard) of this species.

Generally, our results show that large genetic diversity exist in Turkish fir to explain its great ecological plasticity and evolutionary. This results of study showed that the populations are not homogeneous with regard to the morphological characteristics. Populations consist of the trees having more or less different characteristics. The reason of the fact that the grouping and differences existed among the studied population in terms of the morphological characters may explain that there were different origins or varieties forming the Turkish fir stands. Variation in most of these characteristics appeared to be related altitude, divergent gene and genotype frequencies.

As is known, the morphological and physiological characteristics of forest trees are inherited. These features, with the growing effects of climate and environmental conditions can vary very little. As an example; needle length, the number of needles, cones, seed and leaf characteristics, branching characters as show some morphological features. In fact, many researchers in determining the genetic diversity of forest trees, one or a few of the uses of these characters. (Matziris, 1989; Cregg, 1994; Matziris, 1984; Komar, 2000; Fan ve Grossnickle, 1999; Lamhamedi, 2000; Matziris, 1997; John, 1948; Schmidtling et. al., 2005; Kathleen and Furnier 2002; Erkan, 2008; Bilir, 2002; Tylek ve Walczyk 2002, Güney et al., 2011, Kulaç et al. 2011a, 2011b; Turna and Güney, 2009; Turna, 2003, 2004). Seed size, parameters in terms of quality seeds is the most widely used classification also reflects the morphological diversity of values within and between populations. (Güney, 2009). Seeds in the trees, cones and cone elements, least affected by environmental conditions and thus genetic structure of the tree is considered the beginning of the elements that represent the most healthy way. Therefore especially in studies of genetic diversity of seeds, cones, and cones from the studies of the elements has a special place because it results is quite healthy (Turna et al., 2009).

Erkuloğlu (1993) reported that average weight of *Abies bornmulleriana* Mattf. Seeds from Bolu are 57,13 mg. Okada et al., (1973) *Abies sachalinensis* Masters in their study, in Japan, studied on 7 population and thousand grain weight of seed on the basis of population varied between 9.3 g to 12.3 g have identified. Also Skrzyszewska and Chlanda (2009) *Abies alba* Mill. thousand grain weight of seed on the basis of population, have found varied between 38.92 g and 53.27 grams. Edwards (1982), Fowells (1965)'to refer to *Abies lasiocarpa* var. *arizonica* (Merriam) Lemra. subalpine fir, compared to other types of seeds, the seeds of its much larger that represents about 70%. Kolotelo (1998) *Abies amabilis* seed weight (Dougl.) Forbes has changed between 25 mg and 55.6 mg and is the average of 34.5 mg, also *Abies grandis* Lindl. varied between 17.5 mg and 27.6 mg and average is 21.7 mg and *Abies lasiocarpa* (Hook.) Nutt. average of 7.2 mg to 18.5 mg and 12 mg of states that have changed. According to these results, the seeds of Uludag Fir, *Abies alba* Mill., *Abies amabilis* (Dougl.) Forbes, be said to be heavier than the seeds of *Abies lasiocarpa* (Hook.) Nutt. and *Abies grandis* Lindl. Also Sorensen and Franklin (1977) *Abies Procera* Rehd. state that the seed weight varied between 33 mg and 102.6 mg. In our study, the average seed weight ranged between 71.72 mg and 109.11 mg, the average was determined to be 81.58 mg. Gökmen (1970), indicates that *Abies nordmanniana* Mattf. seeds is 1 cm in length. According to these results, seeds of Uludag Fir with *Abies Procera* Rehd. seeds appear to be close to each other in weight. Franklin (1974) *Abies Procera* Rehd. carpel averaging 2.5 x 3 cm in size, seeds indicates that the average size of 12 x 6 mm. Macvean (2007) *Abies guatemalensis* Rehder indicates seed length about 8-10 mm and wings about 15 mm long.

Nowadays; because of its increasing economical value in market and decorative characteristic in landscape architecture, Turkish fir (*Abies nordmanniana* subsp. *bornmulleriana* Mattf.) is taking more importance. In addition to this being an endemic species of Turkey and widely preferred Noel tree in the world. Turkish fir is one of the most important trees to Christmas tree and for these there are very much studies on this species (Frampton and McKinley, 1999; Frampton et al. 2009; Talgø et. al., 2009; Newton et al., 2009; Hart et al., 2009; Langdren et al., 2008; Frampton and Işık, 2008; Talgø and Stensvand, 2008).

In this study, the genetic diversity of Turkish fir determined with respect to the some morphological characteristics. In this sense construction of genetic diversity, basic morphological characteristics, geographical variations and the morphological differences between tree species in the optimal and extreme distribution area of Turkish fir was determined.

Until now a few studies have been conducted about Turkish fir (Kaya et al., 2008; Şimşek, 1991; Velioğlu, 1999; Kaya and Raynal, 2000; Nielsen and Chastagner, 2005). But there is no comprehensive study to disclose the spatial distribution of Turkish fir and provide background information for future studies. In the near future; the studies done with the morphological characters about genetical variations are should also be analysed with DNA markers and isosymes analysis.

5. References

Anonymous, 2006. Orman Varlığımız, T.C. Çevre ve Orman Bakanlığı Orman Genel Müdürlüğü, Orman İdaresi ve Planlama Dairesi Başkanlığı, Ankara, 160 s.

Basu, R.N. 1994. Seed viability. In *Seed Quality: Basic Mechanisms and Agricultural Implications* (ed. A.S Basra), pp. 1-44. Food products press, New York.

Bates, S.L., Borden, J.H., Savoie, A. and Blatt, S.E. 2000. Impact of feeding by *Leptoglossus occidentalis* (Hemiptera: Coreidae) on the major storage reserves of mature Douglas-fir (Pinaceae) seeds. *Canadian Entomologist* 132: 91-102.

Bates, S.L., Lait, C.G., Borden, J.H. and Kermode, A.R. 2001. Effect of feeding by the western conifer seed bug, *Leptoglossus occidentalis*, on the major storage reserves of developing seeds and on seedling vigour of Douglas-fir. *Tree Physiology* 21: 481-487.

Bazzaz, F.A., Ackerly, D.D. and Reekie, E.G. 2000. Reproductive allocation in plants. In *Seeds: the Ecology of Regeneration in Plant Communities* (ed. M. Fenner), 2nd ed., pp. 1-29. CABI Publishing, Wallingford, UK and New York, USA.

Bilgen BB, Kaya N (2007). Allozyme Variations in Six Natural Populations of Scots Pine (*Pinus sylvestris* L.) in Turkey. Biologia, Volume: 62: 697-703.

Bilir, N., 2002. Doğu Karadeniz Bölgesinde Kurulan Toros Sediri (*Cedrus libani* A. Rich) Orijin Denemelerinin İlk Sonuçları, Karadeniz Technical University, Graduate School of Natural And Applied Sciences, Doctorate Thesis, Trabzon, 116 p.

Chalmers, K. J., Waugh, R., Sprent J. I., Sımons, A. J., Powell, W. 1992, Detection of Genetic Variation Between and Within Populations of Gliricidia Sepium and G. maculata using RAPD markers, Heredity 69, 465-472

Clark C. M. Wentworth T. R. O'Malley D. M.. 2000. Genetic Discontinuity Revealed by Chloroplast Microsatellites in Eastern North American *Abies* (Pinaceae). American Journal of Botany *87: 774-782*

Conte, L., Cotti, C., Schicchi, R., Raimondo, F,M. and Cristofolini, G., 2004, Detection of ephemeral genetic sub-structure in the narrow endemic *Abies nebrodensis* (Lojac.) Mattei (Pinaceae) using RAPD markers, Plant Biosystems, 138 (3), 279-289.

Cregg, B, M. 1994. Carbon Allocation, Gas Exchange, and Needle Morphology of *Pinus ponderosa* Genotypes Known to Differ in Growth and Survival Under Imposed Drought, Tree Physiology 14, 883-898.

Dajoz, R. 2000. *Insects and Forests: The role and diversity of insects in the forest environment*. Intercept Ltd, London-Paris-New York. 668 pp.

Diebel KE, Feret, PP. 1991. Isozyme variation within the Fraser fir (*Abies fraseri* (Pursh) Poir.) population on Mount Rogers, Virginia: lack of microgeographic differentiation. Silvae Genetica 40: 79-85.

Edwards D.G.W., 1982. Collection, processing, testing and storage of true fir seeds: a review. In: Oliver CD, Kenady RM, eds. Proceedings, Symposium on the Biology and Management of True Fir in the Pacific Northwest; 1981; Seattle/Tacoma, WA. Contrib. 45. Seattle: University of Washington, Institute of Forest Resources: 113B137.

El Atta, H.A. 1993. The effect of *Caryedon serratus* Oliver (Col., Bruchidae) on viability and germination of seeds of *Acacia nilotica* (L. Willd. Ex Del.) in the Sudan. *Forest Ecology and Management* 57: 169-177.

El-Kassaby, Y.A., Chaisurisri, K., Edwards, D.G.W. and Taylor, D.W. 1993. Genetic control of germination parameters of Douglas-fir, Stika spruce, western red cedar and yellow cedar and its impact on container nursery production. In *Dormancy and Barriers to Germination* (ed. D.G.W. Edwards). pp. 37-42. Proc. on an international symposium of IUFRO project Group P2.04-00 (Seed problems), Victoria, British Colombia, Canada, April 23-26, 1991.

El-Kassaby, Y.A., Edwards, D.G.W., Taylor, D.W., 1992, "Genetic Control of Germination Parameters in Douglas-Fir and İts Importance For Domestication", Silvae Genetica, 41 (1): 49-53

Erkan, S., 2008, Kastamonu Günlüburun Karaçam Tohum Bahçesinde Klonal Farklılıklar, Yüksek Lisans Tezi, Gazi University, Graduate School of Natural And Applied Sciences, Msc Thesis, 96 p.

Erkuloğlu, Ö, S., 1993, Kayın, Göknar ve Sedir Tohumlarını Uzun Süre Saklama Olanakları Üzerine Araştırmalar, İç Anadolu Ormancılık Araştırma Enstitüsü Yayınları, Teknik Rapor Serisi No:62, ISSN 1300-7920, Ankara, s.59,

Fady B, Conkle MT. 1993. Allozyme variation and possible phylogenetic implications in *Abies cephalonica* Loudon and some related eastern Mediterranean firs. Silvae Gentica 42: 351-359.

Fan, S. ve Grossnickle, S, C., 1999, Genetic Variation in Response to Drought of Interior Spruce (*Picea glauca* (Moench) Voss X *P. engelmannii* Parry ex Engelm.), Scandinavian Journal of Forest Research, 251 – 261,

Frampton J. and McKinley C., 1999. Christmas T rees and Greenery in Denmark Production and Tree Improvement American Christmas Tree Journal. 1999. Vol. 43(2):4-11.

Frampton, J. and Işık, F.,2008, A Collection of Turkish (*Abies bornmülleriana*) and Trojan (*Abies equi-trojani*) Fir Seeds, Proceedings of the 8th International Christmas Tree Research & Extension Conference, Forest & Landscape Denmark, Hørsholm. Working Papers No. 26-2008, 27-28.

Frampton, J., Işık, F., Benson, M., Braham, A,M., 2009, Variation in resistance to Phytophthora root rot within Turkish and Trojan fir, Proceedings of the 9th International Christmas Tree Research & Extension Conference, (13-18 September 2009), 45-46, USA.

Franklin, Jerry F. 1974. Abies Mill. fir. In: Schopmeyer, C. S., Technical Coordinator. Seeds of Woody Plants in The United States. Agric. Handb. 450. Washington, DC: U.S. Department of Agriculture, Forest Service: 168-183.

Gökmen, H., 1970, *Abies nordmanniana* (Stev.) Spach. Doğu Karadeniz Göknarı (Kafkas Göknarı), Açıktohumlular Gymnospermae, T.C. Orman Bakanlığı OGM Yayınlarından, 523, 49, 75, Ankara.

Goldstein D. B. Ruiz Linares A. Cavalli-Sforza L. L. Feldman M. W.. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139: 463-471

Güney, D., 2009, Doğu kayınında (*Fagus orientalis* lipsky) Bazı Coğrafik Varyasyonların Morfogenetik Olarak Belirlenmesi, Karadeniz Technical University, Graduate School of Natural And Applied Sciences, Doctorate Thesis, 173 s.

Güney, D., Şevik, H., Kulaç, Ş., 2011. "Kızılçam (*Pinus brutia* TEN.) Tohum Meşcerelerinde Tohum Özelliklerine Bağlı Genetik Çeşitlilik", Türkiye IV. Tohumculuk Kongresi, Bildiriler Kitabı-1, 454-458, Ondokuz Mayıs Üniversitesi Ziraat Fakültesi, 14-17 Haziran 2011, Samsun

Gutterman Y 2000. Maternal effects on seeds during development. In *Seeds: the Ecology of Regeneration in Plant Communities* (ed. M. Fenner). 2nd ed. pp 59-84. CABI Publishing, Wallingford New York.

Hampton, J.G. 2002. What is seed quality? *Seed Science and Technology* 30: 1-10.

Hamrick, J. L., 1989: Isozymes and the analysis of genetic structure in plant populations. In SOLTIS, D. E., SOLITS, E S., (Eds): Isozymes in plant biology pp. 87-105. - Portland, Oregon: Dioscorides Press.

Hamrick, J. L., Godt, M. J. W. and Sherman-Broyles, S. L., 1992. Factors influencing levels of genetic diversity in woody plant species. New Forests 6: 95–124

Hart, J., Landgren, C., Fletcher, R., Moody, J.T., Bondi, M., 2009, The influence of needle location and sampling technique on nutrient concentration, Proceedings of the 9th International Christmas Tree Research & Extension Conference, (13-18 September 2009), 94-98, USA.

John, W, J., 1948. Geographic Variation in Central Oregon Ponderosa Pine (*Pinus ponderosa* Laws.) : Seed Germination; Seed, Wing, and Cone Morphology; Seed Color.

Kathleen L, S. ve Furnier, G, R., 2002, Genetic Variation and Population Structure in Central and Isolated Populations of Balsam Fir, *Abies balsamea* (Pinaceae), American Journal of Botany 89, 5, 783–791.

Kaya, Z., Raynal, D., 2000, Biodiversity and conservation of Turkish forests, Biological Conservation 97 (2001) 131-141

Kaya, Z., Skaggs, A., Neale, D,B., 2008, Genetic Differentiation of Abies equi-trojani (Asch. & Sint. ex Boiss) Mattf. Populations from Kazdağı, Turkey and the Genetic Relationship between Turkish Firs belonging to the *Abies nordmanniana* Spach Complex, Turk J. Bot, 32-(2008),1-10

Kolotelo, D., 1998. Abies Seed Problems, Forest Nursery Association of British Columbia Meetings, Proceedings, 1995, 1996, 1997, p.122-130.

Komar, E, T., 2000. Seed Production, Characteristics and Viability of A. Auriculiformis A. Cunn. From Provenance-Based Seedling Seed Orchard in Wonogiri, Central Java, Indonesia, IUFRO Joint Symposium on Tree Seed Technology, Physiology and Tropical Silviculture, 68-114

Konnert M, Bergmann F. 1995. The geographical distribution of genetic variation of silver fir (*Abies alba*, Pinaceae) in relation to its migration history. Plant Systematics and Evolution 196: 1-2, 19-30.

Kulaç, Ş., Çakar, S., Güney, D., Turna, İ., Şevik, H., 2011a. "Sarıçam (*Pinus sylvestris* L.) Tohumlarında Morfolojik Özelliklerin Çimlenme Üzerine Etkileri", Türkiye IV. Tohumculuk Kongresi, Bildiriler Kitabı-1, 465-471, Ondokuz Mayıs Üniversitesi Ziraat Fakültesi, 14-17 Haziran 2011, Samsun

Kulaç, Ş., Şevik, H., Güney, D., 2011b. "Karaçam (*Pinus nigra* Arnold.) Tohum Meşcerelerinde Yükseltiye Bağlı Genetik Çeşitlilik", Türkiye IV. Tohumculuk Kongresi, Bildiriler Kitabı-1, 448-453, Ondokuz Mayıs Üniversitesi Ziraat Fakültesi, 14-17 Haziran 2011, Samsun

Lamhamedi, S, M., Chamberland, H., Bernier, P, Y., ve Tremblay, F, M., 2000. Clonal Variation in Morphology, Growth, Physiology, Anatomy and Ultrastructure of Container-Grown White Spruce Somatic Plants, Tree Physiology 20, 869–880.

Landgren, C., Chastagner, G., Riley, K., 2008, Provenance and progeny growth and postharvest needle retention test results from Nordmann (*Abies nordmanniana*) and Turkish fir (*A. bornmuelleriana*) trials in Oregon and Washington, Proceedings of the 8th International Christmas Tree Research & Extension Conference, Forest & Landscape Denmark, Hørsholm. Working Papers No. 26-2008, 1-5.

Macvean, A,L,E., 2007. *Abies guatemalensis*. Species Description in The Tropical Tree Seed Manual. 241-242

Matziris, D., 1984. Genetic Variation in Morphological and Anatomical Needle Characteristics in The Black Pine of Poleponnesos, Silvae Genetica, 33, 4-5, 164-166.

Matziris, D., 1997. Variation in Growth, Flowering and Cone Production in a Clonal Seed Orchard of Aleppo Pine Grown in Greece, Silvae Genetica, 46, 4, 224-228.

Matziris, D.I., 1989. Variation in Growth and Branching Characters in Black Pine (*Pinus nigra* Arnold.) of Poleponnesos, Silvae Genetica, 38, 3-4, 77.

Messaoud, Y., Bergeron, Y., ve Asselin, H., 2007. Reproductive Potential of Balsam Fir (*Abies balsamea*), White Spruce (*Picea glauca*), and Black Spruce (*Picea mariana*) at The Ecotone Between Mixed Wood and Coniferous Forests in The Boreal Zone of Western Quebec, American Journal of Botany 94(5): 746–754. 1-7

Newton, L., Hain, F., Frampton, F., 2009, Host resistance screening for balsam woolly adelgid: Early results from 13 fir species, Proceedings of the 9th International Christmas Tree Research & Extension Conference, (13-18 September 2009), 81-84, USA.

Nielsen, U, B. and Chastagner, G, A., 2005, Genetic variation in postharvest needle retention among Nordmann fir families and grafted clones, Scandinavian Journal of Forest Research, 2005; 20: 304_/312

Okada, S., Mukaide, H., ve Sakai, A., 1973, Genetic variation in Saghalien fir from different areas of Hokkaido, Silvae Genetica, 22, 1-2

Owens, J.N., Colangeli, A.M. and Morris, S.J. 1990. The effect of self-, cross, and no pollination on ovule, embryo, seed, and cone development in western red cedar (*Thuja plicata*). *Canadian Journal of Forest Research* 20: 66-75.

Parducci L, Szmidt AE. 1997. Genetic variation among nine Mediterranean *Abies* species based on PCR-RFLP analysis of chloroplast DNA. Proceedings, International Conference on Biodiversity and Bioresources: Conservation and Utilization; Phuket, Thailand: 87.

Parker, W, H., Maze, J., Bradfield G, E. 1981 Implications of Morphological and Anatomical Variation in *Abies balsamea* and *Abies lasiocarpa* (Pinaceae) from Western Canada, Botanical Society of America, 68 (6); 843-854 USA

Planter, E,A., Furnier, G,R. and Eguiarte, L,E., 2000, Low levels of genetic variation within and high levels of genetic differentiation among populations of species of *Abies* from southern Mexico and Guatemala, American Journal of Botany 87(3): 362–371.

Pritam, S. and Singh, P. 1997. Forest tree seed pathogens and their management in sustainable forestry. *Journal of Mycology and Plant Pathology* 27: 138- 147.

Sagnard, F., Barberot, C., Fady, B., 2002, Structure of Genetic diversity in *Abies alba* Mill. from southwestern Alps: multivariate analysis of adaptive and non-adaptive traits for conservation in France, Forest Ecology and Management 157 (2002) 175–189.

Scaltsoyiannes, A., Tsaktsira, M. and Drouzas, A,D., 1999, Allozyme differentiation in the Mediterranean firs *(Abies, Pinaceae)*. A first comparative study with phylogenetic implications, P1. Syst. Evol. 216:289-307

Schmidtling, R.C., Myszewski, J.H. ve McDaniel, C.E., 2005. Geographic Variation in Shortleaf Pine (*Pinus echinata* Mill.) - Cortical Monoterpenes, 28th Southern Forest Tree Improvement Conference: June 21-23, 2005. Raleigh, NC: North Carolina State University, Department of Forestry and Environmental Resources, 161-167.

Şevik, H., 2005, Batı Karadeniz Bölgesi Sarıçam (*Pinus sylvestris* L.) Tohum Meşçerelerinde Populasyonlar Arası Farklılıklar, Gazi University, Graduate School of Natural And Applied Sciences, Msc Thesis, 60p.

Şevik, H., 2010, Structure of Genetic Diversity in The Populations of Turkish fir (Abies nordmanniana subsp. bornmülleriana Mattf.), Karadeniz Technical University, Graduate School of Natural And Applied Sciences, Doctorate Thesis, 150p.

Şevik, H., 2011. Dallanma Karakterleri Bakımından Noel Ağacı Üretimine Uygun Uludağ Göknarı Populasyonlarının Belirlenmesi, Kastamonu Üniversitesi, Orman Fakültesi Dergisi, 11 (1): s. 102-107, Kastamonu.

Şevik, H., Ayan, S., Turna, İ., Yahyaoğlu, Z., 2010a. Genetic diversity among populations in Scotch pine (*Pinus silvestris* L.) seed stands of Western Black Sea Region in Turkey, African Journal of Biotechnology Vol. 9(43), pp. 7266-7272

Şevik, H., Güney, D., Kulaç, Ş., 2011. Uludağ Göknarında (*Abies nordmanniana* subsp. *bornmülleriana* Mattf.) Tohum Temininde Karşılaşılan Güçlükler ve Çözüm Önerileri, Türkiye IV. Tohumculuk Kongresi, Bildiriler Kitabı-1, 436-441, Ondokuz Mayıs Üniversitesi Ziraat Fakültesi, 14-17 Haziran 2011, Samsun

Şevik, H., Yahyaoğlu, Z., Turna, İ., 2010b. "Uludağ Göknarı (*Abies nordmanniana* subsp. *bornmulleriana* Mattf.) Tohumlarında Farklı Ekim Zamanı ve Ekim Koşullarının Çimlenme Üzerine Etkisi", III. Ulusal Karadeniz Ormancılık Kongresi, Bildiriler Kitabı, Cilt:II, s, 780-784, 20-22 Mayıs 2010, Artvin

Shea, K. L. 1990. Genetic variation between and within populations of Engelmann spruce and subalpine fir. *Genome* 33: 1–8.

Shea, L,S. and Furnier G,R., 2002, Genetic variation and population structure in central and isolated populations of balsam fir, A*bies balsamea* (Pinaceae), American Journal of Botany 89(5): 783–791.

Şimşek, Y., 1991, Türkiye orijinli goknar türlerinin (*Abies nordmanniana, Abies bornmuellariana, Abies equi-trojani*) genetik yapıları üzerine araştırmalar. OAE Yayınları, Teknik Bulteni 221.

Skryszewska, K. ve Chlanda, J., 2009. A Study on The Variation of Morphological Characteristics of Silver Fir (*Abies alba* Mill.) Seeds and Their Internal Structure Determined by X-ray Radiography in the Beskid Sądecki and Beskid Niski Mountain Ranges of the Carpathians (southern Poland), Journal of Forest Science, *55*, 9, 403–414.

Sorensen, F.C. ve Franklin, J.F., 1977. Influence of Year of Cone Collection on Seed Weight and Cotyledon Number in *Abies procera*, Silvae Genetica, 26, 1, 41-43.

Streiff, R., T. Labbe, R. Bacilieri, H. Steinkellner, J. Glossl, and A. Kremer. 1998 Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assesed with isozymes and microsatellites. *Molecular Ecology* 7: 317–328

Suyama Y, Tsumura Y, Ohba K. 1992. Inheritance of isozyme variants and allozyme diversity of *Abies mariesii* in three isolated natural populations. Journal of the Japanese Forestry Society 74: 65-73.

Talgø, V. and Stensvand A., 2008, *Delphinella abietis* and *Herpotrichia parasitica* cause needle damage in Norwegian Christmas tree plantations, Proceedings of the 8th International Christmas Tree Research & Extension Conference, Forest & Landscape Denmark, Hørsholm. Working Papers No. 26-2008, 49-54.

Talgø, V., Chastagner, G,A., Thomsen, I,M., Cech, T., Riley, K., Lange, K., Klemsdal, S,S., and Stensvand A., 2009, *Sydowia polyspora* isolated from needles and seeds of true fir is associated with current-season needle necrosis (CSNN), Proceedings of the 9th International Christmas Tree Research & Extension Conference, (13-18 September 2009), 52-57, USA.

Turna, İ,. 2004. Variation of Morphological Characters of Oriental spruce (*Picea orientalis*) in Turkey, Biologia, Bratislava, 59, 4, 519-526.

Turna, İ. ve Güney, D., 2009. Altitudinal Variation of Some Morphological Characters of Scots Pine (*Pinus sylvestris* L.) in Turkey, African Journal of Biotechnology, 8, 2, 202-208.

Turna, İ., 2003, Variation of Morphological and Electrophoretic Characters of 11 Populations of Scots Pine in Turkey, Israel Journal of Plant Sciences, 51.

Turna, İ., Şevik, H. ve Yahyaoğlu, Z., 2009, Uludağ Göknarı (*Abies nordmanniana* subsp. *bornmülleriana* mattf.) Populasyonlarında Morfolojik Özelliklere Bağlı Genetik Çeşitlilik, Bartın Orman Fakültesi Dergisi I.Ulusal Batı Karadeniz Ormancılık Kongresi Bildiriler Kitabı, Özel Sayı, ISSN: 1302-0943, II, 341-347

Turna,I., Z. Yahyaoglu, F. Yüksek, F. A. Ayaz ve D. Guney., 2006. Morphometric and electrophoretic analysis of 13 populations of Anatolian black pine in Turkey, J. of Enviromental Biology, Vol. 27 (3), 491-497.

Tylek, P. ve Walczyk, J., 2002. The Relationship Between The Viability and Geometric Characteristics of Beechnuts *Fagus sylvatica* L. El. J. of Polish Agr. Universities, Forestry, 5, 2.

Velioğlu, E., Çiçek, F. F., Kaya, Z. ve Çengel, B., 1999b. Kaz Dağlarındaki Doğal Kazdağı Göknarı (*Abies equi-trojani* Aschers. Et. Sint.) Populasyonlarında Genetik Çeşitliliğin Yapılanması, Orman Ağaçları ve Tohumları Islah Araştırma Müdürlüğü, 3, 74, 10, 31, Ankara.

Vendramin, G. G., Degen, B. Petit, J. R. Anzidei, M. Madaghiele, A. and Ziegenhagen, B.1999. High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Biology* 8: 1117–1126.

Vicario, F., Vendramin, G.G., Rossi, P., Liò, P. and Giannini, R., 1995, Allozyme, Chloroplast DNA and RAPD Markers for Determining Relationships between Abies alba and the Relic Population of Abies nebrodensis, Theor. Appl. Genet., vol. 90, pp. 1012–1018.

Yahyaoğlu, Z., Demirci, A., Bilir, N. ve Genç, M., 2001. Comparison of 22 Taurus Cedar (*Cedrus libani* A. Rich.) Origins by Seedling Morphological Distance, Turkish Journal of Biology, 25, Ankara, 221, 224.

13

Genetic Diversity in Apricot

Kadir Ugurtan Yilmaz* and Kahraman Gurcan
Erciyes University, Agriculture of Faculty
Department of Horticulture, Kayseri
Erciyes University, Agriculture of Faculty
Department of Agricultural Biotechnology, Kayseri
Turkey

1. Introduction

Apricot is a fruit species adopted to a wide geographical areas (De Poerderlé, 1788; Loudon, 1838; Arakelyan, 1968; Mehlenbacher et al. 1991; Huxley, 1992; Butner, 2001). 173 years ago, Loudon (1838) was first to mention that wild apricots with different shades of pink flowers had been used as ornamental purpose for centuries. Nowadays commercial production areas of apricots are still very limited with a small number of varieties, although they spread across a wide area all over the world. Looking at the statistics, the production value has been observed to show upward trend by years. This increase in production is closely related with breeding studies in different countries. Breeding programs were modified generally according to consumer's demands and also some subjects such as resistance to diseases (Sharka, Monilinia etc.) and frost damages, determination of self-(in)compatibility.

Germplasm collection and characterization is an early essential stage to initiate a breeding program for diversity. Traditionally germplasm collection and characterization had been done discribing phenological, pomological and morphological characteristics such as tree vigor and growth habit (Perez-Gonzales, 1992; Badanes et al. 1998; Asma & Ozturk, 2005), fruit quality features (Rehder, 1940; Bailey & Hough, 1975; Audergon et al. 1990; Souty et al. 1990; Crossa-Raynaud & Audergon, 1991; Parolari et al. 1992; Bassi & Bartolozzi, 1993; Badanes et al. 1998; Gurrieri et al. 2001; Ledbetter & Petterson, 2004; Asma & Ozturk, 2005; Ruiz & Egea, 2008; Milosevic et al. 2010), leaf (Bailey, 1916; Hou, 1983; Rostova & Sokolova, 1992), stone (Felföldi et al. 2009; Malik et al. 2010), flower (Rodrigo et al. 2006; Yilmaz & Paydas-Kargi, 2010), stigma and stylus (Viti et al. 2000) and pollen (Dezhong et al. 1995; Davarynejad et al. 2005; Arzani et al. 2005; Asma, 2008) comparing and combining the results of characaterization researches published by different groups is a difficult tusk since different variety of morphological, phenological and pomological characteristics have been assessed by the research groups. International UPOV and IPGRI criteria was created in order to overcome this unrequred situation and to enable researchers use common descriptor characteristics.

* Corresponding Author

In the last two decades, molecular studies have been integrated into the conventional germplasm characterization researches (Battistini & Sansavini, 1991; Badanes et al. 1996; Mariniello et al. 2002; Hurtado et al. 2001, 2002; Hormaza, 2002; Vilanova et al. 2003; Geuna et al. 2003; Zhebentyayeva & Sivolap, 2000; Zhebentyayeva et al. 2003; Sanchez-Perez et al. 2005; Romero et al. 2003, 2006; Rao et al. 2008; Yilmaz, 2008; Akpinar et al. 2010) and genetic diversity in apricot. Recent studies show that this genetic diversity originated in Central Asia and transfered to Middle Asia and Caucasia. Later on, the apricot was taken to Europe, and recently spreaded from Europe to North America and the rest of world.

2. Systematic and eco-geographical groups of apricot

Apricot belongs to Prunus genus. Some systematicians created different sections under Prunus genus and Prunuphora sub-genara or Rosaceae family and Prunoideae sub-family (Table 1). American apricots are seen Armeniaca sub-section and named *Armeniaca vulgaris* Lam. (Bailey & Hough, 1975).

Bailey (1916) (Ledbetter, 2008)		**Rehder (1940)** (Ledbetter, 2008)	
Genus	*Prunus*	Genus	*Prunus*
Sub-genera	*Prunophora* (plums, prunes & apricot)	Sub-genera	*Prunophora*
Prunus armeniaca L.		Sections	Euprunus (European/Asian Plums)
	Var. *pendulata* Dipp.		Pronocerasus (North American plums)
	Var. *variegata* Hort.		Armeniaca (Apricots)
	Var. *sibirica* Koch	Armeniaca (Apricots)	
	Var. *mandshurica* Maxim.		*P. brigantina* Vill.
	Var. *Ansu* Maxim.		*P. mandshurica* Maxim.
P. mume Sieb. & Zucc.			*P. sibirica* L.
	Var. *Goethartiana* Koehne.		*P. armeniaca* L.
	Var. *albo-plena* Hort.		*P. mume* Sieb. & Zucc.
Other wild forms			*P. dasycarpa* Ehrh.
	laciniata Maxim.	*P. armeniaca* L.	
	microcarpa Makino		*P. armeniaca variegata* Schneid.
	viridicalyx Makino		*P. armeniaca pendula* Jaeg.
	cryptopetala Makino		*P. armeniaca Ansu* Maxim.
P. brigantiaca Vill.		*P. mume* Sieb. & Zucc.	
P. dasycarpa Ehrh.			*P. mume alba* Rehd.
			P. mume Alphandii Rehd.

Lingdi & Bartholomew (2003) (Ledbetter, 2008)	
Family	*Rosaceae*
Subfamily	*Prunoideae*
Genus	*Armeniaca* (Apricots)
Armeniaca vulgaris L.	
Var. *vulgaris* L.	
Var. *zhidanensis* Qiao & Zhu	
Var. *ansu* Maxim	
Var. *meixianensis* Zhang	
Var. *xiongyueensis* Li	
Armeniaca limeixing Zhang & Wang	
Armeniaca sibirica L.	
Var. *sibirica* L.	
Var. *pubescens* Kostina	
Var. *multipetala* Liu & Zhang	
Var. *pleniflora* Zhang	
Armeniaca holosericea Batal.	
Armeniaca hongpingensis Li	
Armeniaca zhengheensis Zhang & Lu	
Armeniaca hypotrichodes Cardot	
Armeniaca dasycarpa Ehrh.	
Armeniaca mandshurica Maxim.	
Var. *mandshurica* Maxim.	
Var. *glabra* Nakai	
Armeniaca mume Sieb. & Zucc.	
Var. *mume* Sieb.	
Var. *pallescens* Franc.	
Var. *cernua* Franc.	
Var. *pubicaulina* Qiao & Shen	

	P. mume albo-plena Bailey
	P. mume Pendula Sieb
	P. mume tonsa Rehd.

Bortiri et al. (2002) (Ledbetter, 2008)	
Sections	*Penarmeniaca*
Penarmeniaca	
	Prunus fremontii S. Wats.
	Prunus andersonii A. Gray.

Hayashi et al. (2008)		
Traditional Japanese classification		
(Mega et al., 1988; Horiuchi et al., 1996)		
Prunus mume		
For fruit production		
For ornamental "kei"		
	Yabai	
		Yabai sub-groups
		Naniwa sub-groups
		Benifude sub-groups
		Aojiku sub-groups
	Koubai	
		Koubai sub-groups
	Bungo	
		Bungo sub-groups
		Anzu sub-groups

Table 1. Apricot systematics according to various researchers

Leaf characteristics were accepted the most important criterion for apricot systematic by various researchers. Bailey (1916) examined the leaves emerged from dormant buds. Rehder (1940) used leaf shape and pubescence for classification of species. Chinese botanists utilized leaves in parallel Western researchers. However, Japanese researchers considered flower and branch color and flower size in Japanese apricots (*Prunus mume*) which is an ornamental plants for Traditional Japanese classification (Mega et al. 1988; Horiuchi et al. 1996). *Prunus fremontii* which is in Penarmeniaca section and reported in some studies (Bortiri et al. 2001, 2002) along with the desert dwelling species *Prunus andersonii* A. Gray. *Prunus fremontii* can be hybridized freely with other apricot species and differs from them (Ledbetter, 2008).

a. *Prunus armeniaca* var. *ansu* (www.flickr.com)
b. *Prunus mume* (http://commons.wikimedia.org/wiki/File:Prunus_mume_Yaekanko.jpg)
c. *Prunus brigantina* (http://luirig.altervista.org/cpm/albums/bot-042/001-prunus-brigantina.jpg)
d. *Prunus armeniaca*
e. *Prunus mandshurica* (http://www.lawyernursery.com)
f. *Prunus sibirica* (http://www.agroatlas.ru/en/content/related/Armeniaca_sibirica/)

Fig. 1. Different apricot species

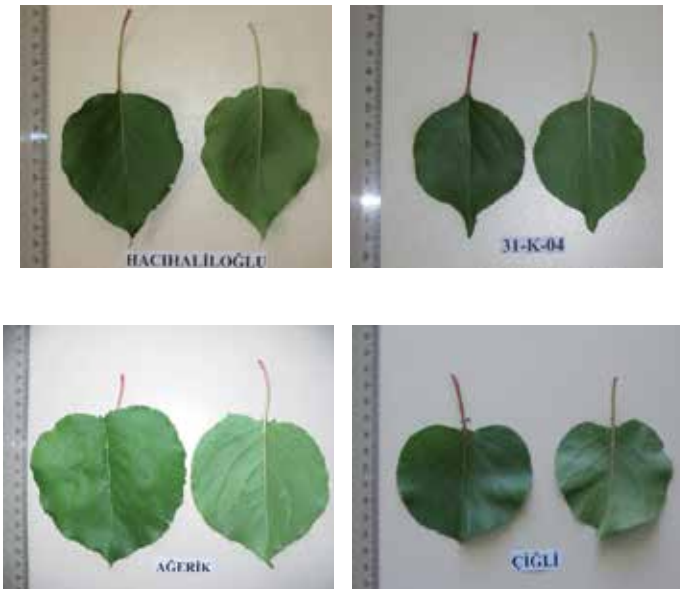


Fig. 2. Leaf shapes in *Prunus armeniaca* in Irano-Caucasian eco-geographical group (Turkish apricots - Yilmaz, 2008)

Apricot has 2n=16 chromosomes. Wide variations emerged in apricots because of seed propagation and growing in different ecological areas in time. Therefore, the systematicians reported that 6-8 eco-geographical groups and 13 regional sub groups occured.

a. Central Asian eco-geographical group (Kostina, 1969)
- Fergana regional sub-group (Bailey & Hough, 1975)
- Upper Zeravshan regional sub-group (Bailey & Hough, 1975)
- Semerkand-Shahrisiabz regional sub-group (Bailey & Hough, 1975)
- Horezm regional sub-group (Bailey & Hough, 1975)
- Kopet-Dagh regional sub-group (Bailey & Hough, 1975)

b. Irano-Caucasian eco-geographical group (Kostina, 1969)
- Iran-Caucasus regional sub-group (Mehlenbacher et al. 1991)
- Dagestan regional sub-group (Mehlenbacher et al. 1991)
- North Africa regional sub-group (Mehlenbacher et al. 1991)

c. European eco-geographical group (Kostina, 1969)
- Western European regional sub-group (Layne et al. 1996)
- Eastern European regional sub-group (Layne et al. 1996)
- Northern European regional sub-group (Layne et al. 1996)

d. Dzhungar-Zailig eco-geographical group (Kostina, 1969)
- Dzhungar regional sub-group (Mehlenbacher et al. 1991)
- Zailig regional sub-group (Mehlenbacher et al. 1991)

e. Northern China eco-geographical group (Bailey & Hough, 1975)

f. Eastern China eco-geographical group (Bailey & Hough, 1975)

g. Tibet eco-geographical group (Bailey & Hough, 1975; Asma, 2011)

h. Northeast China eco-geographical group (Bailey & Hough, 1975; Asma, 2011)

The oldest and richest in diversity is the Central Asian group which includes local apricots from Central Asia, Xinjiang, Afghanistan, Balucistan, Pakistan, and Northern India (Kashmir). This group is mostly self-incompatible and characterized with medium sized fruits and they have a tendency to bloom late spring. The secondary gene center of apricot is the Irano – Caucasian group which extends from Armenia, Georgia, Azerbaijan, Dagestan, Iran, Iraq, Syria, Turkey, to North Africa, and even to Spain and Italy. They are generally self-incompatible, but on contrary, they produce large fruits and blooms earlier than apricots of Central Asia and needs lower chilling hours. Apricots of North American, South African, and Australian are classified as the European group and this group was originated from the apricots of Armenia, Iran, Turkey, and other Arab countries. Apricots of this group are self-compatible, fruits are more precocious and the trees need low chilling. The Dzhungar-Zailig group with mostly small fruits includes selections from regions of Dzharskent, Taldy-Kurgan, Kazakhstan, and Xinjiang (Mehlenbacher et al. 1990; Layne et al. 1996; Faust et al. 1998). Later, two major groups proposed by Bailey & Hough (1975), the Northern China group that includes forms of *Prunus mandshurica* and *Prunus sibirica*, and the Eastern China group that includes forms of *Prunus ansu* (Romero et al. 2003). In addition, some researchers mentioned two more groups named Tibet and Northeast China. While Tibet eco-geographic group includes forms of *Prunus armeniaca* var. *holosericea*, Northeast China eco-geographic group includes varieties and types of *Prunus armeniaca*, *Prunus sibirica* and *Prunus mandshurica* (Bailey & Hough, 1975).

The Central Asian and Irano-Caucausian including Turkish and Iran cultivars eco-geographical groups show the richest phenotypic variability, while European group including cultivars grown in North America, Australia and South Africa is to exibit the least diversity (Mehlenbacher et al. 1991; Halasz et al. 2010).

3. Origin and spread of apricot to the world

According to the famous Russian Botanist Vavilov (1951), there are three important regions as origin of apricots although Armenia had been supposed apricot's origin and named as *Prunus armeniaca*, previously. These are;

a. The Chinese center (China and Tibet)
b. The Central Asian center (from Tien-Shan to Kashmir)
c. The Near-Eastern center (Iran, Caucasus, Turkey)

Also Vavilov (1951) reported that the Near-Eastern center could be secondary gene center because of cultured varieties and absence of wild apricot forms (Bailey & Hough, 1975; Asma, 2011).

The spread of apricots from Central Asia to the rest of world are explained by three different views. The first of these, dried apricot fruits and stones of natural apricot flora in Fergana Valley which is at the border of Uzbekistan, Tajikistan, and Kyrgyzstan and piedmont of Hind Kush and Tian Shan mountains were brought to Anatolia by soldiers on Iran and Transcaucasus during organized Asia campaigns by the Great Alexander in BC 334. Later on, apricot was moved to Europe from Anatolia during the Roman-Persian wars in BC 1 (Layne et al. 1996). The second view; apricot was brought to Anatolia by merchants from China and Center Asia on famous Silk Road, and then Roman soldiers carried apricot to

Italy from Anatolia (Bailey & Hough, 1975). The third view is that Romans removed apricot to west during their expeditions to seize the Near East (Syria, Iran, and Caucasus) in BC 2 (Layne et al. 1996). Apricot gradually spreaded to Africa on Meditarranean countries and Middle East, it was also carried to Balkans by Ottomans in XV. and XVI. century (Suranyi, 1999). Apricot was taken to Southern Europe from Eastern Europe countries (Asma, 2011), to England in 1524 or 1548 from Italy. It was removed to America continent by the Spanish in 1626 (Faust et al. 1998).

4. Apricot production of the world

Despite we observe rich distribution of apricot through the world, the commercial production areas are limited. The majority of production is done in Mediterranean countries and also in Iran, Pakistan, Uzbekistan, Morocco, Algeria, Ukraine and USA (Romero et al. 2003).

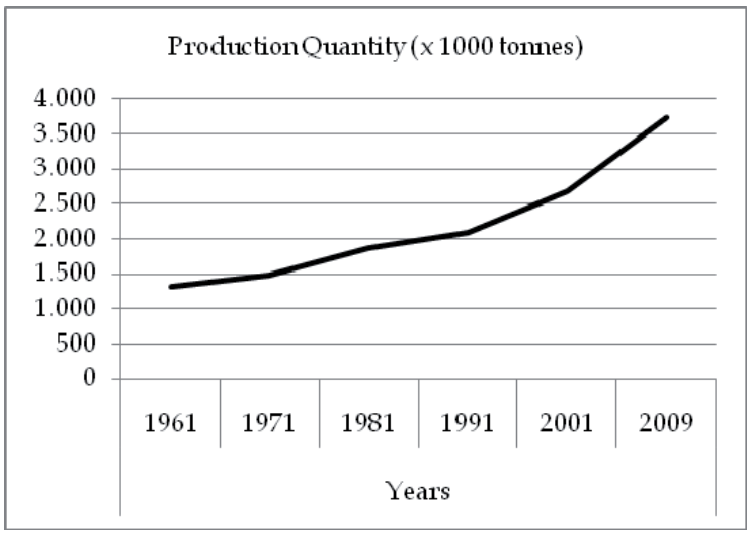


Fig. 3. World apricot production from 1961 to 2009 (FAOSTAT, 2011)

Analyzing the data by years from 1961 until 2009, we observe that the amount of production in the world have increased on a regular basis. During the last 50 years 1.317.607 tons per year production reached about 3.728.083 tons (Fig 3). Despite to this increase in the 50 years amount of production is actually low due to the limited capabilities of adaptation to different environments, limited numbers of variety, self-incompatibility, frost damage, susceptibility to Sharka and Monilinia (*Sclerotinia laxa* Aderh et., Ruhl.).

Turkey is the leading country at the production of apricots. Turkey with its rich genetic resources and high quality dried apricot cultivars has reached to monopolistic position in the world. In recent years, the farming of dried apricot has increased in China, Pakistan, Iran, Syria, Uzbekistan, Afghanistan, too. Especially the city of Malatya in Turkey (Fig 4) and the surrounding areas of the city provides more than half of production of the country. For this reason, the city, Malatya is called also the apricot capital of the world (Asma, 2007). Apricots have been traditionally a part of life and as a symbol of the city at every point

Fig. 4. Malatya province where is the most apricot produce in the world.

Fig. 5. Apricot is a symbol in Malatya.

(Fig 5). Malatya Apricot Research Station established in the city in 1937 played an important role at the development of apricot farming in the city (Fig 6). The station has a rich collection of 285 apricot accessions of different eco-geographical groups. It is third behind the Nikitsky Botanic Garden in Ukraine and Central Asian Experimental Station of the Institute of Plant Industry in Uzbekistan (Fig 7). The institution host thousands of seedlings obtained from constantly maintained crosses. In 2010, Malatya produced 661.000 tons of fresh apricot. Of the that production 101.000 tons dried apricots were obtained (TurkStat, 2010) (Fig 8).

Fig. 6. Apricot Research Station (1937)

Fig. 7. A view of Apricot Genetic Resources Collection Orchard in Apricot Research Station in Malatya / Turkey

Fig. 8. Best apricot competition in Malatya and new apricot cultivar Alkaya for table and dried consumption

Apricot cultivation in the Mediterranean countries generally base on early and middle season table apricots and Spain, Italy and France have authority in the trade of table apricots. USA, South Africa and Australia are producer of dried and flesh apricots. However production of USA decreases sharply.

Apricot rich countries has plenty of genetic diversity due to the production many years with seed. This increase shows that the studies need to be done in apricot. Indeed, in recent years, big, flashy, red-checked with orange flesh and resistant apricot variety has been the target of new development. Especially in Europe and the USA breeding programs released big new cultivars with cheek color of red but because of the low brix and poor flavor the release failed to succeed in the marked.

One of the main priority of apricot breeding programs in the mid of 1990s was to develop cultivars resistant to the late spring frost (Layne et al. 1996; Bassi & Sansavini, 1988). However, unlike almond breeding (Vargas & Romero, 2001), this aim was failed to succeed

(Demirtas et al. 2010). In the wild apricots and germplasm collections the lack of genotypes exhibiting resistance to the late spring frost was the main reason for this fail. But this failure was ignored by producer since apricot is high-profit production and good evaluation of adaptation studies (Occarso, 1977; Durie, 1988; Ogasanovic et al. 1991; Harsanyi, 1991; Ozvardar et al. 1991; Baktir et al. 1992; Bassi et al. 1995; Kaska et al. 1995; Egea et al. 1995; Audergon et al. 1995; Ayanoglu et al. 1995; Paydas et al. 1995; Draganescu & Cociu, 1997; Hofstee et al. 1997; Papanikolaou-Paulopoulo & Poulis, 1997; Yilmaz, 2002; Blanc et al. 2006).

One of the main goals breeding programs was to obtain varieties resistant to Monilinia [*Sclerotinia* (*Monilinia*) *laxa* Aderh et., Ruhl) (Cociu et al. 1991; Gulcan et al. 1994; Bassi et al. 1995; Bassi & Audergon, 2006; Guerriero et al. 2006; Nicotra et al. 2006; Acarsoy et al. 2011). Although this ended up with some success, economically important cultivars were not released into market. Using efficient fungusit prevented also Monilinia breeding programs.

All of these germplasm used in breeding and molecular genetic studies in recent years drawn towards different targets. The most important of these is Sharka (*Plum pox virus*) disease, unfortunately, which caused great losses (Lopez-Moya et al. 2000; Cambra et al. 2006).

In addition, problems related to self-incompatibility became main problem after Sharka, in terms of ensuring the efficiency of production. In this context, the presence of *S* alleles on behalf of researchers to understand the mechanism of conflict directed to this point.

5. Sharka (*Plum pox virus*) resistance in genetic diversity of apricot

Sharka or Plum pox virus (PPV) is the most serious disease of Prunus trees. The disease has spreaded throughout many European countries, especially in Mediterranean countries. Resistant cultivars are limited and only some North American cultivars are known to be resistant to the disease. 'Stark Early Orange' (SEO), 'Goldrich', 'Harlayne', 'Stella', and 'Harcot' are the resistant cultivars mostly used as resistant source for breedings (Martinez-Gomez et al. 2000). The resent focus on the disease increased also cruiosty on the source of resistance of American cultivars. It is believed that North American cultivars originated from a limited number of European cultivars. The source of the resistance is unknown. However, recent studies shows that Central Asian apricots is the most likely source of resistant genes in the North American donors. *Prunus mandshurica* was first to be offered as PPV resistance into North American germplasm (Badenes et al. 1996). Zhebentyayeva et al. (2003) showed that 'Harlayne' and 'Goldrich' clusters with native Central Asian cultivars. 'Stark Early Orange', LE 2904, LE 3276, and 'Vestar' are also grouped with native Chinese material on the genetic diversity study (Zhebentyayeva et al. 2003). Hormaza (2002) also demonstrated that Chinese cultivars contributed to the pedigree of 'Stark Early Orange'. The recent research of Zhebentyayeva et al. (2008) shows that cultivars 'Harlayne', 'Goldrich', and 'Stark Early Orange' has genetic similarity with native Central Asian genotypes. The researchers also showed that *Prunus davidiana* alleles in 'Stark Early Orange' and *Prunus mume* alleles in 'Stark Early Orange' and 'Goldrich' pointed out a contribution of these species to PPV resistance as well (Zhebentyayeva et al. 2008).

6. *S*-genotyping in genetic diversity of apricot

Like to other Prunus species, apricots show gametophytic self-incompatibility controlled by a single locus with multiple genes, *S*-haplotypes (De Nettancourt, 2001). The *S*-haplotype

contains a female determinant, (*S-RNase*) encoding for a ribonuclease enzyme (McClure et al. 1989), and the recently identified male determinant, *S-haplotype-specific F-box* gene (Entani et al. 2003; Romero et al. 2004; Halasz et al. 2010).

The Irano – Caucasian group are usually self-incompatible whereas Europan apricots are mostly self-compatible (Halasz et al. 2005; Kostina, 1970). Mehlenbacher et al. (1991) stated Central Asian apricots are also mostly self-incompatible.

Cross-incompatibility between a pair of cultivars occurs frequently self-incompatible species. Cross-incompatibility was observed among the North American cultivars, Goldrich, Hargrand and Lambertin No.1 (Egea & Burgos, 1996), and also among giant-fruited Hungarian apricots (Szabo & Nyeki, 1991; Halasz et al. 2010). Halasz et al. (2010) determined total 12 cross-incompatiblity groups between Irano-Caucasian eco-geographical groups (Turkish apricots) and European eco-geographical groups (Hungarian and North American apricots) (Table 2).

Cross-incompatibility groups	Cultivars	*S*-genotype
I	Goldrich, Hargrand, Lambertin No.1	S_1S_2
II	Cologlu, Kadioglu, Seftalioglu, Cegledi orias, Ligeti orias	S_8S_9
II	Iri Bitirgen, Moniqui	S_2S_6
IV	Artvin PA, Priana	S_2S_7
V	Alyanak, Ziraat Okulu	S_2S_8
VI	Dortyol-4, Sebbiyiki	S_2S_{19}
VII	Sakit-3, Tokaloglu Izmir	S_3S_{19}
VIII	Cataloglu, Ozal, Soganci	S_6S_9
IX	Zerdali No.1, XI Zerdali	S_6S_{12}
X	Ordubat, X2 Zerdali	S_7S_{12}
XI	Adilcevaz-5, Hacihaliloglu, Kabaasi, Kamelya, Zerdali No.2	S_9S_{13}
XII	Shalakh (Aprikoz), Voski	$S_{11}S_{13}$
XIII	Levent, Sakit-1	S_6S_{19}
XIV	Cekirge 52, X3 Zerdali	S_9S_{20}
0: Universal pollen donors	Canakkale (S_cS_c), Ethembey (S_cS_8), Karacabey (S_cS_2), Mektep (S_cS_8), Pasa Mismisi (S_cS_8), Sam (S_cS_2), Yerli Izmir (S_cS_7)	

Table 2. Cross-incompatibility groups of apricot (Halasz et al. 2010; Egea & Burgos, 1996; Halasz et al. 2005)

To date, 21 *S-RNase* alleles are known in European apricots, 20 of which (S_1-S_{20}) code for self-incompatibility and one (S_c) allowing self-compatibility (Burgos et al. 1998; Halasz, 2007; Halasz et al. 2005, 2007) and recently it was confirmed that S_C haplotype is a pollen part mutant of S_8 haplotype (Halasz et al. 2007). Beside, some additional *S*-alleles have been also identified in Chinese cultivars (Wu et al. 2009; Halasz et al. 2010).

A gradually decreasing allele number was detected in apricot landraces from China to Western Europe, with some allelic exclusivity occuring in certain geographic areas (Halasz, 2007; Halasz et al. 2010).

7. Genetic diversity of apricot based on molecular markers

Apricot is a temperate and subtropical zones fruit. China, the Irano-Caucasian region (Turkey and Iran), Central Asia, Europe and North America are the main producer regions in the world. The Central Asia is the oldest and the primary genetic source of apricot group is the Central Asian accessions are self-incompatible; the Irano-Caucasian apricots which are mostly the cultivated ones are mostly self-incompatible, with large fruits and low chilling requirements. The European and the North American apricots are originated from Irano-Caucasia has relatively narrow genetic diversity and are self-compatible with large fruits (Mehlenbacher et al. 1991). For a long period, genetic diversity in apricot was studied with pomological, morphological and phenological characteristics (Guerriero & Watkins, 1984). DNA-based markers that have been used in the last decade clarify the relationship among the apricot accessions.

For breeding and commercialization of promising apricot cultivars, a precise characterization and discrimination of the cultivars are prerequest. Different types of marker such as morphological, molecular, biochemical systems have been used for genetic analysis in horticultural plants. However, due to the effects of environmental factors, asessment of morphological and pomological traits may be ambiguous. Therefore, markers independent from the environment are necessary for reliable identification and discrimination of genotypes and cultivars. DNA markers are well known independent from environmental interactions and they show high level of polymorphism. Therefore, they are considered invaluable tools for determining genetic relationships/diversity. Various types of DNA markers are now available. Among them, RAPD developed by Williams et al. (1990) has been commonly used method in apricot to assess genetic variability and relationships among cultivars (Takeda et al. 1998; Zhebentyayeva et al. 2000; Hormaza, 2001; Mariniello et al. 2002; Ercisli et al. 2009). More recently, ISSR (Chenjing et al. 2005; Yilmaz, 2008), RFLP (De Vicente et al. 1998), AFLP (Hurtado et al. 2001, 2002; Hagen et al. 2002; Panaud et al. 2002; Geuna et al. 2003; Krichen et al. 2006; Yuan et al. 2007), SSR (Hormaza, 2002; Romero et al. 2003; Zhebentyayeva et al. 2003; Maghuly et al. 2005; He et al. 2006; Maghuly et al. 2006; Ali Khan et al. 2008) and SRAP (Uzun et al. 2010) techniques has also been used in apricot to characterize different cultivars belongs to diverse eco-geographical groups.

The diversity determined between apricot cultivars was probably due to crosses between wild and cultivated apricots and cultivars from different eco-geographic origin (Uzun et al. 2010). Microsatellite analyses suggested that European cultivars might have originated through hybridization among Irano-Caucasian genotypes and also most of the European cultivars have originated by hybridization with genotypes from the Irano-Caucasian group (Maghuly et al. 2005; Faust et al. 1998; Kostina, 1969). The heterozygosity of the apricot genotypes narrowed while apricot transfer from China to Europe. Pedryc et al. (2009) show that Middle European and Chinese apricot are distantly related.

Molecular markes have created new era in genetic diversity researches since early nineties. Restriction fragment length polymorphism (RFLP), and PCR based markers such as randomly amplified polymorphic DNA (RAPDs), sequence-related amplified polymorphism (SRAP), single nucleotide polymorphism (SNPs), micro-satellites or simple sequence repeats (SSRs) are mostly used marker systems in plants and also in apricot genetic divesity researches. Microsattelites among all is a very useful tool for apricot

diversity studies, and most promising to cleary genetic relation among the apricots and travel routes of apricots (Romero et al. 2003; Maghuly et al. 2005).

Amplified fragment length polymorphism (AFLP) molecular markers assesment for the genotyping of 118 commercial apricot accessions and some related apricot species (Geuna et al. 2003). The researchers clustered the apricots into four groups corresponding to their geographic distribution; (1) Mediterranean apricots, (2) Chinese apricots, (3) apricots of continental Europe and (4) Europe-North American apricots. Their data confirmed that the migration of apricot from the East to West. They also showed with molecular markers that *Prunus sibirica* and *Prunus mandshurica* are different from *Prunus armeniaca*, but they group together with Chinese accessions (Geuna et al. 2003). In anohter study Romero et al. (2003) studied apricots by using of SSR markers to determine the genetic relationships among genotypes from different eco-geographical groups. They observed that Western European and North American subgroups clustered together in aggrement with their common origins from ancient European cultivars (Kostina, 1969; Bailey & Hough, 1975; Badanes et al. 1996). However their study placed Hungarian cultivars closer to the Central Asian group than to the other European cultivars.

Hayashi et al. (2008) studied Japanese apricot (*Prunus mume*) germplasm and reported that the genetic diversity and relationships among 127 Japanese apricot germplasms assessed by SSR markers. Their study supported the two hypotheses that Japanese apricot cultivated in Japan had been introduced from China and that fruiting cultivars had been selected from flower-ornamentals.

Turkish germplasm was studied by Yilmaz (2008) and Uzun et al. (2010) and genetic diversity and relationships among the accesessions were determined using RAPD, ISSR, SRAP and SSR markers. The researchers reported the high genetic diversity in Turkish apricots. Four high chilling requiring cultivars originated from Eastern Turkey clustered apart from the rest. European, South African, North American and other Turkish cultivars were not clearly grouped regarding to their geographic districts. Therefore the researchers suggested that these cultivars, despite their different geographic origins, have similar genetic background.

8. References

Acarsoy, N.; Evrenosoglu, Y.; Misirli, A.; Boztepe, O.; Yilmaz, K.U.; Gokalp, K.; Turkoglu, A. & Kokargul, R. (2011). Monilya'ya dayanıklı yeni çeşitlerin ıslahında farklı kombinasyonlarda meyve tutumu, tohum sayısı ve çimlenme oranı. Uluslar arası Katılımlı I. Ali Numan Kıraç Tarım Kongresi Kitabı, Vol. 3, 2537-2543. (in Turkish)

Akpinar, A.E.; Kocal, H.; Ergul, A.; Kazan, K.; Selli, M.E.; Bakir, M.; Aslantas, S.; Kaymak, S. & Saribas R. (2010). SSR-based molecular analysis of economically important Turkish apricot cultivars. *Genetics and Molecular Research* 9, 324-332.

Ali Khan, M.; Maghuly, F.; Borroto-Fernandez, E.G.; Pedryc, A.; Katinger, H. & Laimer, M. (2008). Genetic diversity and population structure of apricot (*Prunus armeniaca* L.) from Northern Pakistan using Simple Sequence Repeats. *Silvae Genetica* 57(3), 157-164.

Arakelyan, B. (1968). Excavations at Garni, 1949–50 in Contributions to the Archaeology of Armenia, (Henry Field, ed.), Cambridge, 1968, page 29.

Arzani, K.; Nejatian, M.A. & Karimzadeh, G. (2005). Apricot (*Prunus armeniaca*) pollen morphological characterisation through scanning electron microscopy, using multivariate analysis. *New Zealand Journal of Crop and Horticultural Science* 33, 381-388.

Asma, B. M. & Ozturk, K. (2005). Analysis of morphological, pomological and yield characteristics of some apricot germplasm in Turkey. *Genetic Resources and Crop Evolution* 52, 305-313.

Asma, B.M. (2007). Malatya: The World's Capital of Apricot Culture", Chronica Hort. 47: 20–24.

Asma, B.M. (2008). Determination of pollen viability, germination ratios and morphology of eight apricot genotypes. *African Journal of Biotechnology* 7, 4296-4273.

Asma, B.M. (2011). Her yönüyle kayısı. Uyum Ajans, Ankara – Türkiye, p. 367. (in Turkish)

Audergon, J.M.; Souty, M. & Breuils, L. (1990). Ame´lioration ge´ne´tique pour l'obtention d'abricots de qualite´. In 9_Colloque sur les recherches fruitie`res. Avignon 4-5-6 de´cembre 1990. Ctifl-INRA, pp 217–228.

Audergon, J.M.; Duffillol, J.M.; Gilles, F. & Signoret, V. (1993). Apricot selection in France: 9 new apricot cultivars for French growers. *Acta Hort.* 384, 237-243.

Ayanoglu, H.; Kaska, N. & Yildiz, A. (1995). Akdeniz Bölgesi'nde erkenci kayısı çeşitlerinin adaptasyonu üzerinde araştırmalar. Türkiye II. Ulusal Bahçe Bitkileri Kongresi Kitabı, Vol. 1, 159-163. (in Turkish)

Badanes, M.L.; Asins, M.J.; Carbonell, E.A. & Llacer, G. (1996). Genetic diversity in apricot, Prunus armeniaca, aimed at improving resistance to plum pox virus. *Plant. Breed.* 115, 133-139.

Badanes, M.L.; Martinez-Calvo, J. & Llacer, G. (1998). Analysis of apricot germplasm from the European ecogeographical group. *Euphytica* 102, 93-99.

Bailey, L.H. (1916). Prunus. In: The standard cyclopedia of horticulture, vol. V.P-R. Mount Pleasant Press, J. Horace McFarland Co., Harrisburg, PA, pp. 2822-2845.

Bailey, L.H. & Hough, L.F. (1975). Apricots. In: Janick J, Moore JN (eds) Advances in fruit breeding. Purdue University Press, West Lafayette, Indiana, pp. 367-383.

Baktir, I.; Ulger, S. & Yayıcı, Z.H. (1992). Yabancı orijinli bazı kayısı çeşitlerinin Antalya koşullarında adaptasyonu ve gelişmeleri üzerine bir araştırma. Türkiye I. Ulusal Bahçe Bitkileri Kongresi Kitabı, Vol. 1, 461-464. (in Turkish)

Bassi, D., & Sansavini, S. (1988). Miglioramento genetico dell'albicocco: stata dele richerche e prospettive. *Frutticoltura* 6, 11-22.

Bassi, D. & Bartolozzi, F. (1993). Il miglioramento genetico dell'albicocco: obiettive e strategie. Rivista di Frutticoltura 5, 41-50.

Bassi, D.; Bellini, E.; Guerriero, R.; Monastra, F. & Pennone, F. (1995). Apricot Breeding in Italy. *Acta Hort.* 384, 47-54.

Bassi, D. & Audergon, J.M. (2006). Apricot Breeding: Update and Perspectives. *Acta Hort.* 701, 279-294.

Battistini, S. & Sansavini, S. (1991). Electrophoretic analysis of isozyme variability in apricot cultivars. *J. Genet. Breed.* 45, 117-122.

Blanc, A.; Broquaire, J.M.; Chauffour, D.; Clauzel, G.; Duffillol, J.M.; Giard, A.; Gilles, F.; Moulon, B & Audergon, J.M. (2006). Soledane, Florilege, Bergarouge ® Avirine, three new apricot cultivars for French Country. *Acta Hort.* 701, 395-398.

Bortiri, E.; Oh, S.H.; Jiang, J.; Baggett, S.; Granger, A.; Weeks, C.; Buckingham, M.; Potter, D. & Parfitt, D.E. (2001). Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Syst. Bot.* 26, 797-807.

Bortiri, E.; Oh, S.H.; Gao, F.Y. & Potter, D. (2002). The phylogenetic utulity of nucleotide sequence analysis of sorbitol 6-phosphate dehydrogenase in *Prunus* (Rosaceae). *Am. J. Bot.* 89, 1697-1708.

Burgos, L.; Perez-Tornero, O.; Ballester, J. & Olmos, E. (1998). Detection and inheritance of stylar ribonucleases associated with incompatibility alleles in apricot. *Sex. Plant Reprod.* 11, 153-158.

Buttner, R. (2001). Armeniaca. In: P. Hanelt (ed) Institute of Plant Genetics and Crop Plant Researches Mansfelds Encyclopedia of Agricultural and Horticultural Crops, pp 523-527.

Cambra, M.; Capote, N.; Myrta, A. & Llácer, G. (2006). Plum pox virus and the estimated costs associated with sharka disease. *EPPO Bulletin* 36, 202-204.

Chenjing, F.; Yuanhui, Z. & Xiuying, X. (2005). Genetic diversity revealed by ISSR marker in apricot. *Journal of Agricultural University of Hebei* 28(5), 52-55.

Cociu, V.; Balan, V. & Oprea, M. (1991). The variability of field resistance against patogenic agents *Monilia laxa* (Aderh. et Ruhl) honey and *Stignina carpophila* (LEV) M.B. Ellis and *Cytospora cincta* Sacc in some hybrids. *Acta Hort.* 293, 203-210.

Crossa-Raynaud, P. (1960). Problems d'arboriculture fruitiere en Tunise. Abricotiers. *Ann. L'Institut National de la Recherche Agronomique de Tunise* 33, 39-63.

Crossa-Raynaud, P. & Audergon, J.M. (1991). Some reflexions on apricot selections. *Acta Hort*. 293, 73-85.

Davarynejad, G.H.; Szabó, Z.; Váci Felhösné, E.; Kun, Z. & Nyéki, J. (2005). Anther and pollen grain characteristics of apricot cultivars. *Acta Hort.* 384, 351-354.

Demirtas, M.N.; Atay, S.; Gokalp, K.; Yilmaz, K.U.; Sahin, M.; Olmez, H.A.; Paydas Kargi, S.; Altindag, M & Ercisli, S. (2010). Improvement of late blooming drying apricot varieties by crossbreeding. *Acta Hort.* 862, 219-223.

De Nettancourt, D. (2001). Incompatibility and incongruity in wild and cultivated plants. 2nd ed. Springer-Verlag, New York.

De Poerderlé, M. le Baron (1788). Manuel de l'Arboriste et du Forestier Belgiques: Seconde Édition: Tome Premier (MDCCLXXXVIII). à Bruxelles: Emmanuel Flon. p. 682. Downloadable Google Books.

De Vicente, M.C.; Truco, M.J.; Egea, J.; Burgos, L. & Arus, P. (1998). RFLP variability in apricot (*Prunus armeniaca* L.). *Plant Breed.* 117(2), 153-158.

Dezhong, T.; Baoming, W.; Gaixiu, D. & Xiafung, F. (1995). Studies on the pollen morphology and ultrastructure of cultivated varieties of apricot (*Armeniaca vulgaris* Lam). *Acta Hort.* 403, 140-144.

Draganescu, E. & Cociu, V. (1997). The modernizing of the apricot assortment of cultivars in Banat Area of Romania. *Acta Hort.* 488(1), 149-151.

Durie, B. (1988). Commercial characteristics of some apricot varieties in Vojvodina. *Hort. Abstr.* 61(4), 2615.

Egea, J.; Garcia, J.E.; Egea, L. & Berenguer, T. (1993). Productive behaviour of apricot varieties in a warm winter area. *Acta Hort.* 384, 129-133.

Egea, J. & Burgos, L. (1996). Detecting cross-incompatibility of three North American apricot cultivars and establishing the first incompatibility group in apricot. *J. Amer. Soc. Hort. Sci.* 121, 1002-1005.

Entani, T.; Iwano, M.; Shiba, H.; Che, S.F.; Isogai, A. & Takayama, S. (2003). Comparative analysis of the self-incompatibility (*S*-) locus region of *Prunus mume*: Identification of a pollen-expressed *F-box* gene with allelic-diversity. *Genes. Cells* 8, 203-213.

Ercisli, S.; Agar, G.; Yildirim, N.; Esitken, A. & Orhan, E. (2009). Identification of apricot cultivars in Turkey (*Prunus armeniaca* L.) using RAPD markers. *Romanian Biotechnological Letters* 14, 4582-4588.

FAOSTAT (2011). FAO web page, www.fao.org

Faust, M.; Suranyi, D. & Nyujto, F. (1998). Origin and Dissemination of Apricot. *Horticultural Reviews* 22, 248-249.

Felföldi, J.; Firtha, F.; Herman, R. & Pedryc, A. (2009). Characterisation of morphological properties of apricot stones by image proccessing. *Bornimer Agrartechnische Berichte*, Helf 69, 162-169.

Geuna, F.; Toschi, M. & Bassi, D. (2003). The use of AFLP markers for cultivar identification in apricot. *Plant. Breed.* 122, 526-531.

Guerriero, R. & Watkins, R. (1984). Revised descriptor list for apricot (*Prunus armeniaca*). IBPGR Secretariat, Rome, CEC Secretariat, Brüssels.

Guerrieri, F.; Audergon, J.M.; Albagnac, G. & Reich, M. (2001). Soluble sugars and carboxylic acids in ripe apricot fruit as parameters for distinguishing different cultivars. *Euphytica* 117, 183-189.

Guerriero, R.; Monteleone, P. & Iacona, C. (2006). Apricot Breeding in Pisa: Some New Selections for Italian Growers. *Acta Hort.* 701, 341-346.

Gulcan, R.; Misirli, A. & Demir, T. (1994). Hacıhaliloğlu Kayısı Çeşidinin Melezleme Yoluyla Monilya (*Sclerotinia* (*Monilinia*) *laxa* Aderh et., Ruhl) Hastalığına Dayanıklılık Islahı Üzerinde Bir Araştırma - TÜBİTAK Proje No. TOAG-806. (in Turkish)

Hagen, L.; Khadari, B.; Lambert, P. & Audergon, J.M. (2002). Genetic diversity in apricot revealed by AFLP markers: species and cultivar comparisons. *Theor. Appl. Genet.* 105(2-3), 298-305.

Halasz, J.; Hegedus, A.; Herman, R.; Stefanovits-Banyai, E. & Pedryc, A. (2005). New self-incompatibility alleles in apricot (*Prunus armeniaca* L.) revealed by stylar ribonuclease assay and *S*-PCR analysis. *Euphytica* 145, 57-66.

Halasz, J. (2007). Molecular background of the S-locus controlled self-incompatibility in apricot. PhD Diss., Corvinus University of Budapest, Budapest – Hungary.

Halasz, J.; Pedryc, A. & Hegedus, A. (2007). Origin and dissemination of the pollen-part mutated *Sc*-haplotype that confers self-compatibility in apricot (*Prunus armeniaca*). *New Phytol.* 176, 793-803.

Halasz, J.; Pedryc, A.; Ercisli, S.; Yilmaz, K.U. & Hegedus, A. (2010). *S*-genotyping supports the genetic relationships between Turkish and Hungarian apricot germplasm. *J. Amer. Soc. Hort. Sci.* 135(5), 410-417.

Harsanyi, J. (1991). Evaluation of apricot varieties at a plain site exposed to frost. *Acta Hort.* 293, 217-220.

Hayashi, K.; Shimazu, K.; Yaegaki, H.; Yamaguchi, M.; Iketani, H. & Yamamoto, T. (2008). Genetic diversity in fruiting and flower-ornamental Japanese apricot (*Prunus mume*) germplasms assessed by SSR markers. *Breeding Science* 58, 401-410.

He, T.M.; Chen, X.S.; Gao, J.S.; Zhang, D.H.; Xu, L. & Wu, Y. (2006). Using SSR markers to study population genetic structure of cultivated apricots native to Xinjiang. *Acta Hort. Sinica* 33, 809-812. (in Chinese with an English abstract)

Hofstee, M.E.; Malona, M. & Howard, C. (1997). Apricot breeding in New Zealand. *Acta Hort.* 488(1), 171-172.

Hormaza, J.I. (2002). Molecular characterization and similarity relationships among apricot (*Prunus armeniaca* L.) genotypes using simple sequence repeats. *Theor. Appl. Genet.* 104(2-3), 321-328.

Horiuchi, S.; Yoshida, M.; Kariya, H.; Nakamura, T.; Hasebe, H.; Suzaki, T. & Sakitani, T. (1996). Nihonnoume Sekainoume. Yokendo.

Hou, H.Y. (1983). Vegetation of China with referance to its geographical distrubition. *Ann. Missouri Bot. Gard.* 70, 509-548.

Hurtado, M.A.; Badenes, M.L.; Llacer, G.; Westman, A.; Beck, E. & Abbott, G.A. (2001). Contribution to apricot genetic analysis with RFLP, RAPD and AFLP markers. *Acta Hort.* 546, 417-420.

Hurtado, M.A.; Westman, A.; Beck, E.; Abbott, G.A.; Llacer, G. & Badenes, M.L. (2002). Genetic diversity in apricot cultivars based on AFLP markers. *Euphytica* 127, 297-301.

Huxley, A. (1992). New RHS Dictionary of Gardening 1: 203–205. Macmillan ISBN 0-333-47494-5.

Karp, A.; Seberg, O. & Buiatti, M. (1996). Molecular techniques in the assessment of botanical diversity. *Ann. Bot.* 78, 143–149.

Kaska, N.; Yildiz, A.; Ayanoglu, H.; Saglamer, M. & Gungor, M.K. (1993). Apricot adaptation studies in the Mediterranean coastal region in Turkey. *Acta Hort.* 384, 67-71.

Kegler, H.; Fuchs, E.; Grüntzig, M. & Schwarz, S. (1998). Some results of 50 years of research on the resistance to plum pox virus. *Acta Virol.* 42, 200-215.

Kostina, K.F. (1969). The use of varietal resources of apricots for breeding. *Trud. Nikit. Bot. Sad.* 40, 45-63. (in Russian)

Kostina, K.F. (1970). Self-fertility studies in apricot. *Trud. Nikit. Bot. Sad.* 45, 7-17. (in Russian)

Krichen, L.; Trifi-Farah, N. & Marrakchi, M. (2006). Variability, organisation and identification of Tunisian apricot (*Prunus armeniaca* L.) cultivars using AFLP markers. *Acta Hort.* 717, 251-254.

Layne, R.E.C.; Bailey, C.H. & Hough, L.F. (1996). Apricots. J. Janick and J.M. Moore (Eds.) In: Fruit Breeding, Vol 1: Tree and tropical fruits, pp. 79-111, John Willey and Sons, New York.

Ledbetter, C.A. & Petterson, S. J. (2004). Utulization of Pakistani apricot (Prunus armeniaca) germplasm for improving brix levels in California adapted apricots. *Plant Genetic Resources Newsletter* 140, 14-22.

Ledbetter, C.A. (2008). Apricots: Hancock, J.F. (eds), Temperate Fruit Crop Breeding: Germplasm to genomics, Chapter 2, pp. 39-82. Springer Science-Business Media B.V.,

Lingdi, L. & Bartholomew, B. (2003). Armeniaca. In: Wu, C.Y., Raven, P.H. (eds.) Flora of China, vol 9 (*Pittosporaceae* through *Connaraceae*). Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis. pp 396-401.

Loopez-Moya, J.; Fernandez-Fernandez, M.; Cambra, M. & Garcia, J. (2000). Biotechnological aspects of plum pox virus. *J. Biotechnol.* 76, 121–136.

Loudon, J.C. (1838). Arboretum Et Fruticetum Britannicum. Vol. II. London: Longman, Orme, Brown, Green and Longmans. pp. 681–684. The genus is given as Armeniaca. Downloadable at Google Books.

Maghuly, F.; Borroto Fernandez, E.; Zelger, R.; Marschall, K.; Katinger, H. & Laimer, M. (2006). Genetic differentiation of apricot (*Prunus armeniaca*) cultivars with SSR markers. *Europ. Journal of Horticultural Science* 71, 129–134.

Maghuly, F.; Borroto Fernandez, E.; Ruthner, S.; Pedryc, A. & Laimer, M. (2005). Microsatellite variability in apricot (*Prunus armeniaca*) reflects their geographical origin and breeding history. *Tree Genetics & Genomes* 1, 151-165.

Malik, S.K.; Chaudhury, R.; Dhariwal, O.P. & Mir, S. (2010). Genetic diversity and traditional uses of wild apricot (*Prunus armeniaca* L.) in high-altitude north-western Himalayas of India. *Plant Genetic Resources: Characterization and Utulization* 8, 249-257.

Marinello, L.; Sommella, M.G. & Sorrentina, A. (2002). Identification of *Prunus armeniaca* cultivars by RAPD and SCAR markers. *Biotechnology Letters* 24, 749-755.

Martinez-Gomez, P.; Dicenta, F. & Audergon, J.M. (2000). Behaviour of apricot (*Prunus armeniaca* L.) cultivars in the presence of Sharka (*Plum pox potyvirus*): a review. *Agronomie* 20, 407–422.

McClure, B.A.; Haring, V.; Ebert, P.R.; Anderson, M.A.; Simpson, R.J.; Sakiyama, F. & Clarke, A.E. (1989). Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. *Nature* 342, 955-957.

Mega, K.; Tomita, E.; Kitamura, S.; Saito, S. & Mizukami, S. (1988). Ume. In: Aoba et al. (Eds.) The Grand Dictionary of Horticulture, Shogakukan, Tokyo, pp. 289-300.

Mehlenbacher, S.A.; Cociu, V. & Hough, L.F. (1991). Apricots (Prunus). In: Moore J. N., Ballington J. R. (eds) Genetic Resources of Temperate Fruit and Nut Crops. *Acta Hort.* 290, 63-109.

Milosevic, T.; Milosevic, N.; Glisic, I. & Krska, B. (2010). Characteristics of promising apricot (*Prunus armeniaca* L.) genetic resources in Central Serbia based on blossoming period and fruit quality. *Hort. Sci. (Prague,* 37(2), 46-55.

Nicotra, A.; Conte, L.; Moser, L.; Fantechi, P.; Corazza, L.; Vitale, S. & Magnotta, A. (2006). Breeding programme for *Monilinia laxa* (Aderh et., Ruhl.) resistance on apricot. *Acta Hort.* 701, 307-311.

Occarso, G. (1977). The apricot in Western Sicily, descripton of interesting very early cultivar. *Hort. Abstr.* 51(10), 7656.

OECD (2002) Consensus document on the biology of Prunus sp. (stone fruits): series on harmonization of regulatory oversights in biotechnology nr. 24. http://www.olis.oecd.org/olis/2002doc.nsf/LinkTo/env-jm-mono

Ogasanovic, D.; Plazinic, R.M. & Papic, U.M. (1991). Results from study of some early apricot cultivars on various interstock. *Acta Hort.* 293, 383-389.

Ozvardar, S.; Onal, K. & Baldiran, E. (1991). Ege Bölgesi'ne uygun kayısı çeşitlerinin seçimi. *Anadolu Ege Tarımsal Araştırma Enstitüsü Dergisi* 2, 36-52. (in Turkish)

Panaud, O.; Chaib, A. & Sarr, A. (2002). Dynamic conservation of apricot Prunus armeniacain Saharian oases: use of AFLP markers to assess genetic diversity in traditional orchards. *Euphytica* 128, 301-305.

Papanikolaou-Paulopoulo, X. & Poulis, J. (1997). Evaluation of fourteen (14) different apricot varieties at the Agricultural Research Station of Rhodes. *Acta Hort.* 488(1), 185-189.

Parolari, G.; Virgili, R. & Bolzoni, L. (1992). Analysis of sensory and instrumental data on apricot purees with pattern recognition techniques. *Analı´tica Chimica Acta* 259, 257-265.

Paydas, S.; Kaska, N. & Durgac, C. (1995). Yerli ve yabancı bazı kayısı çeşitlerinin Pozantı ekolojik koşullarındaki performansları. Türkiye II. Bahçe Bitkileri Kongresi Kitabı, Vol. 1, 169-173. (in Turkish)

Pedryc, A.; Ruthner, S.; Herman, R.; Krska, B.; Hegedus, A. & Halasz, J. (2009). Genetic diversity of apricot revealed by a set of SSR a markers from linkage group G1. *Sci. Hort.* 121, 19-26.

Perez-Gonzales, S. (1992). Associations among morphological and phenological characters representing apricot germplasm in Central Mexico. *J. Amer. Soc. Hort. Sci.* 117, 486-490.

Rao, R.; Bencivenni, M.; Basile, B. & Forlani, M. (2008). Molecular characterization of the apricot cultivars of protected geographical indication (PGI) "Albicocca Vesuviana". XIV. International Symposium on Apricot Breeding and Culture, 16-20 June 2008, Matera (Italy), abstract books.

Rehder, A. (1940). Manual of cultivated trees and shrubs hardy in North America, exclusive of the subtropical and warmer temperate regions, 2nd revised and enlarged edition. Macmillian, New York.

Rodrigo, J.; Herrero, M. & Hormaza, J.I. (2006). Morphological and physiological parameters releated to flower quality in apricot. *Acta Hort,* 717, 89-90.

Romero, C.; Pedryc, A.; Munoz, V.; Llacer, G. & Badenes, M.L. (2003). Genetic diversity of different apricot geographical groups determined by SSR markers. *Genome* 46, 244-252.

Romero, C.; Vilanova, S.; Burgos, L.; Martinez-Calvo, J.; Vicente, M.; Llacer, G. & Badanes, M.L. (2004). Analysis of the *S*-locus structure in *Prunus armeniaca* L.: Identification of *S*-haplotype specific *S-RNase* and *F-box* genes. *Plant Mol. Biol.* 56, 145-157.

Romero, C.; Llacer, G.; Badanes, M.L. & Pedryc, A. (2006). Relationship among apricot cultivars from Hungary and a South European pool determined by SSR markers. *Acta Hort.* 701(1), 233-240.

Rostova, N.S. & Sokolova, E.A. (1992). Variability of anatomical and morphological leaf characters in apricot (*Armeniaca* Scop.) species and varieties. *Bull. Appl. Bot. Genet. Plant Breed.* 146, 74-86.

Ruiz, D. & Egea, J. (2008). Phenotypic diversity and relationships of fruit quality traits in apricot (*Prunus armeniaca* L.) germplasm. *Euphytica* 163, 143-158.

Sanchez-Pe´rez, R.; Ruiz, D.; Dicenta, F.; Egea, J. & Martinez- Gomez, P. (2005). Application of simple sequence repeat (SSR) markers in apricot breeding: Molecular characterisation, protection and genetic relationships. *Scientia* 103, 305-315.

Souty, M.; Audergon, J.M. & Chambroy, Y. (1990). Abricot: les crite`res de qualite´. *L'Arboriculture fruitie`re* 430, 16-24.

Suranyi, D.; 1999. Apricot culture in Hungary: Past and present. *Acta Hort.* 488, 205-210.

Szabo, Z. & Nyeki, J. (1991). Blossoming, fructification and combination of apricot varieties. *Acta Hort.* 293, 295-302.

Takeda, T.; Shimada, T.; Nomura, K.; Ozaki, T.; Haji,T.; Yamaguchi, M. & Yoshida, M. (1998). Classification of apricot varieties by RAPD analysis. *Journal of the Japanese Society for Horticultural Science* 67, 21–27.

TurkStat (2010). Turkish Statistical Institute, www.turkstat.gov.tr

Vargas, F.J., & Romero, M.A. (2001). Blooming time in almond progenies. *Options Me´diterran* 56, 29–34.

Vavilov, N.I. (1951). Phytogeographic basis of plant breeding. The origin, variation, immunity and breeding of cultivated plants. K.S. Chester (translated). *Chron. Bot.* 13, 13-54.

Vilanova, S.; Romero, C.; Abbott, A.G.; Llacer, G. & Badanes, M.L. (2003). An apricot (*Prunus armeniaca* L.) F_2 progeny linkage map based on SSR and AFLP markers, mapping *Plum pox virus* resistance and self-incompatibility traits. *Theor. Appl. Genet.* 107(2), 239-247.

Viti, R.; Bartolini, S. & Minnocci, A. (2000). Morphological structure of the stigma and style of several genotypes of *Prunus armeniaca* L. *Plant Biosystems* 134, 45-51.

Williams, J.G.K.; Kubelik, A.R.; Livak, K.J.; Rafalski, J.A. & Tingey, S.V. (1990). DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18, 6531-6535.

Yilmaz, K.U. (2002). Studies on table apricot growing for export in Kahramanmaras. University of Kahramanmaras Sutcu Imam, Institute of Natural and Applied Science, Department of Horticulture, MSc Thesis. (in Turkish with English abstract)

Yilmaz, K.U. (2008). Phenological, morphological and pomological characteristics of some local apricot genotypes and determination of their genetic relationships and self-incompatibility by molecular markers. Department of Horticulture Institute of Basic and Applied Sciences University of Cukurova, Adana-Turkey, PhD Thesis, (in Turkish with an English abstract).

Yilmaz, K.U. & Paydas-Kargi, S. (2010). A new morphological trait for apricot characteristics. *Acta Hort.* 862, 83-84.

Yuan, Z.; Chen, X.; He, T.; Feng, J.; Feng, T. & Zhang, C. (2007). Population genetic structure in apricot (*Prunus armeniaca* L.) cultivars revealed by fluorescen-AFLP markers in Southern Xinjiang, China. *J. of Genetics and Genomes* 37(4), 1037-1047.

Zhebentyayeva, T.N. & Sivolap, Y.M. (2000). Genetic diversity of apricot determined by isozyme and RAPD analyses. *Acta Hort.* 538, 525-529.

Zhebentyayeva, T.N.; Reighard, G.L.; Gorina, V.M. & Abbott, A.G. (2003). Microsatellite (SSR) analysis for assessment of genetic variability in apricot. *Theor. Appl. Genet.* 106, 435-444.

Zhebentyayeva, T., Reighard, G., Lalli, D., Gorina, V., Krška, B. & Abbott, A.G. (2008). Origin of resistance to plum pox virus in apricot: what new AFLP and targeted SSR data analyses tell. *Tree Genet. Genomes* 4, 403–417.

Zohary, D. & Spiegel Roy, P. (1975). Beginning of fruit growing in the Old World. *Science* 187, 319–327.

14

Sampling the Genetic Diversity of Tall Fescue Utilizing Gamete Selection

Bryan Kindiger
USDA-ARS, Grazinglands Research Laboratory, El Reno, OK USA

1. Introduction

Tall fescue (*Lolium arundinaceum* (Schreb.) S.J. Darbyshire) (2n=6x=42) represents the predominant perennial cool-season grass forage in the USA. Its wide adaptation, excellent spring, summer and fall production, deep root system, tolerance to heat and persistence over summer conditions makes this a highly desirable species for hay, pasture and turf. Tall fescue tolerates short-term flooding, moderate drought and heavy livestock grazing and machinery traffic. It responds well to fertilizer but can maintain itself under limited fertility conditions and is adapted to moderately acid and wet soils (Jennings et al., 2008). Approximately two-thirds of the total annual growth of tall fescue occurs during the spring and about one-third occurs during summer and fall making it nearly a full season forage. A wide array of phenotypic and genotypic variation exists in tall fescue and cultivar development has traditionally used recurrent selection techniques. For this review, we focus primarily on tall fescue as a forage or hay producing species; however, methods discussed can be equally applicable to turf programs. Although tall fescue is a complex hexaploid species, it can be hybridized with annual (2n=2x=14) or perennial ryegrass (2n=4x=28) (*Lolium perenne* L. subsp. *multiflorum* (Lam.) Husnot (syn. *Lolium multiflorum* Lam.) for the development of festuloliums (Yamada and Takamizo, 2004). In some instances, induced androgenesis has been utilized to generate useful germplasm (Vagera et al., 1998).

Gamete selection as originally defined by Stadler (1944) is based on the principal that selection exerted at the gametophytic level can increase desirable allelic frequencies detectable at the sporophytic level. If superior gametes can be recognized with certainty through a selection cycle, then such a system would be theoretically more efficient than one based on zygotic selection (Richey, 1947). In practice, gamete selection ordinarily involves two steps: 1) selection on the basis of outcross performance testing of individual plants of a variety or population, and 2) a similar controlled selection for outstanding individuals exhibiting desirable agronomic attributes. Following the identification of superior genotypes, such individuals would undergo continued selfing, followed by phenotypic selection, to generate a homozygous line, fixed for the desired agronomic characteristics. In instances where haploids can be generated through microspore culture, followed by genome doubling, homozygous or dihaploid lines can be obtained.

Breeding methods that utilize haploids can be varied and often provide materials exhibiting particular usage in genetic analysis and breeding systems when properly exploited (Dunwell, 2010). In addition, gametophytic selection approaches can lead to a correlated sporophytic response that can involve the segregation of important agronomic traits (Mulcahy, 1971). When applicable, gamete selection has emerged as a superior and efficient form of selection and the feasibility of this approach appears to be a promising method for facilitating the incorporation of desirable alleles within a short period of time. Stadler (1944) is among the earliest who reviewed its potential application in corn breeding. More recently, the importance of a gamete form of selection, utilizing early generation selection, has demonstrated successful simultaneous selection of multiple traits, including quantitative trait loci for characteristics such as seed yield, maturity, and tolerance to disease (Singh, 1994; Ravikumar et al., 2004). In these instances, gamete selection was found to be superior in its efficiency to conventional pedigree, single seed descent, and mass selection methods (Singh, 1994). A gamete selection approach may also be equally effective on complex traits such as biotic and abiotic stress, herbicide tolerance, general and specific combining ability, and the selection of complex quantitative traits such as yield, drought tolerance, disease resistance, etc. Traditionally applied to commercial grain crops such as corn (Gordilli and Geiger, 2008; Chang and Coe, 2009), wheat (Barclay, 1975), barley (Kasha and Kao, 1970; Shugar, 1989) and beans (Singh et al., 1998) gamete selection often utilizes the induction of haploids and the subsequent generation of double haploid breeding approaches. With the use of these methods, various degrees of success in gamete selection have been attained. The generation of haploid or monoploid individuals is typically achieved through microspore culture methods (Mohan et al., 1997; Maheshwari et al., 1980; Tuvesson et al., 2008), wide hybridization (Inagaki, 1997), or through a pollen/sperm nucleus utilizing a haploid inducing system (Chang and Coe, 2009). This approach has even been applied to tall fescue (Kasperbauer et al., 1980; Bai & Qu, 2000). Each of these methods relies on the haploid individual normalizing its chromosome number either through the utilization of mitotic inhibitors such as colchicine, or through spontaneous doubling. This chromosome doubling event results in the generation of individuals that are homozygous for the genotype conferred by the single contributing gamete. By using a paternal monoploid or dihaploid generation process, gamete selection is a proven and efficient method of selection that has yet to be applied to forage grass breeding (Stadler 1944; Lu, et al., 1996; Rotarenco and Chalyk, 2000). Dihaploid (DH) breeding methods have been suggested and employed as a means to develop new germplasm in the *Festuca-Lolium* complex of polyploid grasses (Humphreys et al., 2003; Guo and Yamada, 2004; Guo et al., 2005). Previous research utilizing homozygous lines to study the inheritance of palatability has suggested that selection within homozygous lines can represent a useful methodology for the improvement of tall fescue (Henson and Buckner, 1957; Buckner and Fergus, 1960a). As a consequence, when the gamete selection approach is utilized in a breeding program, the generation of DH tall fescue lines should result in the development of efficient breeding advances and superior germplasm (Bouchez and Gallais, 2000). Additional research also indicates that genetic mapping studies focused on the elucidation of a variety of complex genetic traits can be more readily accomplished through the use of homozygous lines (Riera-Lizarazu et al., 2010). The application of gamete selection to tall fescue could be accomplished if an efficient system of haploid induction followed by parthenogenesis were available.

Sporophytically expressed traits are transmitted as genetic information through the gametes (sperm or egg nuclei) and contain half the information that is contained in the sporophytic tissue. In some instances, gametophytic genes can regulate both gamete and sporophytic traits and this is called genetic overlap. In the presented gamete selection approach, a slight modification of the methodology as presented by Stadler and his contemporaries is utilized. Employing the method of gamete selection considered in this text, traits transmitted by the gamete are under the control of genes that are expressed specifically in the sporophyte. A single gamete provided by the tall fescue parent fertilizes the egg of an inducer line (IL) that results in genome loss and a parthenogenic response. Individuals derived from this hybridization event, followed by genome loss and the parthenogenic response, are recovered ryegrass, tall fescue or varied intermediate genotypes of ryegrass-fescue (e.g. festulolium). Selection applied at the F1 sporophyte results in a form of gamete selection.

The USDA-ARS has recently developed an approach utilizing *Lolium multiflorum* lines (IL1 and IL2) that are uniquely characterized by their ability to loose either a ryegrass or tall fescue genome following hybridization and the expression of a low frequency of parthenogenic behavior (Kindiger, 2009). Parthenogenesis involves development of the egg cell within the embryo sac without involvement of the sperm nucleus and such behavior has been characterized across numerous species, especially following wide hybridizations (Kendall, 1934). In this review, a low frequency parthenogenic response represents the first reported incidence of such behavior in ryegrass x tall fescue hybrids. The frequency of this behavior is difficult to determine, however, it is estimated to occur less than 1% which is typical for other species (Kendall, 1934). This frequency can and does vary across tall fescue genotypes hybridized with the IL lines.

Two ryegrass (*Lolium perenne* L. subsp. *multiflorum* (Lam.) Husnot (syn. *Lolium multiflorum* Lam.) (2n=2x=14) genetic stocks, identified as IL1 and IL2 (Kindiger and Singh, 2011; Kindiger, 2011), are characterized by a genome loss phenomenon following hybridization with tall fescue (*Lolium arundinaceum* (Schreb.)) S.J. Darbyshire (syn. *Festuca arundinacea* Schreb.) ($2n = 6x = 42$)\, followed by a low level of parthenogenic development. The IL1 and IL2 genetic stocks exhibit few advantageous agronomic characteristics and are essentially notable only for their ability to induce chromosome or genome loss following hybridization. Both lines are free of the fungal endophytes *Epichloë* sp. or *Neotyphodium* sp. (Carroll, 1988; Moon et al., 1994; Pederson & Sleper, 1988). Recurrent selection remains a widely utilized approach for breeding most cereal and forage species (Gallais, 1993) and has also been applied utilizing DH lines (Bouchez and Gallais, 2000). Incorporation of a gamete selection methodology should provide an equally successful approach for tall fescue improvement.

In the approach, hybrids are generated utilizing genetic stocks IL1 or IL2 as the maternal parent and a single tall fescue individual or population as the paternal parent. Varietal or individual plant sampling within improved tall fescue cultivars or populations will yield a myriad of segregating gametes possessing a wide array of advantageous and disadvantageous alleles. Each pollen sperm nuclei will contribute a unique genotype to each hybrid and as a result, each hybrid represents a random genotypic combination of the IL maternal line and tall fescue pollen parent. Since the IL1 and IL2 lines are narrow based, agronomically inferior, unimproved genetic stocks, any potential advantageous contribution they may provide is considered minimal, notwithstanding unknown potential heterotic

effects in the F1 that may occur between the *Lolium* and *Festuca* genomes. The primary focus of the approach is placed on the gametic contribution of the paternal tall fescue parent and its expression in the hybrid sporophyte. Through this process, the breeder or geneticist is essentially sampling the genetic diversity of the paternal individual on a single pollen grain basis and examining each contribution in the hybrids.

2. Methodology

Pollinations are generated by hand utilizing the appropriate IL genetic stock as the maternal parent. Hybrid seed is easily generated in quantity. Following germination of the hybrid seed, numerous ryegrass-tall fescue hybrids are now available with each F1 hybrid being derived from a single tall fescue pollen grain sperm nucleus fertilizing the IL egg. The chromosome number of the hybrids is typically 2n=2x=28 with each hybrid possessing one genomic contribution from each parent (ryegrass (n=7) and tall fescue (n=21). Hybrids are generally sterile, but a low incidence of fertility can occasionally occur (Buckner, 1960b; Buckner et al., 1961). The generated seed can be sown in low density space planting nurseries, grown in the greenhouse, or planted to a spaced planted nursery where various induced or natural selection pressures can be applied to the hybrid individuals (Figures 1a, 1b, 1c). Multiple locations are desirable to focus selection on the particular attributes such as stress or rust tolerance in a region where the germplasm is planned to be released (Figures 2a, 2b). If disease tolerance is a selection criterion, then the hybrids should be grown in a region where the specific disease under study is prevalent. Once exceptional hybrid individuals are identified, the F1's can be transferred to the greenhouse to exclude any chance of cross-pollination with any tall fescue pollen in the field. If an abundance of tall fescue pollen is not a problem in the nursery, then the hybrids can remain in the nursery for a future harvest of each selected plant inflorescence. Hybrid individuals that do not possess an appropriate genotype contribution from the paternal tall fescue parent are culled from the nursery. If multiple years of selection are to be performed, seed heads are not to be retained until the final year of selection.

Fig. 1. a (left) Low density nursery planting of IL X tall fescue seed; USDA-ARS, Grazinglands Research Laboratory, El Reno, OK, USA. Figure 1b (center) Space planted nursery of IL x tall fescue hybrids. Figure 1c (right) Stress tolerant and stress intolerant hybrids from the IL x tall fescue hybrid space planted nursery.

Fig. 2. a (left) IL x tall fescue F1 drought selection nursery at he Mound Valley Unit, Kansas State University, Southeast Agricultural Research Center, KS in July, 2010. Circled individuals identify F1 plants exhibiting superior drought tolerance and plant vigor. Figure 2b (right) IL x tall fescue F1 rust tolerance selection nursery at Barenbrug Seeds, West Coast Research Center, Albany, OR in August, 2011.

The surviving hybrids are allowed to flower; then inflorescences are gathered at maturity, broken up by hand or machine and sown to trays. A light cleaning is applied to remove stems. The cleaned seed heads are then placed in germination trays for identification and selection of either recovered ryegrass or tall fescue seedlings. Typically, following two weeks of germination, a few seedlings will appear and are allowed to grow to appropriate size for transplanting (Figure 3). The germinating seedlings will generally be ryegrass recoveries possessing a chromosome number of 2n=2x=14, tall fescue recoveries possessing a chromosome number of 2n=6x=42, or various tall fescue DH recoveries with ryegrass introgression or ryegrass recoveries with tall fescue introgression. Since the seedlings obtained from the sterile F1 hybrids are generated through parthenogenic development following spontaneous chromosome doubling, each recovery will possess a full genome contribution of either the ryegrass or tall fescue parent. Genome loss, spontaneous doubling followed by parthenogenic development represent the important and essential contributions of the IL1 and IL2 lines. All recovered lines derived from this process will have all genes, alleles or quantitative loci conferring a trait fixed in the DH recovery (Kindiger and Singh, 2011; Kindiger 2011). Essentially, the process generates homozygous or DH lines from the IL x tall fescue F1 hybrids.

Each recovered individual will be free of the *Lolium* sp. fungal endophytes since the IL1 and IL2 lines do not possess endophyte. Recovered tall fescue DH lines can then be evaluated under additional selection schemes and eventually inter-crossed to perhaps generate a synthetic possessing various advantageous attributes, or be utilized as breeding lines for the development of cultivars.

3. Example 1: Gamete selection for crude protein

The development of forages with superior nutritional qualities is an ongoing and time consuming breeding effort often hindered by the complex genetics associated with each quality trait (Casler, 2001; Bouton, 2009). Since grazing animals have specific and differing

Fig. 3. Germination of dihaplod seedlings from a selected IL x tall fescue F1 hybrid. These individuals will represent DH tall fescue recoveries.

nutritional requirements, the qualities of the forage should be adjusted to fit that particular need (Brummer and Casler, 2009). Intergeneric hybridization between *Lolium* and fescue

has been a common method to understand and improve the genetics underlying the forage quality (Naganowska et al., 2001; Cogan et al., 2005). In addition, applying approaches such as introgression mapping facilitated with either genomic *in situ* hybridization techniques or molecular markers, provide an additional set of powerful tools for forage quality improvements (Cardinal et al., 2003). The close homology between the *Lolium* sp. and *Festuca* sp. genomes can also allow either recombination or a full substitution of a *Lolium* chromosome for a *Festuca* chromosome (Kopecký et al., 2009) and such events could affect the expression of a forage quality. It may also be useful to incorporate molecular markers during the breeding process or during backcross generations to identify any ryegrass segments that might have been introgressed into the fescue genome (Humphreys, 2004).

In a preliminary demonstration utilizing the gamete selection approach, 14 IL x tall fescue hybrids were evaluated for their crude protein (CP) quantity via available nitrogen in plant leaf tissue. The need for forages to possess adequate levels of CP must complement the grazing animal's nutritional requirements. Typically, CP estimates reflect the level of nitrogen and amino acids in forages. CP content is considered a quantitative trait in most forages (Fei et al., 2006.) and is typical of a trait exhibiting low heritability and is, as a consequence, difficult to transfer rapidly and effectively (Vogel et al., 1981). Leaf samples

from 14 IL x tall fescue F1 hybrids were obtained from the nursery in March 2011 and were run on an Elementar varioMacro flash combustion instrument (Elementar Americas, Inc., Mt. Laurel, NJ). In addition, DH lines derived from particular IL x tall fescue hybrids were assayed for N, along with the check tall fescue cultivars Nanyro, Retu, Drover and Barcarella. CP concentration of samples was estimated by multiplying total nitrogen (N) concentration by 6.25 (Hersom, 2007)(Table 1). Multiplying the N concentration by 6.25 to estimate CP level is performed because protein molecules contain an average of 16% N (1/16 = 6.25). This N and CP estimation represents the standard approach utilized for evaluating beef cattle protein requirements (Hersom, 2007). Results presented in Table 1 indicate that hybrids high in CP content provide DH recoveries high in CP content. DH23A, a DH exhibiting low CP content, was obtained from a hybrid exhibiting low CP content. A

IL x Tall Fescue	Crude Protein	Description
LF12-8	18.5	IL X TF Hybrid
LF12-9	10.6	IL X TF Hybrid
LF13-1	14.3	IL X TF Hybrid
LF13-3	11.7	IL X TF Hybrid
LF14-1	17.9	IL X TF Hybrid
LF14-2	13.1	IL X TF Hybrid
LF15-2	17.1	IL X TF Hybrid
LF16-1	11.7	IL X TF Hybrid
LF16-2	16.2	IL X TF Hybrid
LF16-3	14.4	IL X TF Hybrid
LF800-4	15.6	IL X TF Hybrid
LF800-8	12.0	IL X TF Hybrid
LF800-9	19.4	IL X TF Hybrid
LF900-1	19.2	IL X TF Hybrid
Syn1	21.0	Blend of DH lines
DH8B	18.7	LF12-8 recovery
DH9B	18.4	LF12-8 recovery
DH36D	21.1	LF900-1 recovery
DH15B	19.2	L F800-9 recovery
DH14B	18.6	LF14-1 recovery
DH23A	12.2	LF16-1 recovery
Nanyro	21.4	Cultivar Check
Drover	19.8	Cultivar Check
Barcarella	15.5	Cultivar Check
Retu	18.4	Cultivar Check

Table 1. Crude protein estimates across IL x tall fescue hybrids (LF), their dihaploid recoveries (DH) and standard cultivar checks.

program focused on developing a high CP cultivar would simply require the identification of agronomically superior DH lines exhibiting high CP estimations. This approach is superior to traditional recurrent selection as there are no segregating alleles that could confer lower or segregating CP levels. All lines utilized in the breeding process are fixed for the high or low CP genes. This approach removes much of the random segregation of alleles governing CP or other traits of low heritability. Following identification of high CP materials, eight high CP DH lines were combined to generate an experimental synthetic (Syn1). In limited performance trials, Syn1 has demonstrated itself to be a DH synthetic exhibiting high CP and superior agronomic attributes. As a consequence, a selection program focused on identifying hybrid genotypes exhibiting high CP content, then deriving high CP content DH lines from those hybrids should be an efficient and effective method to concentrate quantitative trait loci defining elevated CP content via gamete selection.

4. Example 2: Gamete selection for stress tolerance

A primary objective of the forage research program at the USDA-ARS, Grazinglands Research Laboratory is the development of tall fescue forage possessing tolerance to the environmental extremes of the Southern Plains Region. Typically, these environmental extremes involve heat, drought and low nitrogen inputs. To achieve this goal, a three year program of natural selection was conducted with numerous IL x tall fescue hybrids. Selection criteria were to evaluate the hybrids in a high stress environment on the Southern and Central Plains of the USA consisting of high summer temperatures, high wind, non-irrigated, low nitrogen input (40 lbs/ac) conditions (Figures 1a, 1b, 2a, 2b). Trials were conducted at the USDA-ARS, Grazinglands Research Laboratory, El Reno, OK and at the Mound Valley Research Unit, Kansas State University, Southeast Agricultural Research Center, Parsons, KS. Seed generated from the initial IL x tall fescue crosses were sown in a greenhouse and individuals were transferred to spaced plot field selection nurseries in the fall of 2006 (El Reno, OK) and 2008 (Mound Valley, KS). IL x tall fescue hybrids that did not possess a genotype adaptable to these conditions either died or, if the individual hybrids exhibited unsatisfactory agronomics, were physically removed from the nursery on a year to year basis (Figure 4). In the fall of 2009, seed heads were removed from the El Reno, OK nursery and in the fall of 2010, seed heads were removed from the surviving IL x tall fescue hybrids at the Mound Valley, KS location. Seed heads were threshed as described and sown to germination trays in the greenhouse.

Recovered DH tall fescue lines were removed from the germinating trays and transplanted to pots. In some instances, DH lines were selfed for seed increase. Though tall fescue is considered an obligate out-crossing species we have observed that this is not the case for many recovered DH lines and successful selfing of tall fescue has been observed and utilized in prior tall fescue research programs (Buckner and Fergus, 1960a). Seed generated through selfing recovered DH lines were sown in small, unreplicated plots in the nurseries at the USDA-ARS, Grazinglands Research Laboratory, El Reno, OK. Single DH plant selections were transplanted to the Kansas State University, Southeast Agricultural Research Center, Parsons Unit, Mound Valley, KS in 2010. The selection criteria remained unchanged at both locations for DH evaluation. During the summer of 2011 both locations experienced extended severe to extreme drought and heat conditions (National Oceanic and Atmospheric Administration, National Climatic Data Center, 2011). The DH tall fescue

recoveries have exhibited good to excellent tolerance, superior adaptation to and persistence under these record-setting drought conditions at both locations. This study indicates that the gamete selection approach is effective and efficient in identifying genotypes that can tolerate high environmental stress conditions such as drought, heat and low nitrogen inputs. These and additional DH lines generated from this gamete selection approach will be utilized to form a foundation of tall fescue germplasm with particular adaptation to high stress sites across the Southern Plains and Midwest regions of the USA.

Fig. 4. Example of a stress intolerant IL x tall fescue F1 hybrid (left) and a stress tolerant hybrid (right).

5. Confirming recovery of DH tall fescue lines

There are two methods that can be performed to indicate or verify that the recovered tall fescue materials are DH homozygous recoveries. The first step is to self-pollinate the particular DH individual for seed increase and perform nursery grow outs. If the offspring from the selfing do not segregate for size, inflorescence morphology, maturity or other obvious phenotypic characteristics, they are likely DHs (Figure 5).

A second approach is to utilize molecular markers known to exhibit a banding pattern that is consistent with disomic inheritance and co-dominant expression. That is, each marker exhibits a maximum of two bands for each individual when two alleles at a particular locus are present in a heterozygous state. When one band is present, the allele at that locus would be in a homozygous condition. These markers will be most useful since segregation will be predictable and such markers are easy to score (Stift et al., 2008). Tall fescue, having a polyploid genome constitution (Alderson and Sharp, 1994), has a considerable level of genome duplication. As a consequence, it is important to utilize markers that are known to amplify only a single or specific site in the tall fescue genome. When such markers are utilized, only a single amplification product will be formed at that locus since that locus would be homozygous. If the site were heterozygous, the amplification of multiple alleles would be visualized. Following the approach of Saha et al. (2004), selected DH lines were evaluated utilizing a set of twelve EST-SSR markers with the analyses being performed by DNA LandMarks (Quebec, Canada). The results of the analysis indicated that DH47

exhibited a high level of allelic variability when compared to the other DH recoveries (Table 2). From this analysis, DH47 is indicated to be non-homologous for all marker loci and is likely a rogue. DH lines 41, 42, 44 and 57 are indicated to be true DH recoveries, homozygous at the evaluated marker loci. Similar approaches may be equally effective in polyploids thought to possess a high degree of genome duplication (Esselink et al., 2004).

Fig. 5. A plot in a DH phenotypic uniformity trial at the USDA-ARS, Grazinglands Research Laboratory.

EST-SSR	DH41	DH42	DH44	DH47	DH57
NFA004	285	285	285	298	285
NFA030	199	199	199	199	199
NFA039	293	293	293	291	293
NFA047	244	244	244	241/244/250	244
NFA067	130	151	175	147	136
NFA068	254	254	254	251	254
NFA073	236	236	246	236	246
NFA074	244	244	244	241/244/250	244
NFA095	187	187	187	182/187	187
NFA104	218	218	218	238	218
NFA115	236	236	236	220/230	236
NFA133	150	150	150	145/155	150

Table 2. Base-pair marker size segregation of twelve EST-SSR markers across five potential dihaploid lines. The absence of allelic variation across markers indicates a dihaploid recovery, homozygous for the locus (e.g. DH41, DH42, DH44, DH57). An excessive level of allelic marker variability suggests a non-homozygous condition and the lack of a homozygous DH recovery (DH47). The DH47 line would be discarded.

6. Conclusion

The generation of haploid or dihaploid lines through a gamete selection approach can be of critical importance in breeding and genetic analysis research (Duwell, 2010). The materials and procedures described above apply directly to the breeding and selection of superior tall fescue germplasm and could be expanded to other fescue species. The expansion of research programs utilizing this approach should find the availability of DH lines useful when objectives are focused on plant genomics, fine mapping of DH derived populations and/or marker-assisted selection of genes that confer agronomically favorable traits. One advantage to the described method is that microspore methods are not utilized and the researcher is not limited to genotypes amenable to microspore culture techniques. The sampling of hundreds of thousands of pollen grains, each segregating for a myriad of genotypes from a single tall fescue individual or population requires less labor input and represents a low cost, rapid selection strategy that can be implemented across diverse environments. The simultaneous selection of multiple traits or complementary traits can be applied quite effectively and the prescribed approach also does not require any prior genetic information regarding the inheritance or expression of the quantitative trait. Though molecular markers can be applied, the approach does not require the utilization of molecular markers for marker-assisted selection. Generated DH lines possessing superior genotypes are maintained through selfing, retaining the fixation of the qualitative or quantitative trait of interest.

Though IL1 and IL2 can be utilized to generate DH lines, the frequency of generation is low, less than 1%. However, the ability to generate IL x tall fescue hybrids is rapid and inflorescences on the hybrids are abundant. When numerous F1 inflorescences are utilized from the F1, the recovery of DH lines is quite efficient. Depending on the quality and number of inflorescences, it is not unusual to obtain one to eight seedlings from each IL x tall fescue hybrid. When multiple hybrids are screened and placed in a commercial line production situation, hundreds of DH recoveries may be obtained each season. What the approach may lack in DH generation frequency, is compensated by its efficiency in recovering DH lines. It is anticipated that when applied correctly, this approach would be effective for development of tall fescue lines exhibiting quality traits such as days to flowering, drought tolerance, grazing persistence, CP, NDF, ADF, RFQ, lignin, forage yield, etc. The gamete selection approach should also be advantageous to tall fescue breeders requiring efficient and rapid methods of developing tall fescue germplasm adapted to the unknown parameters of global climate change (Humphreys et al, 2006).

7. References

Alderson, J. & Sharp, W. C. (1994). Grass varieties in the United States. Agric. Handbook 170. Soil Conserv. Serv., USDA. Washington, D.C.

Bai, Y. & Qu, R. (2000) An evaluation of callus induction and plant regeneration in twenty-five turf type tall fescue (*Festuca arundinaceae* Schreb.) cultivars. Grass and Forage Science 55:326-330, ISSN 0142-5242.

Barclay, I.R. (1975). High frequencies of haploid production in wheat (*Triticum aestivum*) by chromosome elimination. Nature 256:410–411, ISSN 0028-0836.

Bouchez, A. & Gallais, A. (2000). Efficiency of the use of doubled-haploids in recurrent selection for combining ability. Crop Sci. 40(1): 23-29, ISSN 1435-0653.

Bouton, J. H. (2009) Molecular breeding to improve forages for use in animal and biofuel production systems. In*: Molecular Breeding of Forage and Turf, The proceedings of the*

5th International Symposium on the Molecular Breeding of Forage and Turf. T. Yamada and G. Spangenberg (Eds.), 1-13, Springer, ISBN 978-0-387-79144-9, New York , NY.

Brummer E. C. & Casler, M. D. (2009). Improving selection for forage, turf and biomass crops using molecular markers. In*: Molecular Breeding of Forage and Turf, The proceedings of the 5th International Symposium on the Molecular Breeding of Forage and Turf* . T. Yamada and G. Spangenberg (Eds.), 193-209, Springer, ISBN 978-0-387-79144-9, New York , NY.

Buckner, R. C. & Fergus, E. N. (1960a). Improvement of tall fescue for palatability by selection within inbred lines. Agron. J. 52:173–176, ISSN 1435-0645.

Buckner, R. C. (1960b). Cross compatibility of annual and perennial ryegrass with tall fescue. Agron J. 52:409-410, ISSN 1435-0645.

Buckner, R. C. Hill, H. D. & Burrus P. B. (1961) Some characteristics of perennial and annual ryegrass x tall fescue hybrids and of the amphidiploid progenies of annual ryegrass x tall fescue. Crop Sci. 1:75-80, ISSN 1435-0653.

Cardinal, A. J.; Lee, M., & Moore, K. J. (2003). Genetic mapping and analysis of quantitative trait loci affecting fiber and lignin content in maize. TAG 106:866–874, ISSN 0040-5752.

Carroll, G. C. (1988). Fungal endophytes in stems and leaves: From latent pathogen to mutualistic symbiont. Ecology, 692-699, ISSN 1365-2745.

Casler, M. D. (2001). Breeding forage crops for increased nutritive value. Adv Agron 71:51–107, ISSN 0065-2113.

Chang, M.T. & Coe, E.H. (2009) Doubled haploids. In: *Molecular Genetic Approaches to Maize Improvement. Biotechnology in Agriculture and Forestry*, Vol. 63 A. L. Kriz and B. A. Larkins (Eds.), 127–142, Springer, ISBN 978-3-642-08844-5, Berlin, Germany.

Cogan, N.; Smith, K. F., Yamada, T., Francki, M. G., Vecchies, A. C., Jones, E. S., Spangenberg, G. C. & Forster, J. W. (2005). QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (*Lolium perenne* L.). TAG 110:364–380, ISSN 0040-5752.

Dunwell, J. W. (2010). Haploids in flowring plants: Origins and exploitation. Plt. Biotech. J. 8:377-424, ISSN 1467-7644.

Esselink, G. D., Nybom, H. & Vosman, B. (2004). Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting – peak ratios) method. TAG 109:402-408, ISSN 0040-5752.

Fei, E.; Brummer, C., Moore, K. J., Barker, R. E., Curley, G. & Warnke, S. (2006). QTL analyses of fiber components and crude protein in an annual - perennial ryegrass interspecific hybrid population. Mol Breeding 18:327-340, ISSN 1923-8266.

Gallais, A. (1993). Efficiency of recurrent selection methods to improve the line value of a population. Plt. Breeding 111:31-41, ISSN 1439-0523.

Gordilli, G. A. & Geiger, H. H. (2008). Alternative recurrent selection strategies using double haploid lines in hybrid maize breeding. Crop Sci. 48:911–922, ISSN 1435-0653.

Guo, Y. & Yamada, T. (2004). High frequency androgenic embryogenesis and genome size variation in pollen-derived Festulolium. In*: Development of a novel grass with environmental stress tolerance and high forage quality through intergeneric hybridization between Lolium and Festuca*. T. Yamada and T. Takamizo (Eds.), 49–54, Natl. Agric. Bio-oriented Res. Org. (NARO), Ibaraki, Japan.

Guo, Y. , Mizukami, Y. & Yamada, T. (2005). Genetic characterization of androgenic progeny derived from *Lolium perenne* × *Festuca pratensis* cultivars. New Phytologist 166:455-464, ISSN 1469-8137.

Henson, L. & Buckner, R. C. (1957). Resistance to *Helminthosporium dictyoides* in inbred lines of *Festuca arundinacea*. Phytopathology 47:523, ISSN 0031-949X.

Hersom, M. (2007). Basic Nutrient Requirements of Beef Cows. University of Florida IFAS Extension Publication #AN190.

Humphreys, M. W.; Canter, P. J. & Thomas, H. M. (2003). Advances in introgression technologies for precision breeding within the *Lolium-Festuca* complex. Ann. Appl. Biol. 143:1–10, ISSN 0003-4746.

Humphreys, M. W. (2004). Development of a novel grass with resistance to environmental stress and high forage quality through intergeneric hybridisation between Lolium and Festuca. In: *Development of a novel grass with resistance to environmental stress and high forage quality through intergeneric hybridisation between Lolium and Festuca*. T. Yamada and T. Takamizo (Eds.), 7-15, National Agric. and Bio-oriented Research Organization (NARO), Ibaraki, Japan,

Humphreys, M. W., Yadav, R. S., Cairns, A. J., Turner, L. B. and J. Humphreys, L. Skøt. (2006). A changing climate for grassland research. New Phytologist 169:9-26, ISSN 1469-8137.

Inagaki M.N. (1997). Technical advances in wheat haploid production using ultra-wide crosses. JIRCAS J 4:51–62, ISSN 35510021-.

Jennings, J. A., Beck, P. & West C. (2008). Tall fescue for forage. University of Arkansas Extension Bulletin, Dept. of Agricultural and Natural Resources. FSA 2133 -PD-1-08RV.

Kasha, K. J. & K. N. Kao, N. (1970). High Frequency Haploid Production in Barley (*Hordeum vulgare* L.). Nature 225:874-876, ISSN 0028-0836.

Kasperbauer, M. J.; Buckner, R. C. & Springer, W. D. (1980). Haploid plants by anther-panicle culture of tall fescue. Crop Sci. 20:103-107, ISSN 1435-0653.

Kendall, J. (1934) A parthenogenetic aberrant tobacco plant. J. Heredity 31:363-366, ISSN 1465-7333.

Kindiger, B. (2009). A *Lolium multiflorum* line inducing genome loss. U.S. patent application 2010083400.

Kindiger, B. & Singh, D. (2011). Registration of Annual Ryegrass Genetic Stock IL2. J. of Plant Reg. 5:254-256, ISSN 1936-5209.

Kindiger, B. (2012). Registration of Annual Ryegrass Genetic Stock IL1. J. of Plant Reg. 6(1): in press), ISSN 1936-5209.

Kopecký, D., Bartoš, J., Zwierzykowski, Z., & Doležel, J. (2009). Chromosome pairing of individual genomes in tall fescue (*Festuca arundinacea* Schreb.), its progenitors, and hybrids with Italian ryegrass (*Lolium multiflorum* Lam.). Cytogenet. Genome Res. 124:170–178, ISSN 0967-3849.

Lu, C.; Shen, L., Tan, Z., Xu, Y., He, P., Chen, Y. & Zhu, L. (1996). Comparative mapping of QTL for agronomic traits in rice across environments using a doubled haploid population. TAG 93:1211–1217, ISSN 0040-5752.

Maheshwari, S. C.; Tyagi, A. K. & Malhotra, K. (1980). Haploids from pollen grains–Retrospect and prospect. Am. J. Bot. 69:865–879, ISSN 0002-9122.

Mohan J. S., Sopory, S. K. & Veilleux, R. E. (1996). *In Vitro Haploid Production in Higher Plants – Volume 4: Cereals*. Kluwer Academic Publishers. ISBN 0-7923-3978-9. Dordrecht, The Netherlands.

Moon, C. D., Tapper, B. A. & Scott, B. (1994). Identification of Epichloe¨ Endophytes in planta by a microsatellite-based PCR fingerprinting assay with automated analysis. Appl. Environ. Micro. 65:1268-1279, ISSN 0099-2240.

Mulcahy D. L. (1971) A correlation between gametophytic characters in Zea mays L. Science 171:1155-1156, ISSN 0036-8075.

Naganowska, B.; Zwierzykowski, Z. & Zwierzykowska, E. (2001). Meiosis and fertility of reciprocal triploid hybrids of *Lolium multiflorum* with *Festuca pratensis*. J. Appl. Genet. 42(3):247-55, ISSN 2190-3883.

National Oceanic and Atmospheric Administration, National Climate Data Center (2011). State of the Climate Drought 2011. http://www.ncdc.noaa.gov/sotc/drought/

Pedersen, J. F. & Sleper, D. A. (1988). Considerations in breeding endophyte-free tall fescue forage cultivars. J. Prod. Agric. 1:127-132, ISSN 0890-8524.

Ravikumar, R. L. & Patil, B. S. (2004). Effect of gamete selection on segregation of wilt susceptibility-linked DNA marker in chickpea. Cur. Sci. 86:642-643, ISSN 0011-3891.

Richey, F. D. (1947) Corn breeding: Gamete selection, the Oenothera method and related miscellany. J. Amer. Soc. Agron. 39:403-411, ISSN 0002-1962.

Riera-Lizarazu, O.; Peterson, C. J. & Leonard, J. M. (2010). Registration of the OS9XQ36 mapping population of wheat (*Triticum aestivum* L.). J. Plant Reg. 4:98–102, ISSN 1936-5209.

Rotarenco, V. A. & Chalyk, S. T. (2000). Selection at the level of haploid sporophyte and its influence on the traits of diploid plants in maize. Genetika 32:479-485 ISSN 0534-0012.

Saha, M.C.; Mian, M. A. R.; Eujayl, I; Zwonitzer, J. C.; Wang, L. & May, G.D. (2004). Tall fescue EST-SSR markers with transferability across several grass species. TAG 109:783–791, ISSN 0040-5752.

Shugar, L. P. (1989). Commercialization of the bulbosum method in barley (Horderum vulgare L.). In: *Current Options for Cereal Improvement*. M. Malusynski (Ed.), pp. 161-179. Kluewer Academic Publisher, ISBN 0-7923-0064-5. Dordrecht, The Netherlands.

Singh, S. P. (1994). Gamete selection for simultaneous improvement of multiple traits in common bean. Crop Sci. 34:352–355, ISSN 1435-0653 .

Singh, S.P.; Cardona, C.; Morales, F.J.; Pastor-Corrales, M.A. & Voysest, O. (1998). Gamete selection for upright carioca bean with resistance to five diseases and a leafhopper. Crop Sci. 38(3):666-672, ISSN 1435-0653 .

Stadler, L. J. (1944). Gamete selection in corn breeding. Jour. Amer. Soc. Agron. 36:988-989, ISSN 0002-1962.

Stift, M., Berenos, C., Kuperus, P., & van Tienderen, P. (2008). Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: A general procedure applied to Rorippa (Yellow Cress) microsatellite data. Genetics 179:2113-2123, ISSN 0016-6731.

Tuvesson, S; Ljungberg, A; Johansson, N.; Karlsson, K. E.; Suijs, L. W. & Josset J. P. (2008). Large-scale production of wheat and triticale double haploids through the use of a single-anther culture method. Plant Breeding 119:455-459, ISSN 1439-0523.

Vagera, J.; Fojtík, A. & Ohnoutková, L. (1998). The induced androgenesis and polyhaploids production in breeding material from Lolium and Festuca species and Festulolium hybrids. - In: *Proceedings of the Conference "Anther and Pollen: from Biology to Biotechnology*". Université de Reims, Francie, April 9-10, 1998, 51, 1998. (in English).

Vogel, K. P.; Gorz, H. J. & Haskins F. A. (1981). Heritability estimates for forage yield, in vitro dry matter digestibility, crude protein and heading date in indiangrass. Crop Sci. 21:35-38, ISSN 1435-0653.

Yamada, T. & Takamizo T. (2004). Development of a novel grass with environmental stress tolerance and high forage quality through intergeneric hybridizaion between Lolium and Festuca. T. Yamada and T. Takamizo (Eds.), National Agric. and Bio-oriented Research Organization, Japan (NARO), Ibaraki, Japan.

15

Genetic Diversity of Rice Grain Quality

Rosa Paula Cuevas and Melissa A. Fitzgerald
Grain Quality and Nutrition Centre, International Rice Research Institute
Philippines

1. Introduction

Rice (*Oryza* spp.) is one of the most important food crops in the world, being planted on almost 11% of the Earth's cultivated land area over a wide number of ecosystems (Khush, 2005; Maclean et al., 2002). Two species, *Oryza sativa* and *O. glaberrima*, are cultivated while other species are wild. Human selection and environmental factors have contributed to the genetic diversity in rice, particularly in *O. sativa* cultivars (Maclean et al., 2002). A significant amount of this genetic diversity is housed in rice gene banks around the world, with a copy also conserved in the snowy depths of Svalbard, Norway. The diversity of rice offers a valuable resource to understand grain quality and how different agronomic backgrounds alter those traits. Rice is consumed mainly as milled, white grains or as brown grains (unpolished), and also as ingredients in food products. The cooking and sensory properties of a variety are key components that affect its acceptability to consumers (Cramer et al., 1993). Consumer preferences shift from low-quality to high-quality rice with increased income and market liberalisation (Cramer et al., 1993; Dawe & Slayton, 2004). Improvements in post-harvest technologies have also contributed to this shift in consumer preference by decreasing the price difference between low- and high-quality rice (Dawe & Slayton, 2004).

With the exponential and breathtaking progress in sequencing and genotyping technologies this decade is bearing witness to in rice, research programs delivering to breeding programs should perhaps focus on understanding the genetic control of different traits of eating quality, and offering genetic markers to breeding programs for inclusion on the new generation chips being developed (Boualaphanh et al., 2011). However, understanding the genetics of eating quality is difficult because consumers are not easily able to describe the sensory experience. Without a way to measure the trait, it is not possible to find an associated locus.

Thus, in order to assess consumer acceptability of breeding materials and rice cultivars, and in the absence of a clear knowledge of sensory properties, quality evaluation and breeding programs rely heavily on three indirect indicators that predict, to some degree, the cooking and sensory properties of rice: amylose content, gelatinisation temperature, and gel consistency. An understanding of the genetics behind these quality indicators can aid in screening early in breeding programs. However, as research has progressed into the genetics of those indicators, the data reveal that one major gene can be found for each trait, but the genes underlying the finer phenotypic classifications for each of those traits remain undiscovered. In this chapter, the genetic diversity of rice for these three indicators of quality and future avenues for research are reviewed.

2. Amylose content

Amylose content (AC) is regarded as the most important indicator in classifying rice varieties (Juliano, 1985) because it influences texture and retrogradation potential of cooked grains (Champagne et al., 2004; Ong & Blanshard, 1995a). It is measured in breeding programs as soon as heterozygosity is minimised and is the first tool used in the selection process for eating quality. Rice varieties are classified into high (>25%), intermediate (20 – 25%), low (10 – 19%), very low (3 – 9%), or waxy (0 – 2%) amylose classes (Kumar & Khush, 1987).

Since waxy rice varieties do not have amylose molecules, they are of particular interest in research because they provide a simple way to unravel the effects of components of the rice grain, other than amylose, that affect cooking and eating properties. However, waxy rice varieties have been reported to have very low amounts of amylose (i.e., Chung et al., 2011; Juliano, 1971; Landers et al., 1991; Sanchez et al., 1988; Varavinit et al., 2003). In an international cooperative test conducted recently, waxy rice samples were reported to have as high as 11% amylose using the iodine colorimetric method (Fitzgerald et al., 2009a). Reports have also indicated the presence of amylose chains within the hilum of immature waxy grains based on the colour reaction with iodine (Badenhuizen, 1956; Baker & Whelan, 1951; Juliano & Villareal, 1987). The difference between the expected amount and the observed amylose values indicate either (1) amylose is truly present in waxy rice varieties or (2) the values obtained by iodine colorimetry stemmed from technical issues relating to the method. Therefore, there is a need to further define the waxy phenotype. In this review, the definition of amylose is based on its genetics, its structure, and its effect on the functional properties of rice.

In order to clarify whether amylose is truly present in the waxy mutants, 53 traditional Lao waxy varieties and IR29 (a waxy mutant) were used along with 75 non-waxy varieties. These non-waxy varieties consisted of 8 very low AC, 7 low AC, 19 intermediate AC, and 42 high AC rice varieties. The set also included milled rice flour of Calmochi-101 (waxy), Calamylow-201 and BL-2 (opaque mutants of Calhikari-201), BR-2, BR-5, and BR-7 (waxy mutants of Calhikari-201) kindly provided by Dr McKenzie (California Cooperative Rice Research Foundation, Inc) (McKenzie et al., 2006a, 2006b) and RS111, a high-amylose content mutant (37%) of a Chinese variety (Shu et al., 2006a) provided by Dr Wu Dianxin (Zhezhiang University, Hangzhou, China). These varieties were tested in many different ways to determine whether or not amylose was present (Cuevas, 2009).

2.1 Genetic definition of amylose

The genetic basis of AC has been studied extensively. The major gene responsible for amylose synthesis in rice is the *Waxy* gene on chromosome 6 (Nagao & Takahashi, 1963; Sano, 1984). Amylose classes associate with polymorphisms in the *Waxy* gene (*Wx*) (Table 1). The functional alleles Wx^a and Wx^b, caused by a single base change at the 5′ splice site of intron 1, distinguish low amylose contents from high and intermediate classes (Hirano et al., 1998; Wang et al., 1995). The G→T mutation in the splice site of intron 1 causes inefficient transcription because of alternate splicing (Bligh et al., 1998; Cai et al., 1998; Isshiki et al., 1998). However, these two alleles are not enough to explain the continuous variation in AC across the different classes nor do they define the waxy class. A single nucleotide polymorphism

(SNP) in exon 6 (A→C), *Wx*^{in}, was associated with intermediate AC (Chen et al., 2008; Larkin & Park, 2003; Mikami et al., 2008). Seeds that are opaque despite the production of amylose, *Wx*^{op} (or *Wx*^{hp}), carry the same SNP in the intron 1 splice site as *Wx*^{a} and an A→G SNP on exon 4 (Liu et al., 2009; Mikami et al., 1999). Further, premature termination of translation caused by a 23-bp duplication in exon 2 of *Wx* causes a shift in the reading frame; this leads to a premature stop codon and the waxy phenotype (Mikami et al., 2008; Wanchana et al., 2003).

Haplotype	Intron 1	Exon 2 (23-bp dupl'n)	Exon 4	Exon 6	Phenotype
1	T[a]	+			Waxy
2	T	-			Low AC
3	G	-	G		Very low AC
4	G	-	A	C	Intermediate AC
5	G	-	A	A	High AC

[a] Waxy rice predominantly carry the *Wx*^{b} allele; however, there are waxy rice varieties that do carry the *Wx*^{a} allele (Yamanaka et al., 2004)

Table 1. Known haplotypes of rice based on mutations in the *Waxy* gene

In rice, Granule-Bound Starch Synthase I (GBSSI) enzyme, the product of the *Wx* gene, is expressed in the endosperm and encodes a 66-kDa protein (Hirose & Terao, 2004). The SNPs defining each haplotype are at important sites for gene expression, therefore each SNP lowers the amount of GBSSI protein produced (Figure 1).

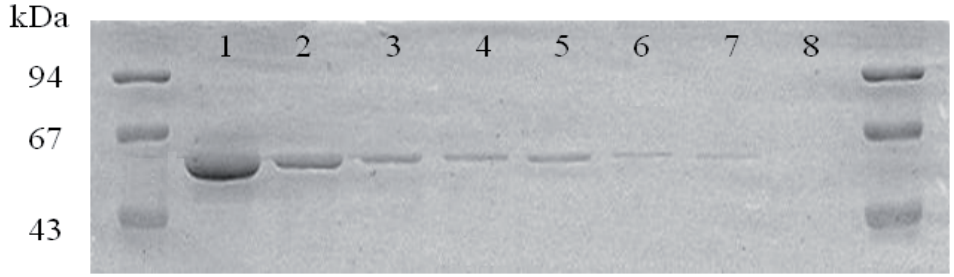


Fig. 1. Comparison of GBSS1 bands in rice varieties with decreasing amounts of amylose: (1) RS 111, (2) IRGC111746 B, (3) IR8, (4) IR64, (5) Calamylow-201, (6) Calhikari-201, (7) BL-2, and (8) Calmochi-101. The ladder is composed of phosphorylase B (94 kDa), albumin (67 kDa), and ovalbumin (43 kDa).

In contrast to non-waxy varieties, Calmochi-101, a waxy rice variety, does not express GBSSI (Figure 1). Bands of GBSSI were also not observed in IR65, a waxy variety used as a standard, nor in the set of 54 waxy varieties (Cuevas, 2009). Hence, it can be concluded that the waxy mutants of rice do not express GBSSI protein.

2.2 The molecules of amylose

GBSSI synthesises amylose molecules in storage organs (Ball & Morell, 2003; Sano & Katsumata, 1984; Smith et al., 1997). Amylose is a linear (sparsely branched) polymer of glucose units linked by α-1,4 linkages (Ball et al., 1996), that reaches up to chain lengths

ranging from 200 to 10,000 degrees of polymerisation (Hizukuri et al., 1989; Takeda et al., 1986; Takeda et al., 1992a, 1992b; Ward et al., 2006) and with molecular weights in the 10^5 range (Ebermann & Praznik, 1975; Roger & Colonna, 1996; Shelton & Lee, 2000). These molecules are distinct from amylopectin molecules, the hyperbranched glucose homopolymer of starch.

Amylose chains form helical coils when suspended in water. Iodine then enters these coils, subsequently forming stable amylose-iodine complexes that are characterised by a blue colour reaction (Rundle et al., 1944). Thus, amylose content in rice varieties can be quantified based on the intensity of the color reaction with iodine. In a previous study (Cuevas, 2009), amylose content of the 54 waxy rice varieties was measured using the a modified method based on the ISO 6647 routine method (International Organization for Standardization, 2007), and the method returned positive values (Table 2).

	Amylose Content (%)
Maximum	8.15
Minimum	3.88
Mean	6.21
Standard Deviation	1.12

Table 2. Range of amylose content values of the waxy rice varieties following modified ISO 6647, using non-waxy standards with known amylose contents (Cuevas, 2009).

The range of amylose contents for the waxy varieties measured in Table 2 fell within the very low amylose class (Kumar & Khush, 1987). Based on these results, it is easy to conclude that waxy rice grains do have some amount of amylose molecules despite the absence of GBSSI expression within the grains of these varieties. It is also easy to classify very low amylose varieties as waxy, or group these two types into one class (Varavinit et al., 2003). However, the results in Table 2 were obtained using calculations based on standard curves that might not be chemically appropriate: 0% amylose was extrapolated from absorbance readings of non-waxy materials, but a waxy variety for the zero reading is more appropriate and delivers matrix control. A pure amylopectin matrix represents 0% amylose and waxy rice is recommended to be this matrix in the AACC Method 61-03 (American Association of Cereal Chemists, 2000).

The difference in structure between the chains that constitute amylose and amylopectin molecules has allowed for their separation by size exclusion chromatography (SEC) after debranching the molecules to obtain the chains from each. Elution volumes of the debranched molecules of the two polymers have been identified: the longer amylose molecules elute first followed by the shorter amylopectin molecules (Ward et al., 2006). Through universal calibration and the use of pullulan standards, the elution volumes could be converted into chain lengths (or degree of polymerisation, X) (Ward et al., 2006). Thus, the proportion of amylose relative to amylopectin can be determined based on the area under the two peaks.

As shown in Figure 2, the waxy variety, Calmochi-101, does not show a peak in the chain-length area where amylose elutes. This is expected because GBSSI is not expressed in this variety (Figure 2). When SEC traces of varieties spanning the range of amylose classes are

compared (those in Figure 1), the area under the amylose peak (X 200 – 10000) increases while the area of the amylopectin peak (X < 200) decreases with increasing amylose content and GBSSI expression (Figure 1). In non-waxy varieties, the amylose peak does not return to the baseline between the amylose and the amylopectin peaks (Figure 2). The distance from the baseline increased with increasing amylose content, indicating that GBSSI synthesises chains between the amylose and the amylopectin chain-length ranges.

Calamylow-201, the opaque mutant, was shown to express GBSSI, albeit weakly (Figure 1). The debranched chain-length distributions of this variety and of BL-2, another opaque mutant, confirm the presence of low amounts of amylose inside the grain (Figure 2) despite the resemblance in physical appearance with waxy grains. Moreover, these mutants appear to fall into haplotype 3 of the *Wx* gene (Table 1), carrying the mutation on exon 4, based on their AC.

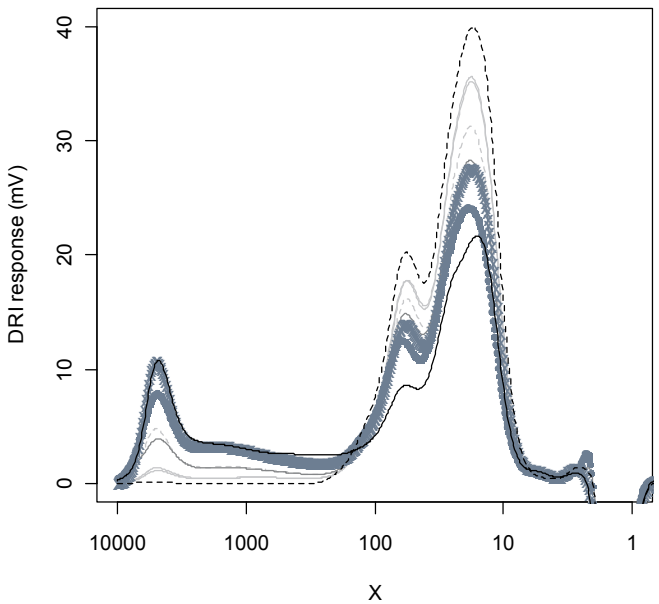


Fig. 2. Debranched chain-length distributions of waxy and non-waxy rice varieties. Calmochi-101 (---); BL-2 and Calamylow-201 (—); Calhikari-201 (---); IR24 (—); IR8 (•); IRGC111746B (x); and RS 111 (—) (Cuevas, 2009).

It has previously been reported that immature waxy rice grains contain amylose chains in the hilum, as indicated by distinctive colour reactions with iodine (Badenhuizen, 1956; Baker & Whelan, 1951; Juliano et al., 1987). The implication is that GBSSI is active at one point in grain-filling, and that waxy rice really does contain amylose, the product of that enzyme. However, genetics and SEC profiles indicate that these linear glucose chains could not be amylose, defined as the product of the *Wx* gene, despite the blue reaction, because of the frame-shift mutation in exon 2 (Mikami et al., 2008; Wanchana et al., 2003) and the absence of a peak consistent with amylose in the mature varieties (Figure 2). To clarify whether amylose exists during grain-filling in a waxy variety, starch from immature rice grains of Thassano 1 (TSN1), a waxy rice variety from Lao PDR, was debranched and the molecules were separated by SEC (Cuevas, 2009). Results showed quite clearly that chains normally assigned as amylose were absent and only amylopectin chains were detected even in the immature stage (Figure 3).

The enzymes that are involved in starch synthesis are reported to have temporal patterns of expression. GBSSI is expressed quite late in the development of the grain, beginning to be detectable five days after flowering (Hirose & Terao, 2004). The iodine reaction then implies that some other enzyme synthesises glucose chains that mimic the response of amylose. There are many possible starch synthases that could produce such chains in immature grains. Early in grain-filling, genes coding enzymes involved in amylopectin synthesis are highly expressed. Starch synthases (SS) II-2 and III-1 are specifically expressed 1 – 5 days after flowering while SSI, II-1, IV-1 and IV-2 are expressed constantly during grain filling (Hirose & Terao, 2004).

The absence of amylose-length chains in immature grains of TSN1 (Figure 3) indicates that GBSSI is indeed absent from waxy grains. The blue core previously observed in the hilum of immature waxy grains (Baker & Whelan, 1951; Badenhuizen, 1956; Juliano et al., 1987) must have been caused by the presence of 'atypical' amylopectin molecules that form complexes with iodine (Buleon et al., 1998). Iodine forms a blue reaction with chains $X \geq 45$ (Bailey & Whelan, 1961); the steady expresser SSIV is associated with the synthesis of chain lengths in this range and is associated to the production of such chains in the hilum (Hirose & Terao, 2004; Roldan et al., 2007). Mutants of *Arabidopsis* that do not express SSIV do not synthesise starch (Roldan et al., 2007). This suggests therefore, that SSIV is implicated in the synthesis of the material that seeds the building of the amyloplast.

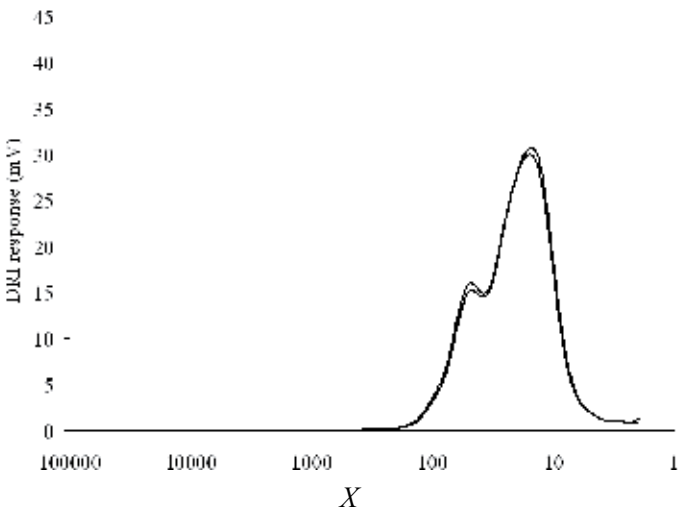
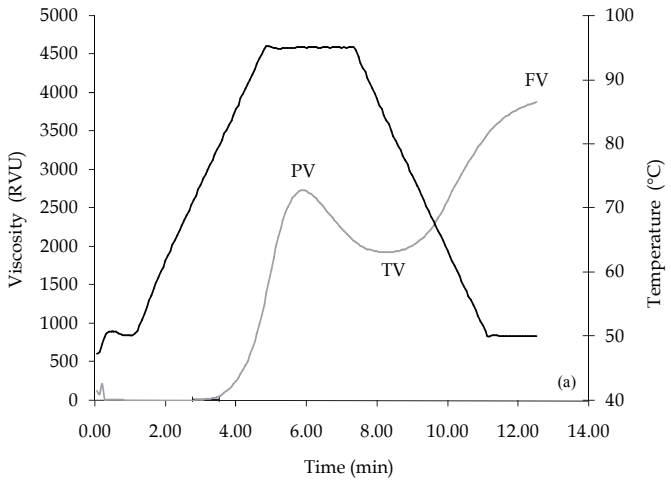


Fig. 3. Debranched chain-length distributions of immature TSN1 grains. The amylose peak is notably absent (Cuevas, 2009).

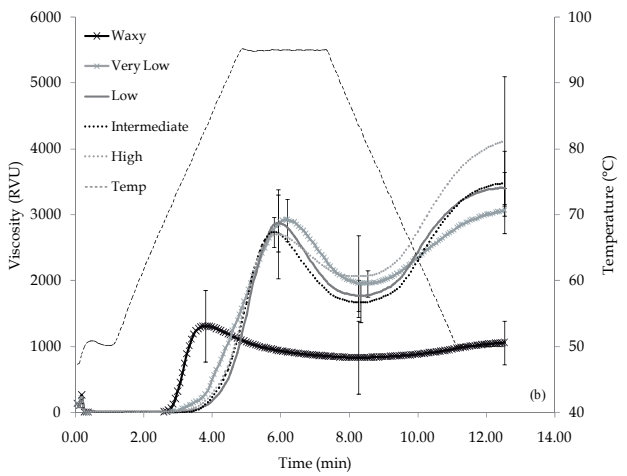
2.3 Effect of amylose on functional properties

The absence of the long linear chains of amylose in waxy rice affects its functional properties. Within the starch granule, amylose is believed to surround and to bind amylopectin blocklets together (Gallant et al., 1997; Morris, 2006; Ridout et al., 2006; Tang et al., 2006) which suppresses swelling of non-waxy starch granules (Lii et al., 1996). The effect on the swelling capacity of starch granules predictably affects the viscosity profiles of non-waxy and waxy starch (Allahgholipour et al., 2006; Fitzgerald et al., 2003); hence, viscosity curves can be used to distinguish waxy from non-waxy rice.

Viscosity parameters (Figure 4a) are usually assessed for late generation material in breeding programs (Fitzgerald et al., 2003; Larkin et al., 2003). One instrument that measures these is the Rapid Visco Analyser (RVA); starch is suspended throughout the test by constant stirring (Lai et al., 2000) and viscosity is denoted by changes in the stirring torque (Miyoshi et al., 2000) during the temperature cycle. Parts of the viscosity curve have been associated with amylose content (Bao et al., 1999; Champagne et al., 1999; Juliano, 1985; Ong & Blanshard, 1995b). The RVA has been deemed an indirect method for determining quality of rice (Yan et al., 2005).



(a)



(b)

Fig. 4. Viscosity curves obtained by Rapid Visco-Analyser. The temperature profile is denoted by the black line. (a) An example of a viscosity curve, showing the different points: PV, peak viscosity; TV, trough viscosity; FV, final viscosity. (b) Comparison of curves of the different amylose classes. Error bars are placed at PV, TV, and FV.

As the temperature increases, starch granules and proteins begin to absorb water and to swell (Batey & Curtin, 2000; Matveev et al., 2000), causing an initial rise in viscosity. The temperature at this point, known as the pasting temperature, is associated with gelatinisation properties of the starch source. Viscosity continues to increase, peaking at the point at which the balance between granule swelling and bursting has been achieved (PV) (Fitzgerald et al., 2003). Amylose and some amylopectin molecules, that leach while the granules are swelling (Tsai & Lii, 2000), form a continuous phase which contributes to the decrease in viscosity when the temperature plateaus (Lii et al., 1995); this is the lowest point in the curve and is known as trough (TV), or hot-paste, viscosity (Bergman et al., 2004; Bhattacharya & Sowbhagya, 1979; Onwulata & Konstance, 2006). As the starch slurry cools, viscosity starts to increase again because the leached amylose molecules form networks (Gidley, 1989; Nguyen et al., 1998). At the end of the test, the RVA records the final viscosity (FV) value as the temperature returns to a constant. From these three points, three additional parameters are derived: breakdown (BD, difference between PV and TV), consistency (CO, difference between FV and TV), and setback (SB, difference between FV and PV) (Bhattacharya & Sowbhagya, 1978; Fitzgerald et al., 2003; Juliano, 2007).

Waxy rice varieties used in this study exhibit distinctive pasting behaviours compared with varieties belonging to different amylose classes (Figure 4b). PV, TV, and FV are the lowest in waxy rice varieties (Figure 4b and Figure 5a-c), consistent with a previous report (Allahgholipour et al., 2006). The low FV in waxy varieties is likely to be explained by the absence of amylose chains. The linear nature of the amylose molecule allows it to retrograde as the RVA cools, forming a gel matrix in which amylopectin molecules and starch ghosts are embedded (Gidley, 1989; Gidley & Bulpin, 1989; Miles et al., 1985). In contrast, because of the abundance of chains *X* 6–9, amylopectin molecules retrograde more slowly (Kanae et al., 2004; Shi & Seib, 1992), taking up to several days or weeks (Colonna et al., 1992). Hence, the non-waxy varieties retrograde a lot faster, resulting in higher FV, than the waxy varieties (Figure 5c).

The effect of AC on PV is not as clear as its contribution to FV. It has previously been reported that AC is positively (Singh et al., 2006; Tran et al., 2001) and negatively (Chung et al., 2011; Juhász & Salgó, 2008; Tan & Corke, 2002) correlated with PV. Using data from IRRI's quality evaluation program, Figure 5a shows that AC does affect PV for non-waxy varieties, in agreement with previous reports (Larkin et al., 2003; Yan et al., 2005). Previous studies all used much smaller sample sets than were available in IRRI's quality evaluation database. While PV appears to be associated with the presence of amylose, it has been determined that varieties within the same amylose class are highly diverse in terms of PV (Figure 6a).

The heights of the viscosity curves appear to be limited by the height of the peak (Figure 6a). For instance, the low value of TV in waxy varieties (Figure 5b) can be attributed to the low PV in the waxy varieties. On the other hand, the other amylose classes had higher PV and had higher TV as well. Since TV is formed by disruption of starch structure due to shear, its level depends on the degree of swelling of the granules (which is associated with PV). Also, the associations of the derived values with amylose seem to depend on PV. BD did not associate with AC (Figure 5d), because both PV and TV did not. SB increased with amylose content (Figure 5f) not due to PV, but rather because FV is affected by AC. Meanwhile, since FV and TV of waxy varieties are lowest among the classes, it follows that the difference between the two parameters is the lowest among the classes as well (Figure 5e).

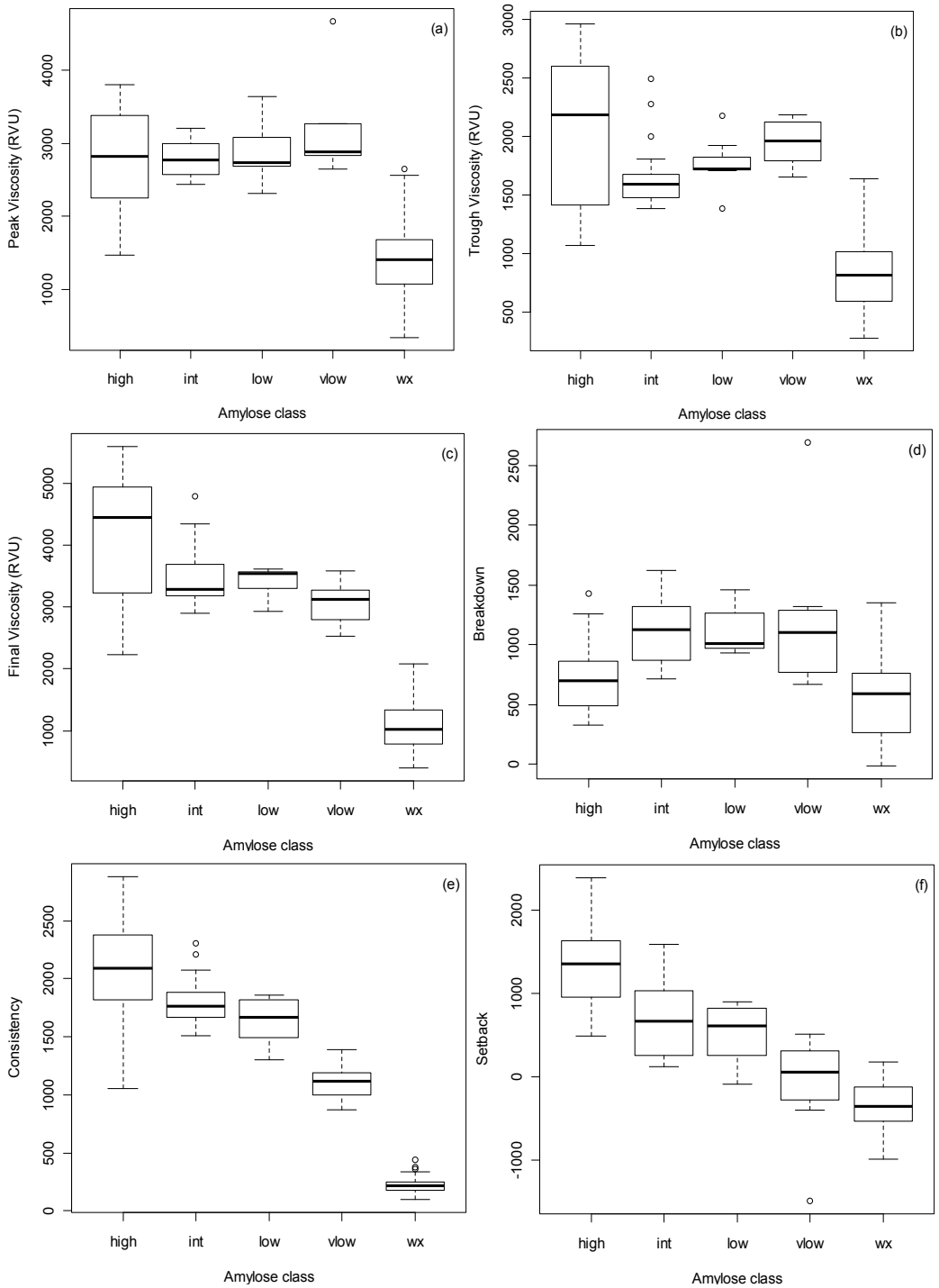


Fig. 5. Comparison of RVA parameters among the different amylose classes: (a) peak viscosity, (b) trough viscosity, (c) final viscosity, (d) breakdown, (e) consistency, (f) setback.

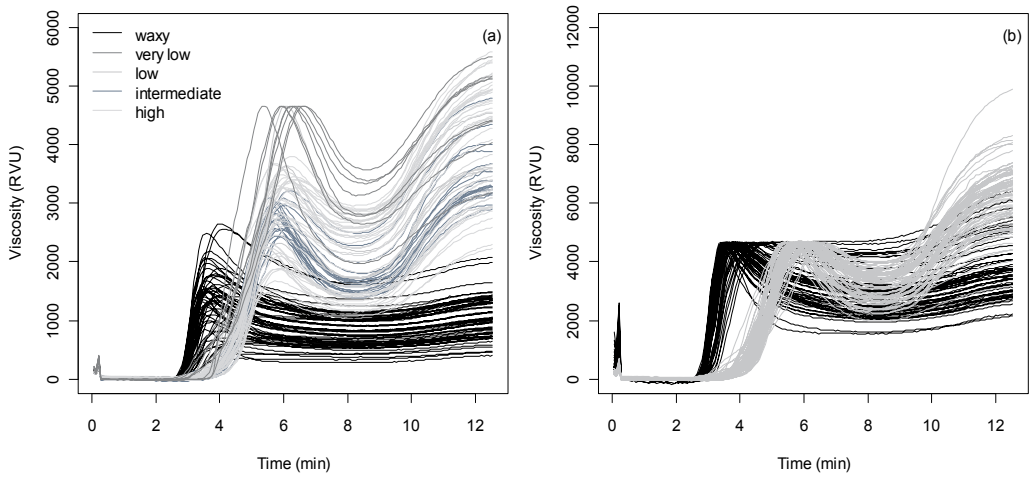


Fig. 6. Comparison of individual RVA curves by (a) amylose class; and (b) after recalculating RVA values to obtain a fixed PV value across all samples: waxy (black lines) and non-waxy (grey lines).

Due to the effect of PV (and of swelling as well) on the other RVA parameters, it could be meaningful to compare RVA curves of different samples if they have a fixed PV value (Bhattacharya & Sowbhagya, 1978). When the viscosities were recalculated so that all samples have the same PV, the samples could be classified into two main groups: one group contained the waxy varieties while the other group contained the non-waxy varieties (Figure 6b), which reaffirms the difference in pasting behaviour between waxy and non-waxy rice flour. Even with a fixed PV value, the varieties within each amylose class showed variability at other points in the RVA curves.

The variability in RVA curves within each amylose class indicates that the *Wx* gene is not the lone predictor of cooking and eating properties of rice. Proteins are reported to affect the RVA curves, particularly because of their contribution to swelling (Baxter et al., 2004; Baxter et al., 2006; Derycke et al., 2005; Fitzgerald et al., 2003; Martin & Fitzgerald, 2002; Teo et al., 2000; Xie et al., 2008). Genes coding for starch branching enzymes (BE) I and III have been associated with variability in viscosity profiles (Han et al., 2004), along with other grain components such as phosphorus (Lin & Czuchajowska, 1998).

Aside from the RVA parameters, waxy rice varieties are distinguishable from non-waxy rice varieties based on the overall shape of the RVA curve (Figure 4 and Figure 6). The increase of viscosity from pasting to peak was much faster in waxy rice than in the non-waxy varieties; this was particularly evident when the RVA curves being compared had a fixed PV value (Figure 6b).

Table 3 shows that the slope from TV to FV increased as the amount of amylose increased. The slope is significantly shallower in the waxy varieties than in the non-waxy varieties. On the other hand, the slope from pasting to PV was clearly significantly higher in waxy rice than in non-waxy rice (except for the very low amylose class) when the slope was computed

based on temperature; when the slope was computed based on the time, the waxy varieties only had significant difference in slope when compared to the high amylose varieties. The results of these comparisons are in agreement with visual comparisons of viscosity profiles in previous reports (Allahgholipour et al., 2006; Horibata et al., 2004; Yan et al., 2005).

Amylose class	N	Slope[1]		
		Paste to PV (x-axis = time)	Paste to PV (x-axis = temp.)	TV to FV
Waxy	54	1562.7[a]	134.4[a]	60.5[d]
Very Low	8	1317.4[a]	96.3[ab]	323.9[c]
Low	7	1424.6[ab]	70.3[b]	489.7[b]
Intermediate	18	1476.4[ab]	87.2[b]	540.3[ab]
High	42	1304.2[b]	74.3[b]	588.5[a]

Table 3. Means of slopes from pasting to PV and from TV to FV for the rice samples used in the study. Independent t-test was conducted to compare the slopes. Data analysed from Cuevas (2009). [1]For each column, numbers followed by the same letters are not significantly different ($\alpha = 0.05$).

Amylose content is affected by the activity of GBSSI, the enzyme encoded by the *Wx* gene. Polymorphisms within the gene have been identified and been linked with the different amylose classes. Rice functional properties, such as pasting behaviour, have also been previously associated with AC. However, within each amylose class, there is a high variability in pasting properties. Thus, other grain components besides amylose and other genes besides the *Wx* gene affect pasting behaviour. To understand the pasting properties of rice, these other factors must be taken into account as well.

3. Gelatinisation temperature

Gelatinisation temperature (GT) is used in varietal development as an indicator of the cooking time of rice samples (Cuevas et al., 2010a). It is an economically important indicator of quality because selecting for shorter cooking times leads to significant potential savings in fuel costs (Fitzgerald et al., 2009b); thus, GT is a significant component of the carbon footprint of rice.

Three classes of GT are recognised in rice breeding programs: high (>74 °C), intermediate (70 – 74 °C), and low (<70 °C) (Jennings et al., 1979; Juliano & Pascual, 1980; Juliano, 2003; Waters et al., 2006). In these programs, three methods are commonly being used to classify grains by GT: deducting 3 °C from the pasting temperature derived from the RVA (Juliano et al., 1965), the alkali spreading value (ASV) (Little et al., 1958), and by differential scanning calorimetry (DSC) (Normand & Marshall, 1989). However, various studies have shown conflicts regarding the relationship among these observations and the associations of these measurements with cooking time (Cuevas et al., 2010a).

Fortunately, the structure and the mechanism leading to high and low GT have already been discovered. Amylopectin, the other polymer of starch, is composed of linear glucose units linked by α-1,4 bonds, which are organised into clusters with α-1,6 bonds at the branch points (Ball & Morell, 2003). At the ultra-structural level, GT is affected by the debranched

chain-length distributions of amylopectin within the crystalline lamella; particularly, the proportions of chains with lengths (measured as *X*) 6 – 12 and *X* 12 – 24 (Nakamura et al., 2002). The synthesis of amylopectin chains has been associated with functionality of several enzymes. By re-plotting amylopectin chain-length distributions using a mechanistic approach originally developed to understand synthetic polymer synthesis, coupled with current biochemical knowledge, one could classify the process of amylopectin synthesis into four stages characterised by rates of chain elongation and of chain termination (Castro et al., 2005).

Starch synthase (SS) I elongates chains of *X* 6 – 12 (Fujita et al., 2006) while SSIIa is reported in various botanical sources to extend the lengths of these chains further to *X* 12 – 24 (Craig et al., 1998; Konik-Rose et al., 2007; Morell et al., 2003; Umemoto et al., 2004; Yamamori et al., 2000). Branching enzyme (BE) IIb, on the other hand, cleaves the elongating chain and transfers the short amylopectin segments (Nishi et al., 2001; Tanaka et al., 2004). Disruptions to activity of any one of these starch synthase and branching enzymes are believed to have pleiotropic effects by altering the dynamics of the processes. In wheat and maize, the enzymes of starch synthesis are organised into complexes (Hennen-Bierwagen et al., 2008; Tetlow et al., 2004, 2008), so a change to one enzyme in the complex alters the composition of the complex and thereby the functionality (Tetlow et al., 2008). Changes in enzyme functionality could lead to changes in the amylopectin chain-length distribution; such changes can be observed by comparing conventional chain-length distribution plots (Hanashiro et al., 1996; Jane et al., 1999) or re-plotting those distributions using the mechanistic approach (Cuevas et al., 2010b). Observed changes in chain-length distributions suggest mutations in the genes coding for the enzymes involved in amylopectin synthesis.

Studies addressing GT have focused on the chain elongation aspect of amylopectin synthesis. Differences in proportions of chains with *X* 6 – 12 and *X* 12 – 24 have been associated with mutations in the gene coding for SSIIa (*SSIIa*), located on chromosome 6 (Umemoto et al., 2004). Four haplotypes have been defined based on two SNPs and a functional nucleotide polymorphism (FNP) in *SSIIa*. These haplotypes generally group rice varieties into high and low GT classes (Cuevas et al., 2010a ; Umemoto & Aoki, 2005; Waters et al., 2006), with a few exceptions (Figure 7).

The four haplotypes of *SSIIa* are grouped into two classes, each spanning 10 °C in GT. Moreover, the distribution of GT by *SSIIa* haplotype does not show a clear population of varieties with intermediate GT (Figure 7) (Cuevas et al., 2010a). It has been reported that any of the four *SSIIa* haplotypes could lead to GT between 70 and 74 °C and improved varieties with intermediate GT by ASV could be classified as high GT by DSC (Cuevas et al., 2010a). Clearly, *SSIIa* does not code for intermediate GT.

Intermediate GT is commonly found in improved varieties cultivated and consumed in Asia (Juliano, 2003). In most programs at the International Rice Research Institute (IRRI), breeders select for lines with intermediate GT (Kaosa-ard & Juliano, 1991). Figure 8 shows that this is still the case in this century. Since intermediate GT has such a high importance in the selection process, the genetic basis for this trait has yet to be identified.

The inability of *SSIIa* to distinguish the intermediate GT class presents an opportunity to understand this functional property more deeply. Other enzymes involved in amylopectin synthesis may be involved. Minor quantitative trait loci (QTLs) that could possibly lower GT

have been mapped to chromosomes 1 and 7 (Bao et al., 2004). Another gene on chromosome 6, called *alk2(t)*, was associated with differences in thermal properties and amylopectin chain-length distributions in varieties with the same *SSIIa* haplotype (Shu et al., 2006b). Other enzymes such as starch synthase or branching enzymes (Butardo et al., 2011), could also be involved in changing the amylopectin structure (Ryoo et al., 2007) leading to intermediate GT.

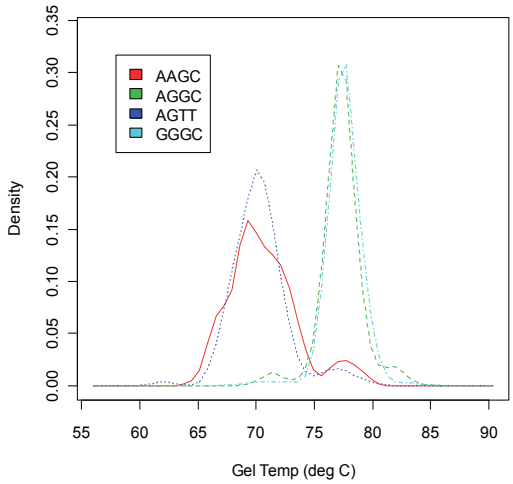


Fig. 7. Comparison of kernel density plots among the four *SSIIa* haplotypes (as represented by the SNP combinations) in 1500 rice varieties. Data analysed from Cuevas et al. (2010a).

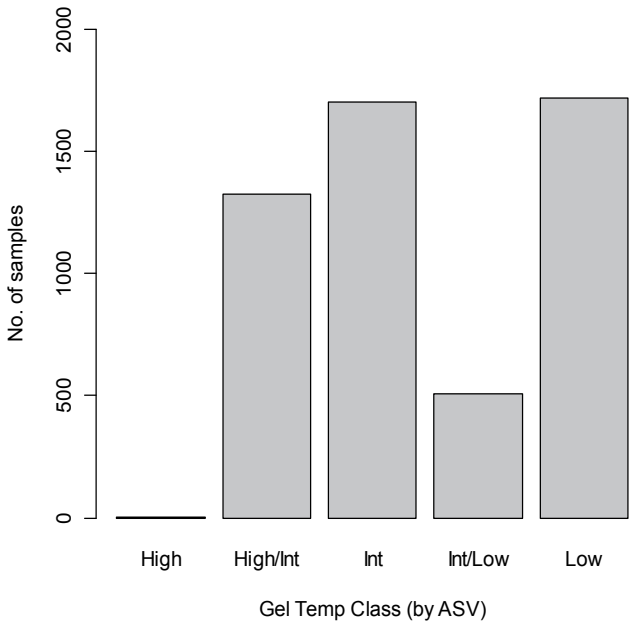


Fig. 8. Distribution of GT classes based on ASV in 29,690 IRRI breeding materials analysed in 2004 – 2007.

Other studies have investigated the role of chain termination on GT. Changes in functionality of BEIIb have been associated with variations in GT by affecting the chain-length distribution of amylopectin in *amylose-extender* mutants (Jiang et al., 2003; Nishi et al., 2001; Tanaka et al., 2004; Yamakawa et al., 2007). Comparisons of amylopectin chain-length distributions between IR36*ae*, an established *BEIIb* mutant (Asaoka et al., 1986; Juliano et al., 1990), and its wildtype IR36, and between Goami 2, a mutant variety with functional properties similar to those of *BEIIb* mutants (Kang et al., 2003), and its wildtype Ilpumbyeo, suggest that BEIIb functions in complex with SSIIa (Cuevas et al. 2010b). A number of gene interactions have been reported to affect GT in an inter-subspecific doubled haploid population (He et al., 2006). However, there is still genetic work to be done to discover the basis of intermediate gelatinisation temperature.

4. Gel consistency

Gel consistency (GC) is a measure of firmness of the rice after cooking and is performed to classify rice varieties of the same AC, particularly in the high AC class, into hard, medium, or soft texture (Cagampang et al., 1973; Kohlwey, 1994). GC is commonly measured by determining the length of a cooled gel made from flour previously cooked in 0.2 M KOH (Cagampang et al., 1973). Variations in the method exist and are used depending on the AC of the samples. For waxy varieties, which form pastes instead of gels (Lii et al., 1995; Tsai et al., 1997), the amount of flour used is higher (Bean et al., 1984) or neutral solutions are used instead of the alkaline solution (Perdon & Juliano, 1975). However, GC is not generally used in rice improvement programs focusing on rice varieties of intermediate and lower AC classes.

GC is a measure of the strength of the gel. The range of GC values to classify rice varieties according to this property is wide. Samples are grouped into arbitrarily set classes based on the length of the gel: hard (length of gel < 40 mm), medium (length of gel 41 – 60 mm), and soft (length of gel > 61 mm) (Graham, 2002). Weak and rigid gels depend on the association of starch polymers in the aqueous phase (Dea, 1989). Since amylose is the main polymer that leaches as starch granules are heated (Tsai & Lii, 2000) and amylose forms networks as the gel starts to cool (Gidley, 1989; Nguyen et al., 1998), GC could well be related to AC. Indeed, correlations between the two properties have been reported in many populations and landraces (Tan & Corke, 2002; Septiningsih et al., 2003; Zheng et al., 2007). In addition, the decrease in AC in irradiated rice was attributed to the softening of the rice gel (Yu & Wang, 2007).

The established correlations support the associations between GC and QTLs mapped to the *Wx* locus (He et al., 2006; Lanceras et al., 2000; Zheng et al., 2007). Studies have even reported that the gene coding for GC is located within the *Wx* locus (i.e., Tang et al., 1991; Tian et al., 2005). It was recently shown that the major gene for GC is in fact the *Wx* gene, and the mutation is a SNP on exon 10, a C → T polymorphism, which groups high AC rice varieties into hard and soft GC classes only (Tran et al., 2011). In the same way that a major gene separates high and low GT, as reviewed above, the extremes of GC are explained by biallelic variability at a single locus, but the intermediate class was not accounted for by that locus (Tran et al., 2011).

The relationship between AC and GC in IRRI breeding materials, however, is not as clear as those previously reported (Figure 9). These results show that GC is spread across all the AC classes in the breeding materials. Varieties from waxy and very low AC tended to have soft

GC (higher GC values) while varieties from the other AC classes had GC readings from hard to soft. Thus, other factors aside from amylose must be contributing to the strength of the cooling gel, and understanding these will assist breeding programs to select more accurately for traits of texture.

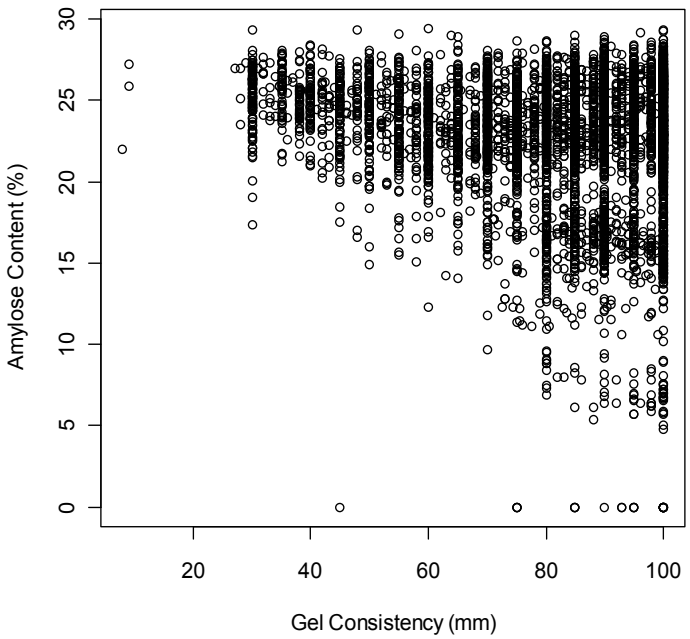


Fig. 9. Relationship of amylose content and gel consistency in IRRI breeding materials analysed in 2004 – 2007 (r = -0.169).

In the IRRI rice breeding programs, materials with soft GC appear to be preferred (Tran et al., 2011). Since the SNP in exon 10 of the *Wx* gene explains extreme variations in GC, it can be used as a tool in selecting breeding lines, particularly in programs working on varieties with high AC.

The diversity in GC in each amylose class indicates that GC may be controlled by several minor genes besides the *Wx* (He et al., 2006; Tang et al., 1991). In backcross populations derived from hard/medium and medium/soft parents, medium GC appeared to be recessive to hard GC while soft GC is recessive to medium GC, indicating three alleles for this trait (Tang et al., 1991). However, the exon 10 SNP in the *Wx* gene does not distinguish the medium type. The appearance of the medium GC trait might be contributed by other genes. For instance, various minor QTLs in chromosomes 1, 2, 6, and 7 have been associated with GC (He et al., 1999, 2006; Lanceras et al., 2000; Bao et al., 2002; Septiningsih et al., 2003; Zheng et al., 2007). Aside from minor genes, pleiotropic effects might be influencing GC as well. Interactions between *Wx* and *BEIII* and between *Wx* and *Pul* have been suggested to contribute to GC (He et al., 2006). With enzymes functioning in complexes, it may be possible that inactivity in one enzyme could affect the functionality of others within the complex, leading to medium GC, if complexes exist in rice.

5. Future directions

Though useful in predicting quality, AC, GT, and GC do not paint the whole picture of rice quality. Varieties identical in these three traits may be grouped into one quality class based on these parameters but consumers easily distinguish a premium variety from a lower quality one (Champagne et al., 2010). This could lead to low rates of adoption of newly developed improved varieties by farmers. Unfortunately, consumers are rarely able to describe the difference when they eat supposedly identical rice varieties (based on AC, GT, and GC if applicable). Hence, the next steps in discovering genes for sensory quality include finding descriptors for the sensory experience and developing phenotyping tools that can be used to quantify these descriptors. Once the phenotype is known, association mapping can begin, using appropriate populations. Such an approach will then lead to the delivery of validated genotyping tools to breeding programs.

An example of the value and need for phenotyping tools is the trait of aroma. Aroma is easy to define as present or absent in cooked rice. It was therefore possible to develop a phenotyping tool, in this case gas chromatography (Bergman et al., 2000), and then use mapping populations, genome-wide genotyping and sequencing of candidate loci to find genes and allelic variation (Kovach et al., 2009). Unfortunately, other sensory properties of rice are more abstruse because they are not as easily described by consumers. To find adjectives for these other sensory properties, descriptive sensory profiling is employed; a trained sensory panel assesses food for aroma, texture, and flavour (Champagne et al., 2010). Comparisons by a trained panel between similar varieties (in terms of AC, GT) but classed as premium and second-best showed that slickness, roughness, and springiness were textural attributes that separated the two classes while sweet taste, popcorn flavour, and metallic mouthfeel were the flavour attributes (Champagne et al., 2010). Without phenotyping tools, these traits could not be associated with genetic loci. Thus, discovering novel sensory quality genes goes in tandem with developing phenotyping tools.

6. Conclusion

Genetics has advanced people's understanding of rice grain quality. The identification of the genes controlling the different quality traits has certainly helped breeders in improving rice varieties, especially with the capacity to deliver those outcomes using genome-wide genotyping tools. In the case of AC, GT, and GC, knowledge of the genetic factors has already translated into molecular markers; these tools are being applied in breeding programs with a caveat: the genetics of the extreme classes of GT and GC are now understood; for programs aimed at targeting for the intermediate phenotype, the gene(s) have yet to be identified. For AC, the SNPs coding for all classes are likely to be known; but there are other factors that contribute to pasting properties. The genetics behind these other factors have yet to be fully understood.

Waxy rice varieties do not have amylose and thus could not be grouped into quality classes based on this property. Within the waxy class, rice varieties are very diverse and new approaches in characterising and classifying them are being developed. The RVA is a promising tool for screening eating and cooking properties of waxy rice and in defining the quality classes. However, since the viscosity curves within each amylose class are different, it is important to understand how the other factors affecting pasting properties interact with AC.

The challenge now is to find new genes for rice quality beyond AC, GC, and GT. Development of high-throughput genotyping technologies is progressing at a hectic pace. The progress, then, of finding novel sensory quality genes hinges on the pace of development of phenotyping tools that target traits that rice consumers find difficult to characterise.

7. Acknowledgment

The authors thank the people who provided rice materials mentioned in Section 2 of this chapter: Dr Kent McKenzie (California Cooperative Research Centre) for the very-low amylose mutants, Dr Wu Dianxin (Zhezhiang University) for RS111, Dr Park (Korean Rural Development Administration) for Ilpumbyeo and Goami 2, and the Genetic Resources Centre (IRRI) for the set of diverse waxy varieties.

The authors also thank the quality evaluation team of the GQNC for providing rice quality phenotyping data of diverse rice samples.

The authors appreciate the financial assistance extended by the Rural Industries Research and Development Corporation.

8. References

Allahgholipour, M., Ali, A. J., Alinia, F., Nagamine, T. & Kojima, Y. (2006). Relationship between rice grain amylose and pasting properties for breeding better quality rice varieties. *Plant Breeding* Vol. 125, pp. 357–362

American Association of Cereal Chemists International. (2000). *Approved methods of the American Association of Cereal Chemists*, American Association of Cereal Chemists International, ISBN 1891127128, St. Paul, MN

Asaoka, M., Okuno K., Sugimoto, K., Yano, M., Omura, T. & Fuwa, H. (1986). Characterization of endosperm starch from high-amylose mutants of rice (*Oryza sativa* L.). *Starch – Stärke*, Vol. 38, No. 4, pp. 114–117

Badenhuizen, N. P. (1956). The structure of the starch granule. *Protoplasma*, Vol. 45, No. 3, pp. 315–326

Bailey, J. M. & Whelan, W. J. (1961). Physical properties of starch 1. Relationship between iodine staining and chain length. *Journal of Biological Chemistry,* Vol. 236, pp. 969–973

Baker, F. & Whelan, W. J. (1951). Birefringence iodine reactions and fine-structure of waxy starches. *Journal of the Science of Food and Agriculture,* Vol. 2, No. 10, pp. 444–446

Ball, S. G., Guan, H. P., James, M. G., Myers, A. M., Keeling, P. L. Mouille, G., Buleon, A., Colonnam P. & Preiss, J. (1996). From glycogen to amylopectin: A model for the biogenesis of the plant starch granule. *Cell*, Vol. 86, No. 3, pp. 349–352

Ball, S. G. & Morell, M. K. (2003). From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule. *Annual Review of Plant Biology*, Vol. 54, pp. 207–233

Bao, J. S., He, P., Xia, Y. W., Chen, Y. & Zhu, L. H. (1999). Starch RVA profile parameters of rice are mainly controlled by *Wx* gene. *Chinese Science Bulletin*, Vol. 44, No. 22, pp. 2047–2051

Bao, J. S., Wu, Y. R.,Hu, B., Wu, P., Cui, H. R. & Shu, Q. Y. (2002). QTL for rice grain quality based on a DH population derived from parents with similar apparent amylose content. *Euphytica*, Vol. 128, No. 3, pp. 317–324

Bao, J. S., Sun, M., Zhu, L. H. & Corke, H. (2004). Analysis of quantitative trait loci for some starch properties of rice (*Oryza sativa* L.): thermal properties, gel texture and swelling volume. *Journal of Cereal Science*, Vol. 39, No. 3, pp. 379–385

Batey, I. L. & Curtin, B. M. (2000). Effects on pasting viscosity of starch and flour from different operating conditions for the Rapid Visco Analyser. *Cereal Chemistry*, Vol. 77, No. 6, pp. 754–760

Baxter, G., Blanchard, C. & Zhao, J. (2004). Effects of prolamin on the textural and pasting properties of rice flour and starch. *Journal of Cereal Science*, Vol. 40, No. 3, pp. 205–211

Baxter, G., Zhao, J. & Blanchard, B. (2006). How do endosperm proteins affect the pasting and texture of rice? A study of the physiochemical properties of rice using a model system, *13th World Congress of Food Science and Technology,* Nantes, France, September 2006

Bean, M. M., Esser, C. A. & Nishita, K. D. (1984). Some physicochemical and food application characteristics of California waxy rice varieties. *Cereal Chemistry*, Vol. 61, No. 6, pp. 475–480

Bergman, C. J., Delgado, J. T., Bryant, R. J., Grimm, C. C., Cadwallader, K. R. & Webb, B. D. (2000). Rapid gas chromatographic technique for quantifying 2-acetyl-1-pyrroline and hexanal in rice (*Oryza sativa* L.). *Cereal Chemistry*, Vol. 77, No. 4, pp. 454–458

Bergman, C. J., Bhattacharya, K. R. & Ohtsubo, K. (2004). Rice end-use quality analysis. In: *Rice Chemistry and Technology*, E. T. Champagne (Ed.), American Association of Cereal Chemists St, Paul, Minnesota

Bhattacharya, K. R. & Sowbhagya, C. M. (1978). On viscograms and viscography, with special reference to rice flour. *Journal of Texture Studies*, Vol. 9, pp. 341–351

Bhattacharya, K. R. & Sowbhagya, C. M. (1979). Pasting behaviour of rice: a new method of viscography. *Journal of Food Science*, Vol. 44, pp. 797–804

Bligh, H. F. J., Larkin, P. D., Roach, P. S., Jones, C. A., Fu, H. & Park, W. D. (1998). Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose rice varieties. *Plant Molecular Biology*, Vol. 38, No. 3, pp. 407–415

Boualaphanh, C., Daygon, V. D., Calingacion, M. N., Sanitchon, J., Jothityangkoon, D., Mumm, R., Hall, R. D. & Fitzgerald, M. A. (2011). Use of new generation single nucleotide polymorphism genotyping for rapid development of near-isogenic lines in rice. *Crop Science*, Vol. 51, pp. 2067-2073

Buleon, A., Colonna, P., Planchot, V. & Ball, S. G. (1998). Starch granules: structure and biosynthesis. *International Journal of Biological Macromolecules*, Vol. 23, No. 2, pp. 85–112

Butardo, V.M., Fitzgerald, M.A., Bird, A.R., Gidley, M.J., Flanagan, B.M., Larroque, O., Resurreccion, A.P., Laidlaw, H.K., Jobling, S.A., Morell, M.K. & Rahman, S. (2011). Impact of down-regulation of starch branching enzyme IIb in rice by artificial microRNA- and hairpin RNA-mediated RNA silencing. *Journal of Experimental Botany*, Vol. 62, No. 14, pp. 4927-4941

Cagampang, G. B., Perez, C. M. & Juliano, B. O. (1973). A gel consistency test for eating quality in rice. *Journal of the Science of Food and Agriculture*, Vol. 24, pp. 1589–1594

Cai, X. L., Wang, Z. Y., Xing, Y. Y., Zhang, J. L. & Hong, M. M. (1998). Aberrant splicing of intron 1 leads to the heterogeneous 5' UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content. *Plant Journal*, Vol. 14, No. 4, pp. 459–465

Castro, J. V., Dumas, C., Chiou, H., Fitzgerald, M. A. & Gilbert, R. G. (2005). Mechanistic information from analysis of molecular weight distributions of starch. *Biomacromolecules*, Vol. 6, No. 4, pp. 2248–2259

Champagne, E. T., Bett, K. L., Vinyard, B. T., McClung, A. M., Barton, F. E., Moldenhauer, K. A., Linscombe, S. & McKenzie, K. S. (1999). Correlation between cooked rice texture and Rapid Visco Analyser measurements. *Cereal Chemistry*, Vol. 76, No. 5, pp. 764–771

Champagne, E. T., Bett-Garber, K. L., McClung, A. M. & Bergman, C. J. (2004). Sensory characteristics of diverse rice cultivars as influenced by genetic and environmental factors. *Cereal Chemistry*, Vol. 81, No. 2, pp. 237–243

Champagne, E., Bett-Garber, K., Fitzgerald, M. A., Grimm, C. C., Lea, J., Ohtsubo, K., Jongdee, S., Xie, L., Bassinello, P., Resurreccion, A. P., Ahmad, R., Habibi, F. & Reinke, R. F. (2010). Important sensory properties differentiating premium rice varieties. *Rice*, Vol. 3, No. 4, pp. 270-281

Chen, M. H., Bergman, C. J., Pinson, S. R. M. & Fjellstrom, R. G. (2008). *Waxy* gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. Journal of Cereal Science, Vol. 47, pp. 536 – 545.

Chung, H.-J., Liu, Q., Lee, L. & Wei, D. (2011). Relationship between the structure, physicochemical properties and *in vitro* digestibility of rice starches with different amylose contents. *Food Hydrocolloids*, Vol. 25, No. 5, pp. 968-975

Colonna, P., Leloup, V. & Buleon, A. (1992). Limiting factors of starch hydrolysis. *European Journal of Clinical Nutrition*, Vol. 46, Suppl. 2, pp. 17–32

Craig, J., Lloyd, J. R., Tomlinson, K. L., Barber, L., Edwards, A., Wang, T. L., Martin, C., Headley, C. L. & Smith, A. M. (1998). Mutations in the gene encoding starch synthase II profoundly alter amylopectin structure in pea embryos. *The Plant Cell*, Vol. 10, No. 3, pp. 413–426

Cramer, G. L., Wailes, E. J. & Shui, S. (1993). Impacts of liberalizing trade in the world rice market. *American Journal of Agricultural Economics*, Vol. 75, No. 1, pp. 219–226

Cuevas, R. P. (2009). Starch microstructure and functional properties in waxy rice (*Oryza sativa* L.). School of Land, Crop and Food Science. Brisbane, University of Queensland. Doctor of Philosophy: 209 p.

Cuevas, R. P., Daygon, V. D., Corpuz, H. M., Reinke, R. F., Waters, D. L. E. & Fitzgerald, M. A. (2010a). Melting the secrets of gelatinisation temperature in rice. *Functional Plant Biology*, Vol. 37, pp. 439–447

Cuevas, R. P., Daygon, V. D., Morell, M. K., Gilbert, R. G. & Fitzgerald, M. A. (2010b). Using chain-length distributions to diagnose genetic diversity in starch biosynthesis. *Carbohydrate Polymers*, Vol. 81, pp. 120–127

Dawe, D. C. & Slayton, T. (2004). Changing structure, conduct, and performance of the world rice market. In: *Rice in Global Markets and Sustainable Production Systems*, Food and Agriculture Organization, Rome, Italy

Dea, I. C. M. (1989). Industrial polysaccharides. *Pure and Applied Chemistry*, Vol. 61, No. 7, pp. 1315-1322

Derycke, V., Veraverbeke, W. S., Vandeputte, G. E., de Man, W., Hoseney, R. G. & Delcour, J. A. (2005). Impact of proteins on pasting and cooking properties of nonparboiled and parboiled rice. *Cereal Chemistry*, Vol. 82, No. 4, pp. 468-474

Ebermann, R. & Praznik, W. (1975). Bestimmung der molekulargewichtsverteilung in nativen stärken durch gelchromatographie. *Starch - Stärke*, Vol. 27, No. 10, pp. 329–333

Fitzgerald, M. A., Martin, M., Ward, R. M., Park, W. D. & Shead, H. J. (2003). Viscosity of rice flour: A rheological and biological study. *Journal of Agricultural and Food Chemistry*, Vol. 51, No. 8, pp. 2295–2299

Fitzgerald, M. A., Bergman, C. J., Resurreccion, A. P., Moller, J., Jimenez, R., Reinke, R. F., Martin, M., Blanco, P., Molina, F., Chen, M. H., Kuri, V., Romero, M. V., Habibi, F., Umemoto, T., Jongdee, S., Graterol, E., Reddy, K. R., Bassinello, P. Z., Sivakami, R., Rani, N. S., Das, S., Wang, Y. J., Indrasari, S. D., Ramli, A., Ahmad, R., Dipti, S. S., Xie, L., Lang, N. T., Singh, P., Toro, D. C., Tavasoli, F. & Mestres, C. (2009a). Addressing the dilemmas of measuring amylose in rice. *Cereal Chemistry*, Vol. 86, No. 5, pp. 492–498

Fitzgerald, M. A., McCouch, S. R. & Hall, R. D. (2009b). Not just a grain of rice: the quest for quality. *Trends in Plant Science*, Vol. 14, No. 3, 133–139

Fujita, N., Yoshida, M., Asakura, N., Ohdan, T., Miyao, A., Hirochika, H. & Nakamura, Y. (2006). Function and characterization of starch synthase I using mutants in rice. *Plant Physiology*, Vol. 140, No. 3, pp. 1070–1084

Gallant, D. J., Bouchet, B. & Baldwin, P. M. (1997). Microscopy of starch: evidence of a new level of granule organization. *Carbohydrate Polymers*, Vol. 32, No. 3–4, pp. 177–191

Gidley, M. J. (1989). Molecular mechanisms underlying amylose aggregation and gelation. *Macromolecules*, Vol. 22, pp. 351–358

Gidley, M. J. & Bulpin, P. V. (1989). Aggregation of amylose in aqueous systems: The effect of chain length on phase behavior and aggregation kinetics. *Macromolecules*, Vol. 22, pp. 341–346

Graham, R. (2002). A proposal for IRRI to establish a Grain Quality and Nutrition Research Center. *IRRI Discussion Paper Series,* No. 44, International Rice Research Institute, Los Banos

Han, Y., Xu, M., Liu, X., Yan, C., Korban, S. S., Chen, X. & Cu, M. (2004). Genes coding for starch branching enzymes are major contributors to starch viscosity characteristics in waxy rice (*Oryza sativa* L.). *Plant Science*, Vol. 166, No. 2, pp. 357–364

Hanashiro, I., Abe, J. I. & Hizukuri, S. (1996). A periodic distribution of the chain length of amylopectin as revealed by high performance anion exchange chromatography. *Carbohydrate Research*, Vol. 283, pp. 151–159

He, P., Li, S. G., Qian, Q., Ma, Y. Q., Li, J. Z., Wang, W. M., Chen, Y. & Zhu, L. H. (1999). Genetic analysis of rice grain quality. *Theoretical and Applied Genetics*, Vol. 98, No. 3–4, pp. 502–508

He, Y., Han, Y., Jiang, L., Xu, C., Lu, J & Xu, M. (2006). Functional analysis of starch-synthesis genes in determining rice eating and cooking qualities. *Molecular Breeding*, Vol. 18, No. 4, pp. 277–290

Hennen-Bierwagen, T. A., Liu, F., Marsh, R. S., Kim, S., Gan, Q., Tetlow, I. J. Emes, M. J., James, M. G. & Myers, A. M. (2008). Starch biosynthetic enzymes from developing maize endosperm associate in multisubunit complexes. *Plant Physiology*, Vol. 146, No. 4, pp. 1892-1908

Hirano, H.-Y., Eiguchi, M. & Sano, Y. (1998). A single base change altered the regulation of the *Waxy* gene at the post-transcriptional level during the domestication of rice. *Molecular Biology and Evolution*, Vol. 15, No. 8, pp. 978–987

Hirose, T. & Terao, T. (2004). A comprehensive expression analysis of the starch synthase gene family in rice (*Oryza sativa* L.). *Planta*, Vol. 220, No. 1, pp. 9–16

Hizukuri, S., Takeda, Y., Maruta, N. & Juliano, B. O. (1989). Molecular structures of rice starch. *Carbohydrate Research*, Vol. 189, pp. 227–235

Horibata, T., Nakamoto, M., Fuwa, H. & Inouchi, N. (2004). Structure and physicochemical characteristics of endosperm starches of rice cultivars recently bred in Japan. *Journal of Applied Glycoscience*, Vol. 51, pp. 303–313

International Organization for Standardization (2007). ISO 6647-2: 2007–Rice—Determination of amylose content—Part 2: Routine methods. p. 10.

Isshiki, M., Morino, K., Nakajima, M., Okagaki, R. J., Wessler, S. R., Izawa, T. & Shimamoto, K. (1998). A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5' splice site of the first intron. *Plant Journal*, Vol. 15, No. 1, pp. 133–138

Jane, J. L., Chen, Y. Y., Lee, L. F., McPherson, A. E., Wong, K. S., Radosavljevic, M. & Kasemsuwan, T. (1999). Effects of amylopectin branch chain length and amylose content on the gelatinization and pasting properties of starch. *Cereal Chemistry*, Vol. 76, No. 5, pp. 629–637

Jennings, P. R., Coffman, W. & Kauffman, H. E. (1979). *Rice Improvement*, International Rice Research Institute,Los Banos, Laguna

Jiang, H., Dian, W. & Wu, P. (2003). Effect of high temperature on fine structure of amylopectin in rice endosperm by reducing the activity of the starch branching enzyme. *Phytochemistry*, Vol. 63, No. 1,pp. 53–59

Juhász, R. & Salgó, A. (2008). Pasting behavior of amylose, amylopectin and their mixtures as determined by RVA curves and first derivatives. *Starch - Stärke*, Vol. 60, No. 2, pp. 70–78

Juliano, B. O. (1971). A simplified assay for milled-rice amylose. *Cereal Science Today*, Vol. 16, No. 10, pp. 334–340

Juliano, B. O. (Ed.). (1985). *Rice Chemistry and Technology*, American Association of Cereal Chemists, St. Paul, MN

Juliano, B. O. (2003). *Rice Chemistry and Quality*, Philippine Rice Research Institute, Munoz, Nueva Ecija

Juliano, B. O. (2007). *Rice Chemistry and Quality*, Philippine Rice Research Institute, Munoz, Nueva Ecija

Juliano, B. O., Onate, L. U. & del Mundo, A. M. (1965). Relation of starch composition, protein content, and gelatinization temperature to cooking and eating qualities of milled rice. *Food Technology*, Vol. 19, No. 6, pp. 116–121

Juliano, B. O. & Pascual, C. G. (1980). Quality characteristics of milled rice grown in different countries. *IRRI Research Paper Series*, Vol. 48, pp. 3–25

Juliano, B. O., Perez, C. M., Kaushik, R. P. & Khush, G. S. (1990). Some grain properties of IR36-based starch mutants. *Starch - Stärke* 42, pp. 256–260

Juliano, B. O. & Villareal, R. M. (1987). Varietal differences in physicochemical properties of waxy rice starch. *Starch - Stärke*, Vol. 39, No. 9, pp. 298–301

Juliano, B. O., Villareal, R. M., Perez, C. M., Villareal, C. P., Takeda, Y. & Hizukuri, S. (1987). Varietal differences in properties among high amylose rice starches. *Starch - Stärke*, Vol. 39, No. 11, pp. 390–393

Kanae, T., Hidetoshi, S., Ken'ichiro, I. & Hayao, T. (2004). The retrogradation of starch gels and starch properties from various kinds of sweet potato starches. *Kagoshimaken Nogyo Shikenjo Kenkyu Hokoku*, Vol. 32, pp. 71–79

Kang, H. J., Hwang, I. K., Kim, K. S. & Choi, H. C. (2003). Comparative structure and physicochemical properties of Ilpumbyeo, a high-quality japonica rice, and its

mutant, Suweon 464. *Journal of Agricultural and Food Chemistry*, Vol. 51, No. 22, pp. 6598–6603

Kaosa-ard, M. & Juliano, B. O. (1991). Assessing rice quality characteristics and prices in selected international markets, In: *Rice Grain Marketing and Quality Issues*, B. O. Juliano, L. R. Pollard & G. Argosino (Eds.), pp. 23-36, International Rice Research Institute, Manila

Khush, G. S. (2005). What it will take to feed 5.0 billion rice consumers in 2030. *Plant Molecular Biology*, Vol. 59, No. 1, pp. 1–6

Kohlwey, D. E. (1994). New Methods for the Evaluation of Rice Quality and Related Terminology, In: *Rice Science and Technology*, W. E. Marshall & J. I. Wadsworth (Eds.), pp. 113 – 137, CRC Press Boca Raton, FL

Konik-Rose, Thistleton, C., J., Chanvrier, H., Tan, I., Halley, P. J., Gidley, M. J., Kosar-Hashemi, B., Wang, H. M., Larroque, O., Ikea, J., McMaugh, S., Regina, A., Rahman, S., Morell, M. K. & Li, Z. (2007). Effects of starch synthase IIa gene dosage on grain, protein and starch in endosperm of wheat. *Theoretical and Applied Genetics*, Vol. 115, No. 8, pp. 1053-1065

Kovach, M. J., Calingacion, M. N., Fitzgerald, M. A. & McCouch, S. R. (2009). The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences*, Vol. 106, No. 34, pp. 14444–14449

Kumar, I. & Khush, G. S. (1987). Genetic analysis of different amylose levels in rice. *Crop Science*, Vol. 27, No. 6, pp. 1167–1172

Lai, K. P., Steffe, J. F. & Ng, P. K. W. (2000). Average shear rates in the Rapid Visco Analyser (RVA) mixing system. *Cereal Chemistry*, Vol. 77, No. 6, pp. 714–716

Lanceras, J. C., Huang, Z.-L., Naivikul, O., Vanavichit, A., Ruanjaichon, V. & Tragoonrung, S. (2000). Mapping of genes for cooking and eating qualities in Thai jasmine rice (KDML105). *DNA Research*, Vol. 7, No. 2, pp. 93–101

Landers, P. S., Gbur, E. E. & Sharp, R. N. (1991). Comparison of two models to predict amylose concentration in rice flours as determined by spectrophotometric assay. *Cereal Chemistry*, Vol. 68, No. 5, pp. 545–548

Larkin, P. D., McClung, A. M., Ayres, N. M. & Park, W. D. (2003a). The effect of the *Waxy* locus (Granule Bound Starch Synthase) on pasting curve characteristics in specialty rices (*Oryza sativa* L.). *Euphytica*, Vol. 131, No. 2, pp. 243–253

Larkin, P. D. & Park, W. D. (2003b). Association of *waxy* gene single nucleotide polymorphisms with starch characteristics in rice (*Oryza sativa* L.). *Molecular Breeding*, Vol. 12, No. 4, pp. 335–339

Lii, C.Y., Shao, Y. Y. & Tseng, K. H. (1995). Gelation mechanism and rheological properties of rice starch. *Cereal Chemistry*, Vol. 72, No. 4, pp. 393–400

Lii, C.Y., Tsai, M.-L. & Tseng, K. H. (1996). Effect of amylose content on the rheological property of rice starch. *Cereal Chemistry*, Vol. 73, No. 4, pp. 415–420

Lin, P. Y. & Czuchajowska, Z. (1998). Role of phosphorus in viscosity, gelatinization, and retrogradation of starch. *Cereal Chemistry*, Vol. 75, No. 5, pp. 705–709

Little, R. R., Hilder, G. B. & Dawson, E. H. (1958). Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chemistry*, Vol. 35, pp. 111–126

Liu, L., Ma, X., Liu, S., Zhu, C., Jiang, L., Wang, Y., Shen, Y., Ren, Y., Dong, H., Chen, L., Liu, X., Zhao, Z., Zhai, H. & Wan, J. (2009). Identification and characterization of a novel *Waxy* allele from a Yunnan rice landrace. *Plant Molecular Biology*, Vol. 71, No. 6, pp. 609–626

Maclean, J. L., Dawe, D. C., Hardy, B. & Hettel, G. P. (Eds.) (2002). *Rice Almanac*, International Rice Research Institute, Los Banos, Philippines

Martin, M. & Fitzgerald, M. A. (2002). Proteins in rice grains influence cooking properties. *Journal of Cereal Science*, Vol. 36, No. 3, pp. 285–294.

Matveev, Y. I., Grinberg, V. Y. & Tolstoguzov, V. B. (2000). The plasticizing effect of water on proteins, polysaccharides and their mixtures. Glassy state of biopolymers, food and seeds. *Food Hydrocolloids*, Vol. 14, No. 5, pp. 425–437

McKenzie, K. S., Johnson, C. W., Jodari, F., Oster, J. J., Hill, J. E., Mutters, R. G, Greer, C. A., Canevari, W. M. & Takami, K. (2006a). Registration of 'Calamylow-201' Rice. *Crop Science*, Vol. 46, No. 5, pp. 2321–2322

McKenzie, K. S., Noble, A. E. & Takami, K. (2006b). Quality characteristics of two amylose lines. *SABRAO Journal of Breeding and Genetics*, Vol. 38, No. 1, pp. 39-46

Mikami, I., Aikawa, M., Hirano, H. H. Y. & Sano, Y. (1999). Altered tissue-specific expression at the *Wx* gene of the opaque mutants in rice. *Euphytica*, Vol. 105, No. 2, pp. 91–97

Mikami, I., Uwatoko, N., Ikeda, Y., Yamaguchi, J., Hirano, H. Y., Suzuki, Y. & Sano, Y.(2008). Allelic diversification at the *wx* locus in landraces of Asian rice. *Theoretical and Applied Genetics*, Vol. 116, No. 7, pp. 979–989

Miles, M. J., Morris, V. J., Orfore, P. D. & Ring, S. G. (1985). The roles of amylose and amylopectin in the gelation and retrogradation of starch. *Carbohydrate Research*, Vol. 135, pp. 271–281

Miyoshi, T., Itoh, M., Akiyama, S. & Kitahara, A. (2000). ALPORAS Aluminum Foam: Production process, properties, and applications. *Advanced Engineering Materials*, Vol. 2, No. 4, pp. 179-183

Morell, M. K., Kosar-Hashemi, B., Cmiel, M., Samuel, M. S., Chandler, P., Rahman, S., Buleon, A., Batey, I. L. & Li, Z. (2003). Barley *sex6* mutants lack starch synthase IIa activity and contain a starch with novel properties. *Plant Journal*, Vol. 34, No. 2, pp. 172–184

Morris, V. J. (2006). Probing nanostructures in food with AFM, *Food is Life: 13th World Congress of Food Science and Technology*, Nantes, France, September 2006

Nagao, S. & Takahashi, M. (1963). Trial construction of twelve linkage groups in Japanese rice: Genetic studies on rice plant, XXVII. *Journal of the Faculty of Agriculture Hokkaido University*, Vol. 53, No. 1, pp. 72-130

Nakamura, Y., Sakurai, A., Inaba, Y., Kimura, H., Iwasawa, N. & Nagamine, T. (2002). The fine structure of amylopectin in endosperm from Asian cultivated rice can be largely classified into two classes. *Starch - Stärke*, Vol. 54, No. 3–4, pp. 117–131

Nguyen, Q. D., Jensen, C. T. B. & Kristensen, P. G. (1998). Experimental and modelling studies of the flow properties of maize and waxy maize starch pastes. *Chemical Engineering Journal*, Vol. 70, No. 2, pp. 165–171

Nishi, A., Nakamura, Y., Tanaka, N. & Satoh, H. (2001). Biochemical and genetic analysis of the effects of *amylose-extender* mutation in rice endosperm. *Plant Physiology*, Vol. 127, No. 2, pp. 459–472

Normand, F. L. & Marshall, W. E. (1989). Differential scanning calorimetry of whole grain milled rice and milled rice flour. *Cereal Chemistry*, Vol. 66, No. 4, pp. 317–320

Ong, M. H. & Blanshard, J. M. V. (1995a). Texture determinants in cooked, parboiled rice. I. Rice starch amylose and fine structure of amylopectin. *Journal of Cereal Science*, Vol. 21, No. 3, pp. 251–260

Ong, M. H. & Blanshard, J. M. V. (1995b). Texture determinants of cooked, parboiled rice. II. Physicochemical properties and leaching behavior of rice. *Journal of Cereal Science*, Vol. 21, No. 3, pp. 261–269

Onwulata, C. I. & Konstance, R. P. (2006). Extruded corn meal and whey protein concentrate: Effect of particle size. *Journal of Food Processing and Preservation*, Vol. 30, No. 4, pp. 475–487

Perdon, A. A. & Juliano, B. O. (1975). Gel and molecular properties of waxy rice starch. *Starch - Stärke*, Vol. 27, No. 3, pp. 69-71

Ridout, M. J., Parker, M. L., Hedley, C. L., Bogracheva, T. Y. & Morris, V. J. (2006). Atomic force microscopy of pea starch: Granule architecture of the *rug3-a*, *rug4-b*, *rug5-a* and *lam-c* mutants. *Carbohydrate Polymers* 65, No. 1, pp. 64-74

Roger, P. & Colonna, P. (1996). Molecular weight distribution of amylose fractions obtained by aqueous leaching of corn starch. *International Journal of Biological Macromolecules*, Vol. 19, No. 1, pp. 51–61

Roldan, I., Wattebled, F., Lucas, M. M., Delvalle, D., Planchot, V., Jimenez, S., Perez, R., Ball, S. G., d'Hulst, C. D. & Merida, A. (2007). The phenotype of soluble starch synthase IV defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. *Plant Journal*, Vol. 49, No. 3, pp. 492–504

Rundle, R. E., Foster, J. F. & Baldwin, R. R. (1944). On the nature of the starch - iodine complex. *Journal of the American Chemical Society*, Vol. 66, No. 12, pp. 2116–2120

Ryoo, N., Yu, C., Park, C. S., Baik, M. Y., Park, I. M., Cho, M. H., Bhoo, S. H., A, G., Hahn, T. R. & Jeon, J. S. (2007). Knockout of a starch synthase gene *OsSSIIIa/Flo5* causes white-core floury endosperm in rice (*Oryza sativa* L.). *Plant Cell Reports*, Vol. 26, pp. 1083–1095

Sanchez, P. C., Juliano, B. O., Laude, V. T. & Perez, C. M. (1988). Nonwaxy rice for tapuy (rice wine) production. *Cereal Chemistry*, Vol. 65, pp. 240-243

Sano, Y. (1984). Differential regulation of *Waxy* gene expression in rice endosperm. *Theoretical and Applied Genetics*, Vol. 68, No. 5, pp. 467–473

Sano, Y. & Katsumata, M. (1984). Differential regulation of *Waxy* gene-expression in rice. *Japanese Journal of Genetics*, Vol. 59, No. 6, pp 652–652

Septiningsih, E. M., Trijatmiko, K. R., Moeljopawiro, S. & McCouch, S. (2003). Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theoretical and Applied Genetics*, Vol. 107, No. 8, pp. 1433–1441

Shelton, D. R. & Lee, W. J. (2000). Cereal Carbohydrates. In: *Handbook of Cereal Science and Technology*, K. Kulp & J. G. Ponte, (Eds.), pp. 385-416, CRC Press, Boca Raton, FL

Shi, Y. C. & Seib, P. A. (1992). The structure of four waxy starches related to gelatinization and retrogradation. *Carbohydrate Research*, Vol. 227, pp. 131–145

Shu, X.-l., Jiao, G., Fitzgerald, M. A., Yang, C., Shu, Q. & Wu, D. (2006a). Starch structure and digestibility of rice high in resistant starch. *Starch – Stärke*, Vol. 58, No. 8,pp. 411–417

Shu, X.-L., Shen, S., Bao, J. S., Wu, D., Nakamura, Y. & Shu, Q. (2006b). Molecular and biochemical analysis of the gelatinization temperature characteristics of rice (*Oryza sativa* L.) starch granules. *Journal of Cereal Science*, Vol. 44, No. 1, pp. 40–48

Singh, N., Kaur, L., Sandhu, K. S., Kaur, J. & Nishinari, K. (2006). Relationships between physicochemical, morphological, thermal, rheological properties of rice starches. *Food Hydrocolloids*, Vol. 20, No. 4, pp. 532–542

Smith, A. M., Denyer, K. & Martin, C. (1997). The synthesis of the starch granule. *Annual Review of Plant Physiology and Plant Molecular Biology*, Vol. 48, pp. 65–87

Takeda, Y., Hizukuri,S., Juliano, B. O. (1986). Purification and structure of amylose from rice starch. *Carbohydrate Research*, Vol. 148, pp. 299–308

Takeda, Y., Maruta, N. & Hizukuri, S. (1992a). Examination of the structure of amylose by tritium labeling of the reducing terminal. *Carbohydrate Research*, Vol. 227, pp.113–120

Takeda, Y., Maruta, N. & Hizukuri, S. (1992b). Structures of amylose subfractions with different molecular sizes. *Carbohydrate Research,* Vol. 226, No. 2, pp. 279–285

Tan, Y. & Corke, H. (2002). Factor analysis of physicochemical properties of 63 rice varieties. *Journal of the Science of Food and Agriculture*, Vol. 82, pp. 745–752

Tanaka, N., Fujita, N., Nishi, A., Satoh, H., Hosaka, Y., Ugaki, M., Kawasaki, S. & Nakamura, Y. (2004). The structure of starch can be manipulated by changing the expression levels of starch branching enzyme IIb in rice endosperm. *Plant Biotechnology Journal*, Vol. 2, No. 6, pp. 507–516

Tang, H., Mitsunaga, T., Kawamura, Y. (2006). Molecular arrangement in blocklets and starch granule architecture. *Carbohydrate Polymers*, Vol. 63, No. 4, pp. 555–560

Tang, X., Khush, G. S., Juliano, B. O. (1991). Genetics of gel consistency in rice (*Oryza sativa* L.). *Journal of Genetics*, Vol. 70, No. 2, pp. 69–78

Teo, C. H., Karim,A. A., Cheah, P. B., Norziah, M. H. & Seow, C. C. (2000). On the roles of protein and starch in the aging of non-waxy rice flour. *Food Chemistry*, Vol. 69, No. 3, pp. 229–236

Tetlow, I. J., Beisel, K. G., Cameron, S., Makhmoudova, A., Liu, F., Bresolin, N. S., Wait, R., Morell, M. K. & Emes, M. J. (2008). Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes. *Plant Physiology*, Vol. 146, No. 4, pp. 1878–1891

Tetlow, I. J., Wait, R., Lu, Z., Akkasaeng, R., Bowsher, C. G., Esposito, S., Kosar-Hashemi, B., Morell, M. K. & Emes, M. J. (2004). Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. *The Plant Cell*, Vol. 16, No. 3, pp. 694–708

Tian, R., Jiang, G. H., Shen, L. H., Wang, L. Q. & He, Y. Q. (2005). Mapping quantitative trait loci underlying the cooking and eating quality of rice using a DH population. *Molecular Breeding*, Vol. 15, No. 2, pp. 117–124

Tran, N., Daygon, V. D., Resurreccion, A. P., Cuevas, R. P., Corpuz, H. M. & Fitzgerald, M. A. (2011). A single nucleotide polymorphism on the *Waxy* gene explains a signficant component of gel consistency. *Theoretical and Applied Genetics*, Vol. 123, No. 4, pp. 519-525

Tran, U. T., Okadome, H., Murata, M., Homma, S. & Ohtsubo, K. (2001). Comparison of Vietnamese and Japanese rice cultivars in terms of physicochemical properties. *Food Science and Technology Research*, Vol. 7, No. 4, pp. 323–330

Tsai, M.-L., Li, C. F. & Lii, C. Y. (1997). Effects of granular structures on the pasting behaviours of starches. *Cereal Chemistry*, Vol. 74, No. 6, pp. 750–757

Tsai, M.-L. & Lii, C. Y. (2000). Effect of hot-water-soluble components on the rheological properties of rice starch. *Starch - Stärke* 52, No. 2–3, pp. 44–53

Umemoto, T. & Aoki, N. (2005). Single-nucleotide polymorphisms in rice *starch synthase IIa* that alter starch gelatinisation and starch association of the enzyme. *Functional Plant Biology* 32, No. 9, pp. 763–768

Umemoto, T., , Lin, H., Nakamura, Y., Inouchi, N., Sato, Y., Yano, M., Hirabayashi, H. & Maruyama, S. (2004). Natural variation in rice *starch synthase IIa* affects enzyme and starch properties. *Functional Plant Biology*, Vol. 31, pp. 671–684

Varavinit, S., Shobsngob, S., Varanyanond, W., Chinachoti, P. & Naivikul, O. (2003). Effect of amylose content on gelatinization, retrogradation and pasting properties of flours from different cultivars of Thai rice. *Starch - Stärke*, Vol. 55, No. 9, pp. 410–415

Wanchana, S., Toojinda, T., Tragoonrung, S. & Vanavichit A. (2003). Duplicated coding sequence in the waxy allele of tropical glutinous rice (*Oryza sativa* L.). *Plant Science*, Vol. 165, No. 6, pp. 1193–1199

Wang, Z. Y., Zheng, F. Q., Shen, G. Z., Snustad, D. P., Li M. G., Zhang, J. L. & Hong, M. M. (1995). The amylose content in rice endosperm is related to the post-transcriptional regulation of the *Waxy* gene. *Plant Journal*, Vol. 7, No. 4, pp. 613–622

Ward, R. M., Gao, Q., de Bruyn, H., Gilbert, R. G. G. & Fitzgerald, M. A. (2006). Improved methods for the structural analysis of the amylose-rich fraction from rice flour. *Biomacromolecules*, Vol. 7, No. 3, pp. 866–876

Waters, D. L. E., Henry, R. J., Reinke, R. F. & Fitzgerald, M. A. (2006). Gelatinization temperature of rice explained by polymorphisms in *starch synthase*. *Plant Biotechnology Journal*, Vol. 4, No. 1, pp. 115–122

Xie, L., N., Chen, N., Duan, B., Zhu, Z. & Liao, X. (2008). Impact of proteins on pasting and cooking properties of waxy and non-waxy rice. *Journal of Cereal Science*, Vol. 47, No. 2, pp. 372–379

Yamakawa, H., Hirose, T., Kuroda, M. & Yamaguchi, T. (2007). Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. *Plant Physiology*, Vol. 144, No. 1, pp. 258–277

Yamamori, M., Fujita, S., Hayakawa, K., Matsuki, J. & Yasui, T. (2000). Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose. *Theoretical and Applied Genetics*, Vol. 101, No. 1–2, pp. 21–29

Yamanaka, S., Nakamura, I., Watanabe, K. N. & Sato, Y. (2004). Identification of SNPs in the *waxy* gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. *Theoretical and Applied Genetics*, Vol. 108, No. 7, pp. 1200–1204

Yan, C., Li, X., Zhang, R., Sui, J. M., Liang, G. H., Shen, X. P., Gu, S. L. & Gu, M. H. (2005). Performance and inheritance of rice starch RVA profile characteristics. *Rice Science*, Vol. 12, pp. 39–47

Yu, Y. & Wang, J. (2007). Effect of γ-ray irradiation on starch granule structure and physicochemical properties of rice. *Food Research International*, Vol. 40, No. 2, pp. 297-303

Zheng, X., Wu, J. G., Lou, X. Y. & Shi, C. H. (2007). The QTL analysis on maternal and endosperm genome and their environmental interactions for characters of cooking quality in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, Vol. 116, No. 3, pp. 335–342

Part 3

Conservation of Germplasms

16

Genetic Diversity in *Gossypium* genus

Ibrokhim Y. Abdurakhmonov et al*
Center of Genomic Technologies, Institute of Genetics and Plant Experimental Biology Academy of Sciences of Uzbekistan, Yuqori Yuz, Qibray Region, Tashkent Uzbekistan

1. Introduction

Cotton (*Gossypium* spp.) is the unique, most important natural fiber crop in the world that brings significant economic income, with an annual average ranging from $27 – 29 billion worldwide from lint fiber production (Campbell et al., 2010). The worldwide economic impact of the cotton industry is estimated at ~$500 billion/yr with an annual utilization of ~115-million bales or ~27-million metric tons (MT) of cotton fiber (Chen et al., 2007). In 2011 and 2012, global cotton production is projected to increase 8% (to 26.9 million MT). This will be the largest crop since 2004 and 2005 (International Cotton Advisory Committee [ICAC], 2011).

Cotton is also a significant food source for humans and livestock (Sunilkumar et al., 2006). Cotton fiber production and its export, being one of the main economic resources, annually brings an average of ~$0.9 to 1.2 billion economic income for Uzbekistan (Abdurakhmonov, 2007) that represented 22% of all Uzbek exports from 2001-2003 (Campbell et al., 2010). The economic income from cotton production accounts for roughly 11% of the Uzbekistan's GDP in 2009 (http://www.state.gov/r/pa/ei/bgn/2924.htm, verified on September 15, 2011).

The level of genetic diversity of crop species is an essential element of sustainable crop production in agriculture, including cotton. The amplitude of genetic diversity of *Gossypium* species is exclusively wide, encompassing wide geographic and ecological niches. It is conserved *in situ* at centers for cotton origin (Ulloa et al., 2006) and preserved *ex situ* within worldwide cotton germplasm collections and materials of breeding programs. Cotton

*Zabardast T. Buriev[1], Shukhrat E. Shermatov[1], Alisher A. Abdullaev[1], Khurshid Urmonov[1], Fakhriddin Kushanov[1], Sharof S. Egamberdiev[1], Umid Shapulatov[1], Abdusttor Abdukarimov[1], Sukumar Saha[2], Johnnie N. Jenkins[2], Russell J. Kohel[3], John Z. Yu[3], Alan E. Pepper[4], Siva P. Kumpatla[5] and Mauricio Ulloa[6]
[1]*Center of Genomic Technologies, Institute of Genetics and Plant Experimental Biology, Academy of Sciences of Uzbekistan, Yuqori Yuz, Qibray Region, Tashkent, Uzbekistan*
[2]*United States Department of Agriculture-Agriculture Research Service, Crop Science Research Laboratory, Starkville, Mississippi State, USA*
[3]*United States Department of Agriculture-Agriculture Research Service, Crop Germplasm Research Unit, College Station, Texas, USA*
[4]*Department of Biology, Texas A&M University, College Station, Texas, USA*
[5]*Department of Biotechnology Regulatory Sciences, Dow AgroSciences LLC, Indianapolis, Indiana, USA*
[6]*United States Department of Agriculture-Agriculture Research Service, Western Integrated Cropping Systems, USA*

productivity and the future of cotton breeding efforts tightly depend on 1) the level of the genetic diversity of cotton gene pools and 2) its effective exploitation in cotton breeding programs. Elucidating the details of genetic diversity is also very important to determine timeframe of cotton agronomy, develop a strategy for genetic gains in breeding, and conserve existing gene pools of cotton.

During past decades, because of advances in molecular marker technology, there have been extensive efforts to explore the molecular genetic diversity levels in various cotton gene pools and genomic groups, varietal and breeding collections, and specific germplasm resources. These efforts reinforced a serious concern about the narrow genetic base of cultivated cotton germplasm, which has obviously been associated with a "genetic bottleneck" occurred during historic cotton domestication process (Iqbal et al., 2001). A narrow genetic base of cultivated germplasm was one of the major factors causing the recent cotton yield and quality declines (Esbroeck & Bowman, 1998; Paterson et al., 2004).These declines, however, are largely due to challenges and the lack of innovative tools to effectively exploit genetic diversity of *Gossypium* species. The most effective utilization of genetic diversity of *Gossypium* species further requires modern genomics technologies that help to reveal the molecular basis of genetic variations of agronomic importance. Sequencing the cotton genome(s) (Chen et al., 2007;) is a pivotal step that will facilitate the fine-scale mapping and better utilization of functionally significant variations in cotton gene pools (Abdurakhmonov, 2007). Once exploited effectively, these wide ranges of genetic diversity of the genus, in particular reservoir of potentially underutilized genetic diversity in exotic wild cotton germplasm, are the 'golden' resources to improve cotton cultivars and solve many fundamental problems associated with fiber quality, resistance to insects and pathogens and tolerance to abiotic stresses (Abdurakhmonov, 2007). In this chapter, we describe cotton germplasm resources, the amplitude of morphobiological and agronomic diversity of *Gossypium* genus and review efforts on molecular genetic diversity of cotton gene pools as well as highlight examples, challenges and perspectives of exploiting genetic diversity in cotton.

2. Description of cotton gene pools and worldwide germplasm collections

Although wild cottons (*Gossypium* spp) are perennial shrubs and trees, the domesticated cottons are tropic and sub-tropic annual crops cultivated since prehistoric times of the development of human civilization. The *Gossypium* genus of the *Malvaceae* family contains more than 45 diploid species and 5 allotetraploid species (Fryxell et al., 1992; Percival et al., 1999; Ulloa et al., 2007). These species are grouped into nine genomic types ($x = 2n = 26$, or $n = 13$) with designations: AD, A, B, C, D, E, F, G, and K (Percival et al., 1999). The species are largely spread throughout the diverse geographic regions of the world. Based on the usage of these *Gossypium* species in cotton breeding and their genetic hybridization properties, they can be grouped into 1) primary gene pool, which includes the two species from the New World, *G. hirsutum* L. and *G. barbadense* L., as well as remaining three wild tetraploid species, *G. tomentosum* Nuttall *ex* Seemann, *G. mustelinum* Miers *ex* Watt and *G. darwinii* Watt; 2) secondary gene pool, including A, B, D and F genome diploid cotton species; 3) tertiary gene pool, including C, E, G, K genome *Gossypium* species (Stelly et al., 2007; Campbell et al., 2010).

Diploid cottons, referred as Old World cottons, are classified into eight (A-G to K) cytogenetically defined genome groups that have African/Asian, American, and Australian origin (Endrizzi et al., 1985). Two of these Old World cottons from Asian origin, *G. arboreum* L. and *G. herbaceum* L., with a spinnable seed fiber, were originally cultivated in Asian continent. Today, Old World cultivated cottons remain primarily for non-industrial consumption in India and adjacent Asian countries.

The New World diploid *Gossypium* comprises of 14 (one undescribed taxon US-72) D genome species (Ulloa et al., 2006; Alvarez and Wendel, 2006; Feng et al., 2011). Taxonomically, these species are recognized as the *Houzingenia* subgenus (Fryxell, 1979, 1992). Twelve of the 14 species of this group are distributed in Mexico and extending northward into Arizona. Five species are adapted to the desert environments of Baja California [*G. armourianum* Kearney ($D_{2\text{-}1}$), *G. harknessii* Brandegee ($D_{2\text{-}2}$), and *G. davidsonii* Kellogg ($D_{3\text{-}d}$)] and NW mainland Mexico [*G. turneri* Fryxell (D_{10}) and *G. thurberi* Todaro (D_1)]. An additional seven species [*G.* sp. US-72, *G. aridum* (Rose & Standley) Skovsted (D_4), *G. lobatum* Gentry (D_7), *G. laxum* Phillips (D_9), *G. schwendimanii* Fryx. & Koch (D_{11}), *G. gossypioides* (Ulbrich) Standley (D_6), and *G. trilobum* (Mociño & Sessé ex DC.) Skovsted (D_8)] are located in the Pacific coast states of Mexico and, with the exception of the last species, are arborescent in growth habit (Ulloa et al., 2006). The other two species with disjunct distributions, *G. raimondii* Ulbrich (D_5) is endemic to Peru, while *G. klotzschianum* Andersson ($D_{3\text{-}k}$) is found in the Galápagos Islands. The D-genome species (subgenus *Houzingenia*) are classified into six sections: Section *Houzingenia* Fryxell (D_1 and D_8); Section *Integrifolia* Todaro ($D_{3\text{-}d}$ and $D_{3\text{-}k}$); Section *Caducibracteolata* Mauer ($D_{2\text{-}1}$, $D_{2\text{-}2}$, and D_{10}); Section *Erioxylum* Rose & Standley (US-72, D_4, D_7, D_9, and D_{11}); Section *Selera* (Ulbrich) Fryxell (D_6); and Section *Austroamericana* Fryxell (D_5) (Percival et al., 1999).

Until recently, evaluation of the New World D-genome species of *Gossypium*, especially Section *Houzingenia* and Section *Erioxylum*, has been limited by the lack of resource material for *ex situ* evaluation. In recent years, the United States Department of Agriculture and the Mexican Instituto Nacional de Investigaciónes Forestales Agricolas y Pecuarias (INIFAP) have sponsored joint *Gossypium* germplasm collection trips by U.S. and Mexican cotton scientists (Ulloa et al., 2006; Feng et al., 2011). As a result of these efforts, a significant number of additional *Gossypium* accessions of the subgenus *Houzingenia* from various parts of Mexico are now available for evaluation, including several accessions of each of the arborescent species (Ulloa et al., 2006). Although none of these diploid species produces cotton fibers, the D genome is one of the parental lineages of the modern allotetraploid cultivated cottons, Upland and Pima (Ulloa, 2009). Studying these D genome species is the first critical step to fulfill the pressing need to document the *in situ* conservation, to assess the genetic diversity in *Gossypium* species for the preservation of the D genome species, and to facilitate their use for cotton improvement. *In situ* conservation of some of these species is threatened by population growth and industrialized agriculture. These *Gossypium* species are donors of important genes for cotton improvement (Ulloa et al., 2006).

Hybridization between A-genome (Old World cottons) and D-genome (New World cottons) diploids and subsequent polyploidization about 1.5 million years ago created the five AD allotetraploid lineages belonging to the primary gene pool that are indigenous to America and Hawaii (Phillips, 1964; Wendel & Albert, 1992; Adams et al., 2004). These New World allotetraploid cottons include the commercially important species, *G. hirsutum* and *G.*

barbadense, which are extensively cultivated worldwide (Abdurakhmonov, 2007; Campbell et al., 2010).

G. hirsutum (also called Acala or Upland, short stapled, Mocó, and Cambodia cotton) is the most widely cultivated (90%) and industrial cotton among all *Gossypium* species. It includes the Upland cotton cultivars and other early maturing, annually grown herbal bushes. The center of origin for *G. hirsutum* is Mesoamerica (Mexico and Guatemala), but it spread throughout Central America and Caribbean. According to archaeobotanical findings, *G. hirsutum* probably was domesticated originally within the Southern end of Mesoamerican gene pool (Wendel, 1995; Brubaker et al., 1999). Consequently, two centers of genetic diversity exist within *G. hirsutum*: Southern Mexico-Guatemala and Caribbean (Brubaker et al., 1999); Mexico-Guatemala gene pool is considered the site of original domestication and primary center of diversity. Within this range, *G. hirsutum* exhibits diverse types of morphological forms, including wild, primitive to domesticated accessions. According to Mauer (1954), there are four groups of sub-species of *G. hirsutum*: (1) *G. hirsutum* ssp. mexicanum, (2) *G. hirsutum* ssp. paniculatum, (3) *G. hirsutum* ssp. punctatum, and (4) *G. hirsutum* ssp. euhirsutum (domesticated cultivars). These four groups of sub-species include within themselves a number of wild landraces and primitive predomesticated forms such as yucatanense, richmondi, punctatum, latifolium, palmeri, morilli, purpurascens and their accessions as well as a number of domesticated variety accessions from 80 different cotton growing countries worldwide (Sunilkumar et al., 2006; Lacape et al., 2007; Abdurakhmonov, 2007).

G. barbadense (also called as long staple fibered Pima, Sea Island or Egyptian cotton), accounting for about 9% of world cotton production, was originally cultivated in coastal islands and lowland of the USA and became known as Sea Island cotton. Sea Island cottons, then, were introduced into Nile Valley of Egypt and widely grown as Egyptian cotton to produce long staple fine fibers (Abdalla et al., 2001). The wide-distribution of *G. barbadense* included mostly South America, southern Mesoamerica and the Caribbean basin (Fryxell, 1979). *G. barbadense* can be divided into two botanical races *brasilense* (with kidney-seed trait) and *barbadense* (with nonaggregated seeds) that both widely present as semi-domesticated forms in Brazil (de Almeida et al., 2009). The *brasilense* race, considered to have been domesticated in the Amazonian basin (de Almeida et al., 2009) is considered a locally domesticated form for *G. barbadense* cotton (Brubaker et al., 1999; de Almeida et al., 2009).

The other three AD tetraploid species of cotton, *G. mustelinum* with specific distribution in the Northeast Brazil (Wendel et al., 1994), *G. darwinii* endemic to Galapagos Islands (Wendel & Percy, 1990), and *G. tomentosum* Nutall ex Seemann endemic to Hawaiian Islands (DeJoode and Wendel, 1992; Hawkins et al., 2005), are truly wild species (Westengen et al., 2005).

The main *ex situ* cotton germplasm collections are in the US, France, China, India, Russia, Uzbekistan, Brazil, and Australia. Although there are a few other cotton germplasm collections present in other countries of the world, these eight countries represent the majority of the world's cotton germplasm resources. Each country has a germplasm storage and conservation program in place (Campbell et al., 2010). The history of collecting an initial cotton germplasm through the specific expeditions of cotton scientists to the centers of *Gossypium* origins are well described by Ulloa et al. (2006) that were the basis, perhaps, for the majority of the current cotton germplasm collections worldwide. Consequently, to protect the world-wide economic value of cotton and cotton byproducts, cotton germplasm collections worldwide were enriched with numerous cotton germplasm accessions and

breeding materials/lines as source of the genetic diversity through continuous research efforts of specific cotton breeding programs and mutual germplasm exchange over the last 100 years (Abdurakhmonov, 2007; Campbell et al., 2010).

The brief descriptions for some of worldwide cotton germplasm collections were highlighted in several documents by Abdurakhmonov (2007), Chen et al. (2007), Stelly et al. (2007), Ibragimov et al. (2008), Wallace et al. (2009) and Campbell et al. (2010). In particular, a recent report of cotton researchers published in Crop science journal (Campbell et al., 2010) has widely described the current status of global cotton germplasm resources. Campbell et al. (2010) provided information regarding: 1) members of the collection, 2) maintenance and storage procedures, 3) seed request and disbursement, 4) funding apparatus and staffing, 5) characterization methodology, 6) data management, and 7) past and present explorations.

The contents and distribution of cotton germplasm accessions across the eight collections is summarized by Campbell et al. (2010), so we will not review the details of each collection to avoid redundancy, but rather found appropriate to list brief information in regards to the overall content and specificity of these world cotton collections. Based on a number of preserved cotton accessions in the collection, the eight major world collections can be positioned as follows: Uzbekistan (18971 accessions), India (10469 accessions), USA (10318 accessions), China (8837 accessions), Russia (6276 accessions), Brazil (4296 accessions), CIRAD (France; 3070 accessions) and Australia (1711 accessions). The main content of these collections consists of accessions for two cultivated cotton species, *G. hirsutum* and *G. barbadense*. Uzbekistan (2680 accessions), India (2283 accessions) and USA (1923 accessions) collections are the richest ones to maintain a great number of accessions for Asian diploid cottons, *G. herbaceum* and *G. arboreum* belonging to the secondary gene pool. If the collection of wild species belonging to primary, secondary and tertiary gene pools are considered Brazil (889 accessions), USA (509 accessions) and CIRAD (295 accessions) are the richest cotton collections in the world.

3. Spectra of morphological and agronomic diversity in cotton

The amplitude of genetic diversity of cotton (*Gossypium* spp), including all its morphological, physiological and agronomic properties, is exclusively wide (Mauer, 1954). There is a great deal of genetic diversity in the *Gossypium* genus with characteristics such as plant architecture, stem pubescence and color, leaf plate shape, flower color, pollen color, boll shape, fiber quality, yield potential, early maturity, photoperiod dependency, and resistance to multi-adversity environmental stresses that are important for the applied breeding of cotton. The glimpse of genetic diversity on some morphological traits is demonstrated in Figs. 1 and 2.

Besides morphological diversity in *Gossypium* genus, representatives of different genomic groups have diverse characteristics in many agronomically useful traits. Considering only *G. hirsutum* accessions, exotic and cultivar germplasm represent a wide range of genetic diversity in yield and fiber quality parameters. For example, in the analyses of ~1000 *G. hirsutum* exotic and cultivated accessions in the two different environments, Mexico and Uzbekistan, we found a wide range of useful agronomic diversities (Abdurakhmonov et al., 2004, 2006, 2008, 2009). In one or two environments, the cotton boll mass varies in a range of

1-9 grams per boll, 1000 seed mass varies in a range of 50-170 grams, the lint percentage varies in a range of 0-45%, Micronaire varies in a range of 3-7 mic, the fiber length varies in a range of 1-1.28 inch, and fiber strength varies in a range of 26-36 g/tex. There was also a wide range of variation in photoperiodic flowering (day neutral, weak to strong photoperiodic dependency) and maturity (Abdurakhmonov, 2007). This wide phenotypic diversity of cotton shows the extensive plasticity of cotton plants and potential of their wide utilization in the breeding programs as an initial material.

4. Some examples of exploiting genetic diversity through traditional breeding

Above mentioned genetic diversity, preserved in germplasm collections worldwide, are the golden resources to genetically improve the cotton cultivars. There are numerous examples on the utilization of such genetic variations in solving many fundamental problems in cotton breeding and production (Abdurakhmonov, 2007). For instance, the exploration for genetic diversity for *Verticillium* wilt fungi from the exotic *G. hirsutum ssp mexicanum var nervosum* germplasm and its on-time mobilization into the elite cultivars solved wilt epidemics in 1960's and saved Uzbekistan's cotton production, and so the economy of the country (Abdullaev et al., 2009). As a result, the wilt resistant variety series named as "Tashkent" were developed (Abdukarimov et al., 2003; Abdurakhmonov, 2007). Later, salt tolerant genotype AN-Boyovut-2 was selected from Tashkent cultivar biotypes demonstrating a continuation of a 'genetic diversity imprint' introgressed from the wild landrace stock (Abdukarimov et al., 2003). This is one of the success stories on exploiting genetic diversity and its impact from the single landrace stock germplasm, *G. hirsutum* ssp. mexicanum (Abdurakhmonov, 2007). A number of other examples on the creation of natural defoliation, disease and pest resistance, tolerance to multi-adversity stresses, improved seed oil content and fiber quality parameters, utilizing the exotic germplasm genetic diversity in Uzbekistan have been well documented (Abdukarimov et al., 2003; Abdurakhmonov et al., 2005, 2007).

Successful photoperiodic conversion program in cotton was developed to mobilize day-neutral genes into the primitive accessions of *G. hirsutum.* Day-neutral genes were introgressed into 97 primitive cotton accessions by a large backcrossing effort (McCarty et al., 1979; McCarty & Jenkins, 1993, Liu et al., 2000). This converted cotton germplasm is an important reservoir for potential genetic diversity and can be used as a source to introgress genes into breeding germplasm (Abdurakhmonov, 2007).

Similarly, using genetic diversity existing in *Gossypium* genus, reniform nematode resistance, which is one of the high cost ($100 million/year) problems in US cotton production, was addressed. Scientists succeeded in introgressing high resistance to the nematode from *G. longicalyx* into *G. hirsutum* through genetic bridge crossing of two trispecies hybrids of *G. hirsutum*, *G. longicalyx*, and either *G. armourianum* or *G. herbaceum* (Robinson et al., 2007). Later, a gene of interest was mapped (Dighe et al., 2010). Resistance to root-knot nematode was also solved with the use of genetic diversity in *Gossypium* genomes (Roberts & Ulloa, 2010). Additionally, Hinze et al. (2011) developed four diverse populations based on US germplasm collection that helped to utilize a large amount of 'still underutilized' genetic variability in cotton breeding that should be useful in sustainable cotton production with superior quality. There are many other examples recorded in different cotton breeding programs, but we limit this section with above examples and move to address the challenges behind these success stories and future perspectives in this direction.

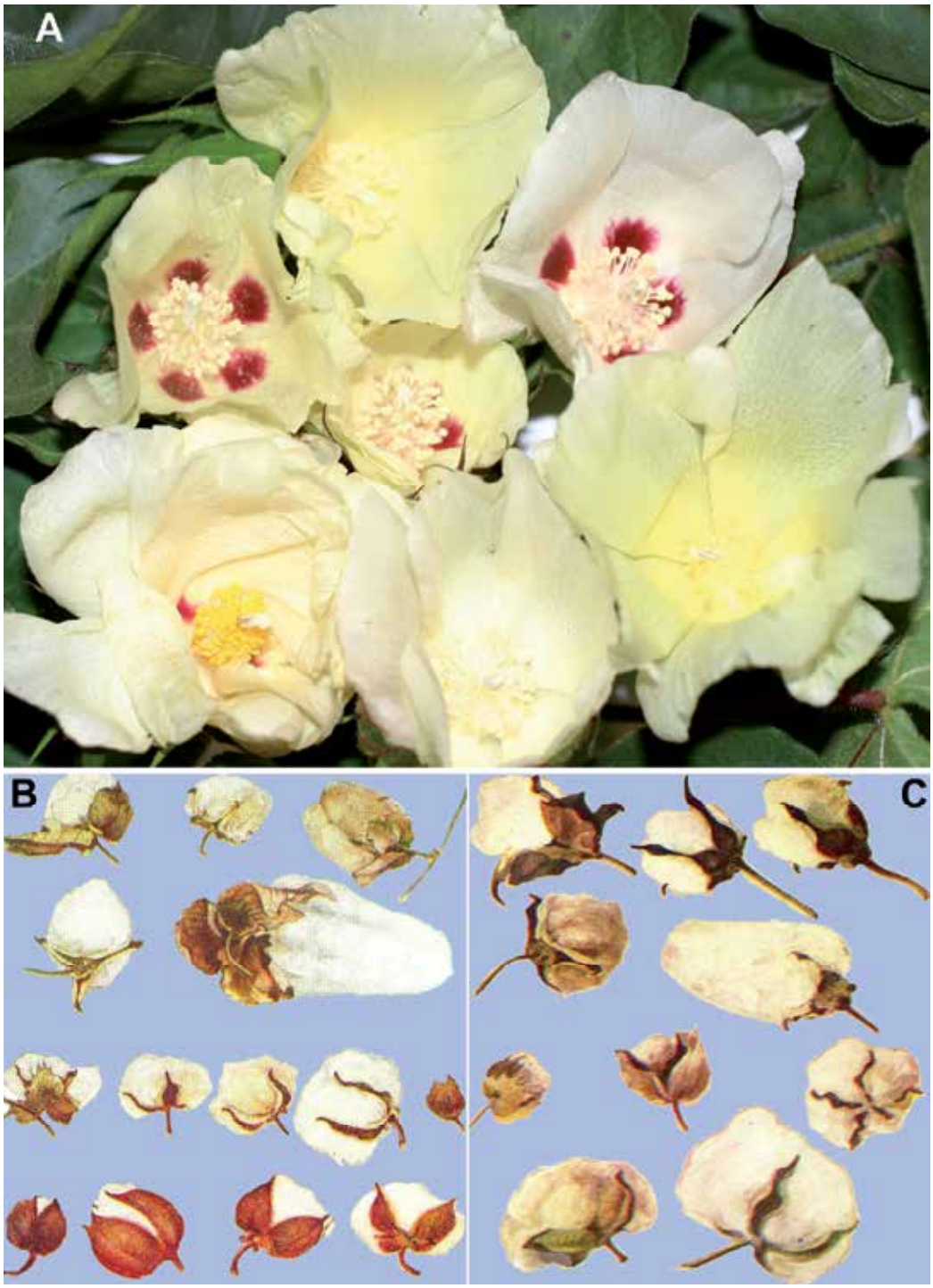


Fig. 1. Morphological trait diversity: (A)-pollen color, petal color and spot; (B)-matured bolls in diploid species, and (C)-matured bolls in tetraploid species.

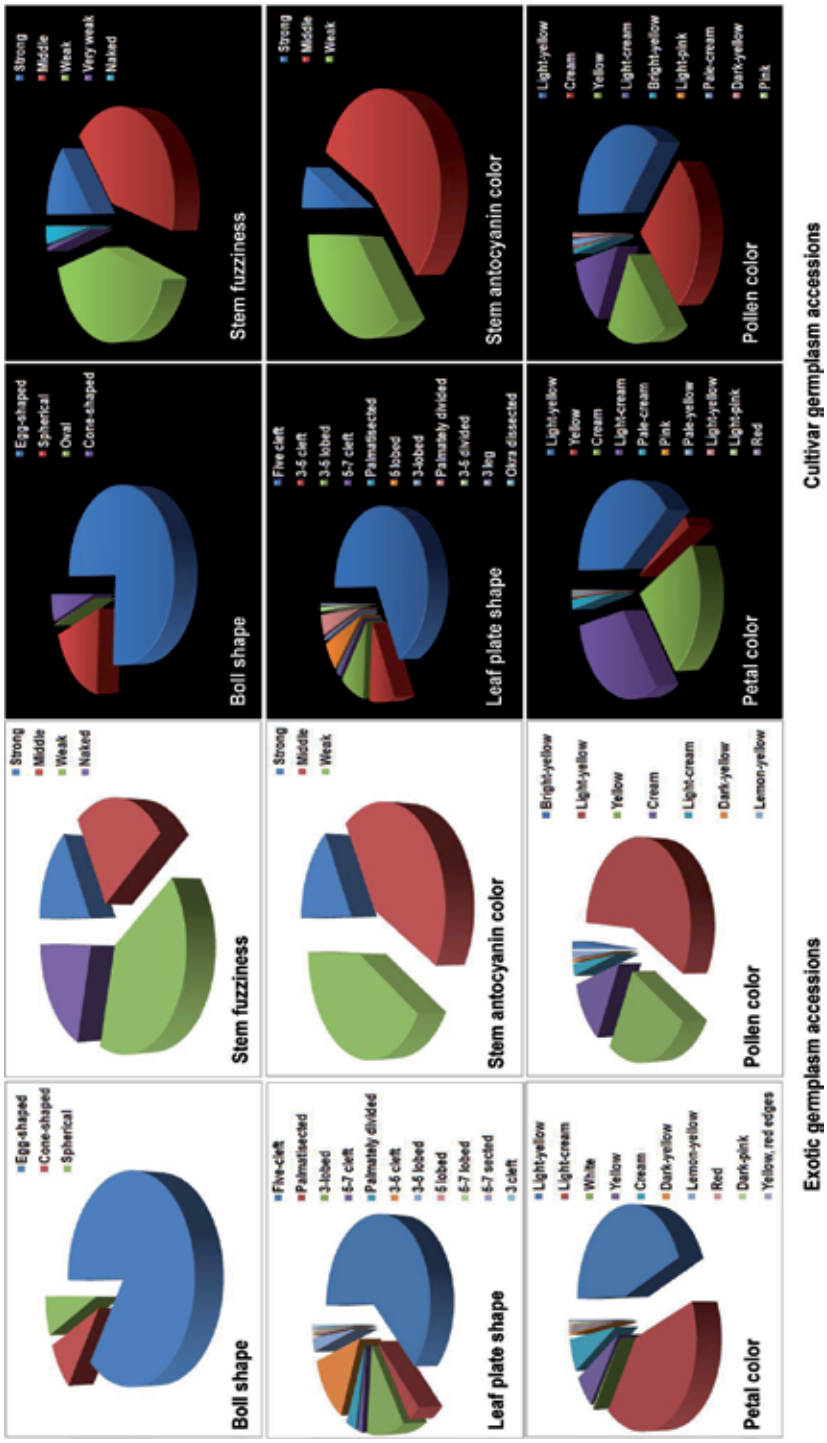


Fig. 2. Diversity on several morphological traits in exotic and cultivar germplasm of *G. hirsutum* accessions from Uzbekistan cotton collection (Abdurakhmonov et al., 2004, 2006).

5. Challenges and perspectives of exploiting diversity of different gene pools

The introduction of genetic diversity into elite cotton germplasm is difficult and the breeding process is slow. When breeders use new and exotic germplasm sources, which possess desirable genes for crop trait improvements, large blocks of undesirable genes are also introgressed during the recombination between the two parental lines (linkage drag). This linkage drag has limited the use of such germplasm. Therefore, the utilization of useful genetic diversity of the wild germplasm using traditional breeding efforts is challenging due to: 1) hybridization issues between various cotton genomes, 2) sterility issues of interspecific multi-genome hybrids, 3) segregation distortion, 4) photoperiodic flowering of wild cottons and 5) long timescale (10-12 years of efforts) required for successful introgression and recovering superior quality homozygous genotypes using traditional breeding approaches (Abdurakhmonov, 2007). This underlies necessity for the development of new innovative genomics approaches to support and accelerate the traditional efforts of exploiting the genetic diversity in cotton breeding. Continuing the introduction of genetic diversity into cultivated plants is important for reducing crop vulnerability and improving important traits such as yield, fiber quality traits, and disease and pest resistance of the cotton crop.

The most effective utilization of the genetic diversity of *Gossypium* species further requires (1) characterization of candidate gene(s) underlying the phenotypic and agronomic diversities based on genomic information in other species, (2) estimation of molecular diversity, genetic distances, genealogy and phylogeny of gene pools and germplasm groups, (3) acceleration of linkage mapping and marker-assisted selection, (4) development of efficient cotton transgenomics, and (5) sequencing cotton genome(s) (Abdurakhmonov, 2007). Furthermore, (6) it is very important to characterize and describe the existing cotton germplasm collections for both phenotypic and genomic diversity. Consequently, (7) incorporation of information into electronic web-based cotton databases such as cotton DB (http://cottondb.org), Cotton Portal (http://gossypium.info), and the Cotton Diversity Database (http://cotton.agtec.uga.edu; Gingle et al., 2006) as well as further improvement of data management tools are pivotal to facilitate an effective exploitation of the genetic diversity of cotton in the future. Cotton germplasm exchange (8) among collections and research groups is also an imperative part toward this goal (Abdurakhmonov, 2007).

6. Characterization of molecular genetic diversity in *Gossypium* genus

Molecular diversity using protein and DNA marker technologies has extensively been studied for accessions from primary and secondary gene pools. Molecular genetic diversity of tertiary gene pool cotton species is poorly explored using molecular marker technology .

6.1 Molecular diversity within primary gene pool

6.1.1 Upland germplasm

As mentioned above, application of modern molecular marker technologies, such as isozymes (Wendel & Percy, 1990; Wendel et al., 1992), random amplified polymorphic DNAs – RAPDs (Multani & Lyyon, 1995; Tatineni et al., 1996; Iqbal et al., 1997; Mahmood et al., 2009; Chaudhary et al., 2010), restricted fragment length polymorphisms – RFLPs (Wendel & Brubaker, 1993), amplified fragment length polymorphisms – AFLPs (Pilay & Myers, 1999; Abdalla et al., 2001; Iqbal et al., 2001; Rana et al., 2005; Lukonge et al., 2007) and

Simple Sequence Repeats – SSRs (Liu et al., 2000, Gutierrez et al., 2002; Rungis et al., 2005; Zhang et al., 2005a; Bertini et al., 2006; Zhang et al., 2011a; Kalivas et al., 2011) generally revealed a low level of genetic diversity within Upland cultivars. There were little variations in estimation of molecular diversity among Upland cultivars (*G. hirsutum)*; however, in general, the genetic distance reported for Upland cultivars was in the range of 0.01-0.28 (Abdurakhmonov, 2007).

Recently, we analyzed a large number of *G. hirsutum* variety and exotic accessions from Uzbek cotton germplasm collection (Fig.2) with SSR markers (Abdurakhmonov et al., 2008, 2009). Analysis of a large number of *G. hirsutum* accessions from exotic germplasm and diverse ecotypes/breeding programs with SSR markers confirmed the narrow genetic base of Upland cotton cultivar germplasm pool (with the genetic distance (GD) range of 0.005-0.26) and provided an additional evidence for the occurrence of a genetic 'bottleneck' during domestication events of the Upland cultivars at molecular level (Iqbal et al., 2001). Molecular diversity analysis of germplasm accessions using principal component analysis (PCA) suggested that germplasm resources could be broadly grouped into three large clusters (Fig.3) of exotic (1), USA-type (2) and Uzbekistan (3). First three eigenvalues of PCA analysis accounted for a ~52% variation and demonstrated existence of wide genetic diversity within the exotic germplasm, including germplasm accessions from Mexican and African origin (GD=0.02-0.50; Fig.3). We recorded a plenty of private SSR alleles within each group of accessions, specific to the germplasm groups, breeding ecotypes or exotic accessions.

A wider genetic diversity in the land race stocks of *G. hirsutum* was reported by previous studies (Liu et al., 2000, Lacape et al., 2007), suggesting the existence of sufficient genetic diversity in the exotic germplasm for future breeding programs. Rana et al. (2005) also reported a wider genetic diversity (30-87%) within *G. hirsutum* breeding lines using AFLP markers. Some recent studies have reported a relatively higher genetic diversity with an average genetic distance of up to ~37-77% in *G. hirsutum* cultivars, based on the analysis of specific germplasm resources from Brazil (Bertini et al., 2006), Pakistan (Khan et al., 2009; Azamat & Khan, 2010), India (Chaudhary et al., 2010) and China (Liu et al., 2011; Zhang et al., 2011a) breeding programs. Results of these studies were inferred from SSR or combination of a SSR and/or RAPD marker polymorphisms.

Similarly, using SSR and RAPD markers, Sapkal et al. (2011) reported moderately high level of genetic diversity (up to 57%) for 91 Upland cotton accessions with genetic male sterility maintainer and restorer properties. This suggested the existence of useful genetic diversity both in exotic and breeding line resources, useful to broaden the genetic base of Upland cotton cultivars. There is a need for evaluation of molecular genetic diversity level (Zhang et al., 2011a) and its effective exploitation in breeding programs that will address current concerns on narrowness of genetic base of widely grown Upland cotton cultivars (Hinze et al., 2011).

6.1.2 Sea Island germplasm

The molecular genetic diversity within *G. barbadense* germplasm accessions was also studied using molecular markers such as allozymes (Wendel & Percy, 1990) and AFLPs (Abdalla et al., 2001; Westengen et al., 2005). These studies revealed a narrow genetic base within *G. barbadense* accessions with a genetic distance of 7-11% (Abdalla et al., 2001; Westengen et al., 2005)

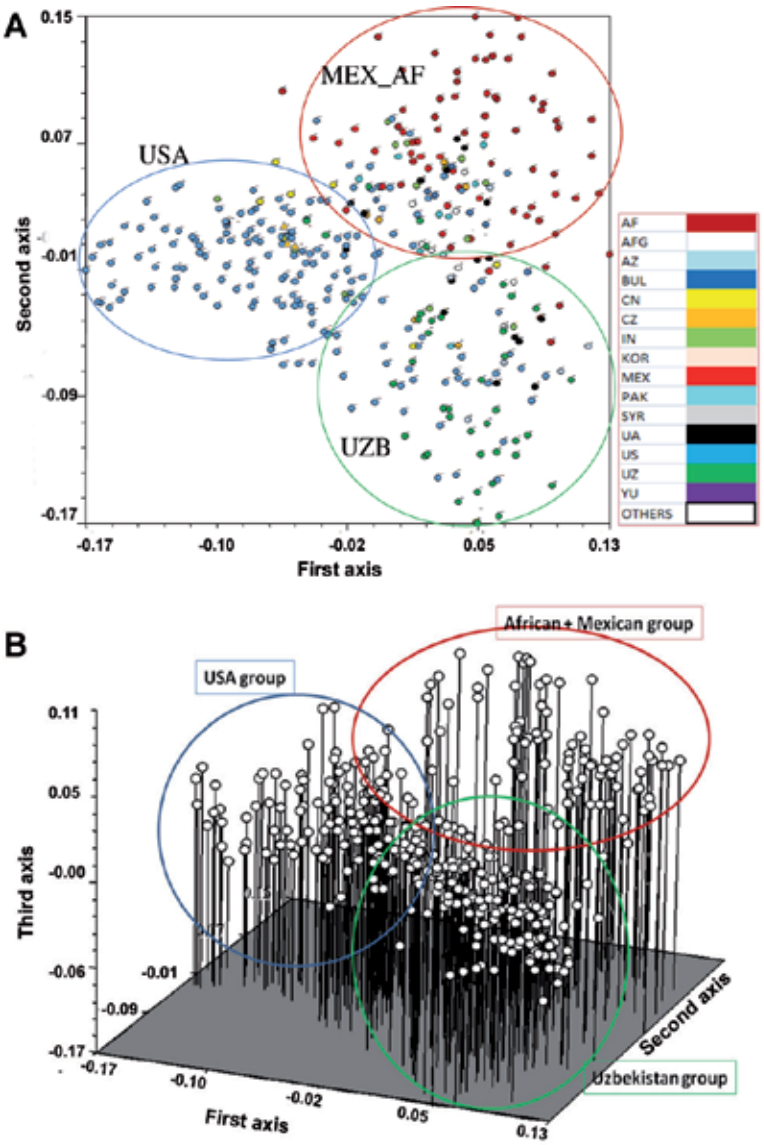


Fig. 3. Principal coordinate analysis of Upland cotton (*G. hirsutum*) accessions from Uzbek cotton germplasm collection analyzed with SSR markers. Two (A) and three (B) dimensional view for accessions from Africa (AF), Afghanistan (AFG), Bulgaria (BUL), China (CN), Czechoslovakia (CZ), India (IN), Korea (KOR), Mexico (MEX), Pakistan (PAK), Syria (SYR), Ukraine (UA), United States (US), Uzbekistan (UZB), Yugoslavia (YU), and others (Turkey, Iraq and Azerbaijan).

as was observed within the Upland cotton germplasm. In contrast, Boopathi et al. (2008) have identified highly diverse pairs of *G. barbadense* accessions using SSR marker analysis, which is useful for breeding of high quality Pima type cotton cultivars. Recently, de Almeida et al. (2009) have studied the molecular diversity level of *G. barbadense* populations

in situ preserved in the two states of Brazil, Ampa and Para. The genetic analysis using SSR markers of plant populations in these two states revealed 1) high homozygosity in each genotype tested, 2) high total genetic diversity (H_e=39%) in *G. barbadense* populations studied and 3) high level of population differentiation (F_{st}=36%) between cotton plants from these two Brazilian states. Results suggested the existence of noticeable genetic diversity preserved in *in situ* populations of *G. barbadense* in Brazil that should be further maintained within an *ex situ* germplasm collection to guarantee its long term preservation (de Almeida et al., 2009). Similarly, there is useful genetic diversity in *ex situ* preserved *G. barbadense* germplasm collections worldwide. For instance, the molecular diversity analysis of *G. barbadense* accessions using SSR markers revealed that moderately higher genetic diversity (up to 34%) exists within former USSR (that includes collections of Uzbekistan and Russia), China, USA, and Egypt germplasm collections (Wu et al., 2010). In that, USSR collection demonstrated the extraordinary genetic diversity compared with other collections whereas Egyptian collection had the least genetic diversity.

6.1.3 Wild allotetraploid germplasm

The molecular diversity revealed by AFLP markers was low within *G. tomentosum* germplasm with a genetic distance range of 2-11% (Hawkins et al., 2005). However, recent efforts on the characterization of genetic diversity level of three *in situ* preserved *G. mustelinum* population from Brazil using SSR markers suggested 1) high level of homozygosity within each population studied and 2) existence of high level of total genetic differentiations (58.5%) between them, which is due to geographic isolations and genetic founder effects (Barrosso et al., 2010). Wendel & Percy (1990) analyzed 58 *G. darwinii* accessions from six islands using 17 isozyme markers encoded by 59 genetic loci and identified high genetic diversity level within its accessions and relationships with *G. barbadense* and *G. hirsutum* genomes. This classical study suggested that *G. darwinii* is closely related to *G. barbadense* despite having gene flow imprints from *G. hirsutum;* however, *G. darwinii* has a large number of unique alleles to be considered a distinct genome (Wendel & Percy, 1990).

6.2 Molecular diversity within secondary gene pool

The genomic diversity of the A-genome diploid cottons has also been studied using molecular marker technology (Liu et al., 2006; Guo et al., 2006; Kebede et al., 2007; Rahman et al., 2008; Kantartzi et al., 2009; Patel et al., 2009; Azamat & Khan, 2010). The genetic distance within 39 *G. arboreum* L (A_2A_2-genome) accessions, analyzed with SSR markers, ranged from 0.13-0.42 (Liu et al., 2006) demonstrating the existence of wider genomic diversity in the A-genome diploids compared to the Upland cultivar germplasm. Kebede et al. (2007) reported, however, moderate level of genetic diversity within each A_1 and A_2-genome cottons that ranged from 0.03-0.20 with an average of 0.11 within *G. herbaceum* and 0.02-0.18 with an average of 0.11 for *G. arboreum* (A_2). The overall genetic distance between A_1 and A_2 genomes was up to 36-38% (Kebede et al., 2007; Mahmood et al., 2010). In fact, *G. arboreum* arose from the primitive perennial form of *G. herbaceum* spread in India and there is a single reciprocal chromosomal translocation in *G. arboreum* genome compared to *G. herbaceum* (Guo et al., 2006). Molecular diversity revealed by SSR markers was higher within *G. arboreum* accessions (an average of 25%) compared to *G. herbaceum* accessions (an average of 4%; Patel et al., 2009), suggesting differences in two closely related cotton genome

germplasm resources. This is an interesting finding but is in contrast to the report by Kebede et al. (2007) where an average genetic diversity within A_1 and A_2 genome accessions was equal.

Rahman et al. (2008) studied 32 *G. arboreum* accessions specific to Pakistan with RAPD markers and found up to 53% genetic diversity between studied accessions with very narrow diversity within cultivated *G. arboreum* accessions compared to non-cultivated ones. Analyzing 96 *G. arboreum* accessions with SSR markers, Kantartzi et al. (2009) reported that genetic distance within these geographically diverse A_2 genome accessions ranged up to 51%. In a more recent study, Azamat & Khan (2010) also reported wider genetic diversity in *G. arboreum* cultivar germplasm revealed by RAPD (GD=0.371) and SSR markers (GD=0.41). Although variable genetic distance estimates are presented, these reports collectively suggest that A genome representatives of secondary gene pool have sufficient molecular diversity useful for breeding programs.

Studying a large number of accessions for D genome cotton such as *G. aridum* (D_4), *G. davidsonii* ($D_{3\text{-}d}$), *G. klotzschianum* ($D_{3\text{-}k}$), G. laxum (D_9), G. lobatum (D_7), *G. schwendimanii* (D_{11}) with AFLP markers Alvarez & Wendel (2006) have reported 7 to 54% genetic diversity among D-genome accessions studied. A wider range of genetic diversity was observed among 12 D-genome diploid cottons with the genetic similarity of 0.08-0.94 (Guo et al., 2007a), suggesting existence of diverse variations in D-genome cotton germplasm useful for breeding programs. Recently, Feng et al. (2011) have studied 33 arborescent D-genome accessions, including 23 accessions of *G. aridum* with RAPD and AFLP markers. They found high molecular diversity among accessions studied, varying from 32% to 84%. This study suggests for continual efforts to study these D-genome American *Gossypium* species (subsection *Erioxylum*) to resolve genetically distant geographical ecotypes useful for cotton improvement (Feng et al., 2011)

6.3 Molecular diversity within tertiary gene pool

There is a limited information on molecular diversity estimates for tertiary germplasm pool accessions, including C, E, G and K-genomic species. Recently, Tiwari & Stewart (2008) reported AFLP marker-based molecular diversity analysis results for 57 accessions of C- and G-genome species, including *G. australe* F. Mueller *(G)*, *G. nelsonii* Fryxell (G_3), *G. bickii* Prokhanov (G_1) and *G. sturtianum* J.H. Willis (C_1). Results showed that within *G. australe* accessions, the pairwise mean genetic distance was in a range of 3-15%, suggesting narrow genetic diversity within *G. australe* accessions that could be due to relatively recent seed dispersal over large growing area of this species (Tiwari & Stewart, 2008). However, there was moderately high molecular diversity between *G. australe* and *G. nelsonii* accessions, ranging from ~17-31%.Higher molecular diversity of up to ~43% was found between *G. australe* and *G. bickii* accessions. The genetic distance between *G. bickii* and *G. nelsonii* varied from 25% to 35% and as expected C_1-genome accessions were most distantly related ones to these three G-genome species (Tiwari & Stewart, 2008). There is no report on molecular diversity studies on other representatives of tertiary gene pool.

6.4 Molecular diversity among cotton gene pools

The genetic diversity among different gene pools was also estimated in many studies using various marker systems. AFLP marker analyses studies (Iqbal et al., 2001, Abdalla et al.,

2001, Westengen et al., 2005) revealed that the genetic distance between *G. barbadense* and *G. hirsutum* was in the range of 21-33%. The other wild AD tetraploids (*G. mustelinum, G. tomentosum*) were close to the cultivated AD cottons sharing 75-84% similarity, where *G. tomentosum* was closer to *G. hirsutum* genome (GD=0.16) than the other allotetraploid species (Westengen et al., 2005). At the same time, as mentioned above, *G. darwinii* was closer to *G. barbadense* than *G. hirsutum* (Wendel & Percy, 1990).

Based on AFLP marker analysis, the genetic distance between the widely cultivated AD cottons (*G. barbadense* and *G. hirsutum*) and A-genome diploids varied from 45 to 69%, and that between the cultivated AD cottons and the D-genome varied from 55 to 71%. The genetic distance between the wild AD tetraploids and the A-genome was in the range of 46-52%, and between the wild AD cottons and the D-genome was 58-59%. The genetic distance between the A- and D-genome cottons was in the range of 0.72-0.82 when analyzed with AFLPs (Iqbal et al., 2001, Abdalla et al., 2001, Westengen et al., 2005).

The use of SSR markers revealed that the genetic distance between *G. hirsutum* and *G. barbadense* was in a range of 42-54% (Kebede et al., 2007). However, Lacape et al. (2007) reported higher genome dissimilarity values (D=0.89-0.91%) between *G. hirsutum* and *G. barbadense* within their material. Also, high mean dissimilarity values were reported between *G. hirsutum* and *G. tomentosum* (D=0.71-0.75) and between *G. barbadense* and *G. tomentosum* (D=0.80) using highly polymorphic sets of SSRs (Lacape et al., 2007). The genetic distance among the AD tetraploids was also in the range of 0.80-0.88 (Liu et al., 2000) with moderate closeness of *G. tomentosum* to the Upland cotton than *G. barbadense* cultivars that was also supported by other studies with different marker systems (Dejoode & Wendel, 1992; Hawkins et al., 2005). Based on SSR marker analysis, the genetic distance between the cultivated AD cottons and the A-genome was in the range of 31-43%, and that between the cultivated AD cottons and the D-genome was in the range of 35-46% (Kebede et al., 2007). The genetic distance between A-and D-genome cottons varied in the range of 29-42% (Kebede et al., 2007).

7. Perspectives of 21st century cotton genomics efforts in characterizing and exploiting the genetic diversity of *Gossypium* species

During the past two decades, the international cotton research community has made extensive efforts to utilize the genetic diversity in cotton, which are imperative for the future of trait improvements of the cotton crop. There are many marker systems such as isozymes, RAPDs, RFLPs, AFLPs (extensively referenced herein), and their various modifications (Zhang et al., 2005b) successfully used in cotton. However, the development of a large collection of robust, portable, and PCR-based molecular marker resources such as Simple Sequence Repeats (SSRs; www.cottonmarker.org) and Single Nucleotide Polymorphisms (SNPs) for cotton were one of the tremendous accomplishments of cotton research community (Chen et al., 2007; Van Deynze et al., 2009). This accelerated studies on genetic diversity in cotton at genomic level. Cotton marker resources were made available for cotton research community through cotton marker database (CMD) (Blenda et al., 2006) that are being extensively used to create cotton genetic linkage maps and to map important agronomic QTLs (Abdurakhmonov, 2007; Chen et al., 2007; Zhang et al., 2008). In addition to available DNA marker systems, recently, Reddy et al. (2011) developed a diversity array technology (DArT) marker platform for the cotton genome and evaluated the use of DArT

markers compared with AFLP markers in mapping populations. These studies are very important to elucidate molecular basis of genetic diversities in cotton that are vital to mobilize useful genes of agronomic importance to the elite cultivars through marker-assisted breeding programs.

Furthermore, researchers have reported several potential candidate genes of many agronomic traits in cotton. Tremendous efforts were made to study molecular basis of one of the most complex, but important traits – cotton fiber development (Abdurakhmonov, 2007; Chen et al., 2007; Zhang et al., 2008). These efforts, including many more recent reports on the dissection of candidate genes that are specifically expressed in developing fibers are undoubtedly imperative for future exploitation of genetic diversity in cotton fiber traits using transgenomics approaches (Arpat et al., 2004; Ruan et al., 2003; Zhang et al., 2011b).

Despite wide spectra of genetic diversity in *Gossypium* genus and extensive cotton genomics efforts, cotton lags behind other major crops for marker-assisted breeding due to limited polymorphism in the cultivated germplasm. This underlies broadening of cultivar germplasm genetic base through mobilization of useful gene variants from other gene pools into cultivated germplasm. There is a need for application of modern innovative genomics tools such as association mapping to identify genetic causatives of natural variations preserved in cotton germplasm resources and their use in plant breeding. Efforts on turning the gene-tagging efforts from bi-parental crosses to natural population or germplasm collections, and from now classical QTL-mapping approach to modern linkage disequilibrium (LD)-based association study should lead to elucidation of *ex situ* conserved natural genetic diversity of worldwide cotton germplasm resources and its effective utilization. LD refers to a historically reduced (non-equilibrium) level of the recombination of specific alleles at different loci controlling particular genetic variations in a population (Abdurakhmonov & Abdukarimov, 2008). Although novel to cotton research, the association genetics strategy is, in fact, highly applicable to the identification of markers linked to fiber quality and yield through the examination of linkage disequilibrium (LD) of DNA-based markers with fiber quality and yield traits in a large, diverse germplasm collection (Abdurakhmonov et al., 2004, 2008, 2009).

Application of association mapping strategy in gene mapping and germplasm characterization gained wider use in cotton. For example, Kantartzi & Stewart (2008) conducted association analysis for the main fiber traits in 56 *G. arboreum* germplasm accessions introduced from nine regions of Africa, Asia and Europe using 98 SSR markers. Association mapping strategy was also applied for tagging fiber traits in the exotic germplasm derived from multiple crosses among *Gossypium* tetraploid species (Zeng et al., 2009). Both of these studies did not quantify the LD level in the population and used marker-trait associations to tag genetic variations contributing to the trait of interest.

Alternatively, to better assess and exploit a molecular diversity of cotton genus, we conducted molecular genetic analyses in a global set of ~1000 *G. hirsutum* L. accessions, one of the widely grown allotetraploid cotton species, from Uzbek cotton germplasm collection. This global set represented at least 37 cotton growing countries and 8 breeding ecotypes as well as wild landrace stock accessions. The important fiber quality (fiber length and strength, Micronaire, uniformity, reflectance, elongation, etc.) traits were measured in two distinct environments of Uzbekistan and Mexico. This study allowed us to quantify the linkage disequilibrium level in the genome of Upland cotton germplasm and to design an

"association mapping" study to find biologically meaningful marker-trait associations for important fiber quality traits that accounts for population confounding effects (Yu et al., 2006; Abdurakhmonov & Abdukarimov, 2008). Several SSR markers associated with major fiber quality traits along with donor accessions were identified and selected for MAS programs (Abdurakhmonov et al., 2008, 2009).

Further, with the specific objective of introducing and enriching the currently-applied traditional breeding approaches with more efficient modern MAS tools in Uzbekistan, we began marker-assisted selection efforts based on our association mapping studies mentioned above. For this purpose, we selected (1) a set of twenty three major (Micronaire, fiber strength and length, and elongation) fiber trait-associated DNA markers as a tool to manipulate the transfer of QTL loci during a genetic hybridization; and (2) thirty-seven (11 wild race stocks and 26 variety accessions from diverse ecotypes) donor cotton genotypes that bear important QTLs for fiber traits. These donor genotypes were crossed with 9 commercial cultivars of Uzbekistan (as recipients) in various combinations with the objective of improving one or more of fiber characteristics of these recipients. These 9 parental recipient genomes were first screened with our DNA-marker panel to compare with 37 donor genotypes. The polymorphic status of marker bands between donor and recipient genotypes were recorded. The hybrid plants generated from each crossing combination were tested using DNA-markers at the seedling stage, and hybrids bearing DNA-marker bands from donor plants were selected for further backcross breeding (Abdurakhmonov et al., 2011).

Testing the major fiber quality traits using HVI in trait-associated marker-band-bearing hybrids revealed that mobilization of the specific marker bands from donors really had positively improved the trait of interest in recipient genotypes (data not shown). Currently, we developed a second generation of recurrent parent backcrossed hybrids (F_1BC_2), bearing novel marker bands and having superior fiber quality compared to original recipient parent (lacking trait-associated SSR bands). These results showed the functionality of the trait-associated SSR markers detected in our association mapping efforts in diverse set of Upland cotton germplasm. Using these effective molecular markers as a breeding tool, we aim to pyramid major fiber quality traits into single genotype of several commercial Upland cotton cultivars of Uzbekistan. Our efforts will not only help rapid introgression of novel polymorphisms, broadening the genetic diversity of cotton cultivars and accelerating the breeding efforts for future sustainable cotton production in Uzbekistan but also exemplify effective exploitation of the natural genetic diversity *ex situ* preserved in cotton germplasm collections (Abdurakhmonov et al., 2011).

In spite of successful application of association mapping in cotton, there is a great challenge with assigning correct allelic relationships (identity by decent) of multiple band amplicons when diverse, reticulated, and polyploid cotton germplasm resources lacking historical pedigree information are investigated. Besides, there is the issue of rare and unique alleles that is problematic for conducting association mapping (Abdurakhmonov & Abdukarimov, 2008). While these issues can be solved using many available methodologies and approaches (Abdurakhmonov & Abdukarimov, 2008); however, recent studies in model crops suggested a new methodology to minimize these issues with the creation of segregating populations, performing genetic crosses between several reference populations with known allele frequencies for functional polymorphisms. Such an approach is referred to as nested

association mapping (NAM) and NAM populations would greatly enhance the power of association mapping in plants (Stich and Melchinger, 2010). The usefulness and feasibility of NAM population based genetic mapping studies were successfully demonstrated in maize (Kump et al., 2011; Poland et al., 2011; Tian et al., 2011;) and should be adopted for other crops with complex genome and diverse germplasm resources like cotton. Therefore, creation of NAM populations for cotton on the basis of germplasm evaluation and characterization studies is the task of high priority for future characterization and mapping of biologically meaningful genetic variations in cotton. This requires further efforts and investments that facilitate fine-scale association mapping studies in cotton. This will ultimately lead to cloning and characterization of genetic causatives controlling the genetic diversities and its effective exploitation in plant breeding.

8. Conclusions

In conclusion, by having a wide geographic and ecological dispersal, the *Gossypium* genus represents and preserves large amplitude of morphobiological and genetic diversity within its *ex situ* worldwide germplasm collections and *in situ* occupation sites. Because of the development of molecular marker technologies, and their application in genetic diversity studies of germplasm resources, various gene pools and specific cultivar groups, researchers found a genetic bottleneck in cultivated cotton germplasm resources. However, there is moderately high molecular diversity present in some specific cultivar germplasm analyzed worldwide, suggesting a need for continual efforts on searching the diverse cultivar germplasm resources using molecular markers. There is a need to extend molecular marker-based diversity studies for tertiary gene pool accessions of cotton. There also is high genetic diversity available within exotic land race stocks, wild AD cottons, and putative A-and D-genome ancestors to AD cottons that have potential to search for genetic variations useful in future improvement of cotton. The variations observed in genetic distance estimations between different studies could be due to (1) germplasm resources chosen for the study, (2) number of accessions analyzed with molecular markers, (3) number of markers and marker types used, (4) genomic regions screened, and (5) subjective features of data analyses process, e.g., considering or removing unique or rare alleles, largely influencing the genetic distance measures.

Further, the narrowness of genetic diversity in cultivar germplasm was associated with recent and possibly future declines in cotton production and its quality, which was a timely warning to accelerate efforts on broadening the genetic base of cultivar germplasm resource via mobilizing novel genetic variants from wild, primitive, pre-domesticated primary, secondary and tertiary gene pools. Traditional efforts have succeeded in introgressing many new genetic variations into cultivar germplasm from other gene pools, but it is still challenging and the breeding process is slow due to a number of genetic barriers and obstacles, as highlighted above, to accomplish the goal. This underlies the importance of development of innovative tools to exploit the biologically meaningful genetic variations, existing in *Gossypium* genus. The most effective utilization of genetic diversity of *Gossypium* species further requires characterization of candidate gene(s) underlying the phenotypic and agronomic diversities, acceleration of linkage mapping, map-based cloning and marker-assisted selection that underlie development of modern genomics technologies such as high-resolution, cost effective LD-based association mapping for cotton with its optimization

through development of modern nested association mapping populations. The development of efficient cotton transgenomics tools and complete sequencing of cotton genome(s) will further accelerate exploitation of genetic diversity in highly specific manner and with clear vision. Future application of whole genome-association strategy with epigenomics perspectives, which currently is widely being applied in human and the other model plants such as Arabidopsis, will have a significant impact on identifying true functions of genes controlling available genetic diversity, and consequently, its effective utilization.

9. Acknowledgements

We thank the Office of International Research Programs (OIRP) of United States Department of Agriculture (USDA) for continual funding of our collaborative research on cotton germplasm characterization and genetic diversity analysis. We acknowledge Civilian Research Development Foundation (CRDF), USA for project coordination and Academy of Sciences of Uzbekistan for their continual in-house support of the research efforts.

10. References

Abdalla, A.M.; Reddy, O.U.K.; El-Zik, K.M. & Pepper, A.E. (2001). Genetic Diversity and Relationships of Diploid and Tetraploid Cottons Revealed Using AFLP. *Theoretical and Applied Genetics*, Vol. 102, No. 2-3, (February 2001), pp. 222-229, ISSN 1432-2242

Abdukarimov, A.; Djataev, S. & Abdurakhmonov, I.Y. (2003). Cotton Research in Uzbekistan: Elite Varieties and Future of Cotton Breeding, *Proceedings of World Cotton Research Conference-3*, pp. 5-16, Cape Town, South Africa, March 9-13, 2003

Abdullaev, A.A.; Klyat, V.P. & Rizaeva, C.M. (2009). Cotton Introduction in Uzbekistan - History and Perspectives of Using of Plant Introduction: Problems and Perspectives. (2009), *Proceedings of 4th National scientific-applied conference*, pp. 59-61, Tashkent, Uzbekistan, July 3-4, 2009 (in Russian)

Abdurakhmonov, I.Y. (2007). Exploiting Genetic Diversity, *Proceedings of World Cotton Research Conference-4*, P2153, Lubbock, Texas, USA, September 10-14, 2007

Abdurakhmonov, I. Y. & Abdukarimov A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics,* Vol. 2008, No. 574927, (June 2008), pp.1-18, ISSN 1687-5389

Abdurakhmonov, I.Y.; Abdullaev, A.; Rizaeva, S.; Buriev, Z.; Adylova, A.; Abdukarimov, A.; Saha, S.; Kohel, R.; Yu, J. & Pepper, A.E. (2004). Evaluation of *G. hirsutum* Exotic Accessions From Uzbek Cotton Germplasm Collection for Further Molecular Mapping Purposes, *Proceedings Beltwide Cotton Improvement Conference,* pp. 1133-1142, San Antonio, Texas, USA, January 5-9, 2004

Abdurakhmonov, I.Y.; Abdullaev, A.A.; Saha, S.; Buriev, Z.T.; Arslanov, D.; Kuryazov, Z.; Mavlonov, G.T.; Rizaeva, S.M.; Reddy, U.K.; Jenkins, J.N.; Abdullaev, A. & Abdukarimov, A. (2005). Simple Sequence Repeat Marker Associated With a Natural Leaf Defoliation Trait in Tetraploid Cotton. *Journal of Heredity*, Vol. 96, No. 6, (September 2005), pp. 644-653, ISSN 1465-7333

Abdurakhmonov, I.Y.; Buriev, Z.T.; Salakhuddinov, I.B.; Rizaeva, S.M.; Adylova, A.T. Shermatov, S.E.; Adukarimov, A.; Kohel, R.J.; Yu, J.Z.; Pepper, A.E .; Saha, S. & Jenkins, J.N. (2006). Characterization of *G. hirsutum* Wild and Variety Accessions

from Uzbek Cotton Germplasm Collection for Morphological and Fiber Quality Traits and Database Development, *Proceedings Beltwide Cotton Improvement Conference,* P5306, San Antonio, Texas, USA, January 3-6, 2006

Abdurakhmonov, I.Y.; Kohel, R.J.; Yu, J.Z.; Pepper, A.E.; Abdullaev, A.A.; Kushanov, F.N.; Salakhutdinov, I.B.; Buriev, Z.T.; Saha,S.; Scheffler, B.E.; Jenkins, J.N. & Abdukarimov, A. (2008). Molecular Diversity and Association Mapping of Fiber Quality Traits in Exotic *G. hirsutum* L. Germplasm. *Genomics,* Vol. 92, No. 6, (October 2008), pp.478-487, ISSN 0888-7543

Abdurakhmonov, I.Y.; Saha, S.; Jenkins, J.N.; Buriev, Z.T.; Shermatov, S.E.; Scheffler, B.E.; Pepper, A.E.; Yu, J.Z.; Kohel, R.J. & Abdukarimov, A. (2009). Linkage Disequilibrium Based Association Mapping of Fiber Quality Traits in G. hirsutum L. Variety Germplasm. *Genetica*, Vol. 136, No. 3, (July 2009), pp. 401-417, ISSN *1573-6857*

*Abdurakhmonov, I.Y.; Buriev, Z.T.; Shermatov, Sh.E.; Kushanov, F.N.; Makamov, A.; Shopulatov, U.; Turaev, O.; Norov, T.; Akhmedov, Ch.; Mirzaakhmedov, M. & Abdukarimov, A. (2011). Utilization of Natural Diversity in Upland Cotton (*G. hirsutum*) Germplasm Collection for Pyramiding Genes via Marker-Assisted Selection Program,* Proceedings of 5th meeting of Asian Cotton Research and Development Network*, Lahore, Pakistan, February 23-25, 2011*

Adams, K.L.; Percifield, R. & Wendel, J.F. (2004). Organ-specific Silencing of Duplicated Genes in a Newly Synthesized Cotton Allotetraploid. *Genetics,* Vol. 168, No. 4, (December 2004), pp. 2217-2226, ISSN 0016-6731

Alvarez, I. & Wendel, J.F. (2006). Cryptic Interspecific Introgression and Genetic Differentiation Within *Gossypium aridum* (Malvaceae) and Its Relatives. *Evolution*, Vol. 60, No. 3, (March 2006), pp. 505-517, ISSN 1558-5646

Arpat, A.B.; Waugh, M.; Sullivan, J.P.; Gonzales, M.; Frisch, D.; Main, D.; Wood, T.; Leslie, A.; Wing, R.A. & Wilkins, T.A. (2004). Functional Genomics of Cell Elongation in Developing Cotton Fibers. *Plant Molecular Biology*, Vol. 54, No. 6, (April 2004), pp. 911–929, ISSN 1573-5028

Azmat, M.A. & Khan, A.A. (2010). Assessment of Genetic Diversity among the Varieties of Gossypium arboreum and Gossypium hirsutum through Random Amplification of Polymorphic DNA (RAPD) and Simple Sequence Repeat (SSR) Markers. *Pakistan Journal of Botany*, Vol. 42, No. 5, (October 2010), pp. 3173-3181, ISSN 0556-3321

Barroso, P.A.V.; Hoffman, L.V.; de Freitas, R.B.; de Batista, A. C.E.; Alves, M.F.; Silva, U.C. & de Andrade, F.P. (2010). In situ Conservation and Genetic Diversity of Three Populations of *Gossypium mustelinum* Myers ex Watt. *Genetic Resources and Crop Evolutio*n, Vol. 57, No. 3, (March 2010), pp. 343-349, ISSN 0925-9864

Bertini, C. H. C. M.; Schuster, I.; Sediyama, T.; Barros E.G. & Moreira, M.A. (2006). Characterization and Genetic Diversity Analysis of Cotton Cultivars Using Microsatellites. *Genetics* and Molecular Biology, Vol. 29, No. 2, (2006), pp. 321-329, ISSN 1415-4757

Blenda, A.J.; Scheffler, B.; Scheffler, M.; Palmer, J.M.; Lacape, J.Z.; Yu, C.; Jesudurai, S.; Jung, S.; Muthukumar, P.; Yellambalase, S.; Ficklin, M.; Staton, R.; Eshelman, R.; Ulloa, M.; Saha, S.; Burr, B.; Liu, S.; Zhang, T.; Fang, D.; Pepper, A.; Kumpatla, S.; Jacobs, J.; Tomkins, J.; Cantrell, R. & Main, D. (2006). CMD: A Cotton Microsatellite Database Resource for *Gossypium* Genomics. *BMC Genomics*, Vol. 7, No. 1, (December 2006), pp. 132, ISSN 1471-2164

Boopathi, N.M.; Gopikrishnan, A.; Selvam, N.J.; Ravikesavan, R.; Iyanar, K.; Muthuraman, S. & Saravanan, N. (2008). Genetic Diversity Assessment of G. barbadense Accessions to Widen Cotton (*Gossypium* spp.) Gene Pool for Improved Fibre Quality. *Journal of Cotton Research and Development*, Vol. 22, No. 2, (July 2008), pp. 135-138, ISSN 0972-8619

Brubaker, C.L.; Paterson, A.H. & Wendel, J.F. (1999). Comparative Genetic Mapping of Allotetraploid Cotton and Its Diploid Progenitors. *Genome*, Vol. 42, No. 2, (April 1999), pp. 184-203, ISSN 1480-3321

Campbell, B.T.; Saha, S.; Percy, R.; Frelichowski, J.; Jenkins, J.N.; Park, W.; Mayee, C.D.; Gotmare, V.; Dessauw, D.; Gband, M.; Du, X.; Jia, Y.; Constable, G.; Dillon, S.; Abdurakhmonov, I.Y.; Abdukarimov, A.; Rizaeva, S.M.; Abdullaev, A.A.; Barrose, P.A.V.; Padua, J.G.; Hoffman, L.V. & Podolnaya, L. (2010). Status of Global Cotton Germplasm Resources. *Crop Science*, Vol. 50, No. 4, (July 2010), pp. 1161-1179, ISSN 1435-0653

Chaudhary, L.; Sindhu, A.; Kumar, M. & Saini, M. (2010). Estimation of Genetic Divergence Among Some Cotton Varieties by RAPD Analysis. Journal of Plant Breeding and Crop Science, Vol. 2, No. 3, (March 2010), pp. 039-043, ISSN 2006-9758

Chen, Z.J.; Scheffler, B.E.; Dennis, E.; Triplett, B.A.; Zhang, T.; Guo, W.; Chen, X.; Stelly, D.M.; Rabinowicz, P.D.; Town, C.D.; Arioli, T.; Brubaker, C.; Cantrell, R.G.; Lacape, J.M.; Ulloa, M.; Chee, P.; Gingle, A.R.; Haigler, C.H.; Percy, R.; Saha, S.; Wilkins, T.; Wright, R.J.; Van Deynze, A.; Zhu, Y.; Yu, S.; Abdurakhmonov, I.; Katageri, I.; Kumar, P.A.; Mehboob-Ur-Rahman, Z.Y.; Yu, J.Z.; Kohel, R.J.; Wendel, J.F. & Paterson, A.H. (2007). Toward Sequencing Cotton (*Gossypium*) Genomes. *Plant Physioogyl*, Vol. 145, No. 4, (December 2007), pp. 1303-1310, ISSN 1532-2548

De Almeida, V.C.; Hoffman, L.V.; Yokomizo G.K.I.; da Costa, J.N.; Giband, M. & Barroso, P.A.V. (2009). In situ Genetic Characterization of *Gossypium barbadense* Populations from the States of Para and Amapa, Brazil. *Pesquisa Agropecuaria Brasileira*, Vol. 44, No. 7, (July 2009), pp. 719-725, ISSN 0100-204X

Dejoode, D.R. & Wendel, J.F. (1992). Genetic Diversity and Origin of the Hawaiian-Islands Cotton, *Gossypium tomentosum*. *American Journal of Botany,* Vol. 79, No. 11, (November 1992), pp. 1311-1319, ISSN 0002-9122

Dighe, N.D.; Robinson, A.F.; Bell, A.A.; Menz, M.A.; Cantrell, R.G & Stelly, D.M. (2009). Linkage Mapping of Resistance to Reniform Nematode in Cotton Following Introgression from *Gossypium longicalyx* (Hutch.& Lee). *Crop Science*, Vol. 49, No. 4, (July-August 2009), pp.1151-1164, ISSN 0011-183X

Endrizzi, J.; Turcotte, E. & Kohel, R. (1985). Genetics, Cytology and Evolution of *Gossypium*. *Advances in Genetics*, Vol. 23, pp. 271-375, ISBN 9780120176236

Esbroeck, G. & Bowman, D. (1998). Cotton Germplasm Diversity and its Importance to Cultivar Development. *Journal of Cotton Science*, Vol. 2, No. 3, (July 1998), pp. 121-129, ISSN 1523-6919.

Feng, Ch.; Ulloa, M.; Perez-M, C. & Stewart, J.McD. (2011). Distribution and Molecular Diversity of Arborescent *Gossypium* Species. *Botany*, Vol. 89, No. 9, (September 2011), pp.615-624, ISSN 1916-2790

Fryxell, P.A. (1979). Natural History of the Cotton Tribe. Texas A&M University Press, College Station, Texas, USA

Fryxell, P.A. (1992). A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheedea*, Vol. 2, No. 1, (January 1992), pp. 108-165, ISSN: 0971-2313

Gingle, A.R.; Yang, H; Chee, P.W.; May, O.L.; Rong, J.; Bowman, D.T.; Lubbers, E.L.; Day, J.L.; & Paterson, A.H. (2006). An Integrated Web Resource for Cotton. *Crop Science*, Vol. 46, No. 5, (September 2006), pp. 1998-2007, ISSN 1435-0653

Guo, W.; Sang, Z.Q.; Zhou, B.L. & Zhang, T. (2007). Genetic Relationships of D-genome Species Based on Two Types of EST-SSR Markers Derived from *G. arboreum* and *G. raimondii* in *Gossypium*. *Plant Science,* Vol. 172, No. 4, (April 2007), pp. 808-814, ISSN 0168-9452

Guo, W.Z.; Zhou, B.L.; Yang, L.M.; Wang, W. & Zhang, T.Z. (2006). Genetic Diversity of Landraces in Gossypium arboretum L. Rase *sinense* Assessed with Simple Sequence Repeat Markers. *Journal of Integrative Plant Biology*, Vol. 48, No. 9, (September 2006), pp. 1008-1017, ISSN 1744-7909

Gutierrez, O.A.; Basu, S.; Saha, S.; Jenkins, J.N.; Shoemaker, D.B.; Cheatham, C.L. & McCarty, J.C. (2002). Genetic Distance Among Selected Cotton Genotypes and Its Relationship With F2 Performance. *Crop Science*, Vol. 42, No. 6, (November–December 2002), pp. 1841-1847, ISSN 1435-0653

Hawkins, J.S.; Pleassants, J. & Wendel, J.F. (2005). Identification of AFLP Markers That Discriminate Between Cultivated Cotton and the Hawaiian Island Endemic, *Gossypium tomentosum* Nuttall ex Seeman. *Genetic Resources* and *Crop Evolution*, Vol. 52, No. 8, (December 2005), pp. 1069-1078, ISSN 1573-5109

Hinze, L.L.; Kohel, R.J.; Campbell, B.T. & Percy, R.G. (2011). Variability in Four Diverse Cotton (Gossypium hirsutum L.) Germplasm Populations. *Genetic Resources and Crop Evolution*, Vol. 58, No. 4, (April 2011), pp. 561-570, ISSN 0925-9864

Ibragimov, P.S.; Avtonomov, V.A.; Amanturdiev, A.B.; Namazov, S.E.; Zaurov, D.E.; Molnar, T.J.; Eisenman, S.W.; Orton, T.J.; Funk, C.R.; Percival, J. & Edward, A. (2008). Uzbek Scientific Research Institute of Cotton Breeding and Seed Production: Breeding and Germplasm Resources. *Journal of Cotton Science,* Vol. 12, No. 2, (April 2008), pp. 62-72, ISSN 1524-3303

International Cotton Advisory Committee [ICAC]. (2011). Small Increase in Global cotton consumption expected in 2011/2012. Press Release, (September 2011), Washington, DC, USA.

Iqbal, J.; Reddy, O.U.K.; El-Zik, K.M. & Pepper, A.E. (2001). A Genetic Bottleneck in the 'Evolution Under Domestication' of Upland Cotton *Gossypium hirsutum L.* Examined Using DNA Fingerprinting. *Theoretical and Applied Genetics*, Vol. 103, No. 4, (September 2001), pp. 547-554, ISSN 1432-2242

Iqbal, M.J.; Aziz, N.; Saeed, N.A.; Zafar, Y. & Malik, K.A. (1997). Genetic Diversity Evaluation of Some Elite Cotton Varieties by RAPD Analysis. *Theoretical and Applied Genetics*, Vol. 94, No. 1, (July 1997), pp. 139-144, ISSN 1432-2242

Kalivas, A.; Xanthopoulos, F.; Kehagia, O. & Tsaftaris, A.S. (2011). Agronomic Characterization, Genetic Diversity and Association Analysis of Cotton Cultivars using Simple Sequence Repeat Molecular Markers. *Genetics and Molecular Research*, Vol. 10, No. 1, (February 2011), pp. 208-217, ISSN 1676-5680

Kalivas, A.; Xanthopoulos, F.; Kehagia, O. & Tsaftaris, A.S. (2011). Agronomic Characterization, Genetic Diversity and Association Analysis of Cotton Cuitivars

using Simple Sequence Repeat Molecular Markers. *Genetics and Molecular Research*, Vol. 10, No. 1, (February 2011), pp. 208-217, ISSN 1676-5680

Kantartzi, S.K. & Stewart, J.McD. (2008). Association Analysis of Fibre Traits in *Gossypium arboreum* Accessions. *Plant Breeding,* Vol. 127, No. 2, (April 2008), pp. 173-179, ISSN 1439-0523

Kantartzi, S.K.; Ulloa, M.; Sacks, E. & Stewart, J.M. (2009). Assessing Genetic Diversity in Gossypium arboreum L. Cultivars Using Genomic and EST-derived Microsatellites. *Genetica*, Vol. 136, No.1, (May 2009), pp. 141-147, ISSN *1573-6857*

Kebede, H.; Burow, G.; Dani, R.G. & Allen, R.D. (2007). A-genome Cotton as a Source of Genetic Variability for Upland Cotton (*Gossypium hirsutum*). *Genetic Resources and Crop Evolution,* Vol. 54, No. 4, (June 2007), pp. 885-895, ISSN 1573-5109

Khan, A.I.; Fu, Y. & Khan, I. (2009). Genetic Diversity of Pakistani Cotton Cultivars as Revealed by Simple Sequence Repeat Markers. *Communications in Biometry and Crop Science*, Vol. 4, No. 1, (2009), pp. 21-30, ISSN 1896-0782

Kump, K.L.; Bradbury, P.J.; Wisser, R.J.; Buckler, E.S.; Belcher, A.R.; Oropeza-Rosas, M.A.; Zwonitzer, J.C.; Kresovich, S.; McMullen, M.D.; Ware, D.; Balint-Kurti, P.J. & Holland, J.B. (2011). Genome-Wide Association Study of Quantitative Resistance to Southern Leaf Blight in the Maize Nested Association Mapping Population. *Nature Genetics*, Vol. 43, No. 2, (January 2011), pp. 163-168, ISSN 1546-1718

Lacape, J.M.; Dessauw, D.; Rajab, M.; Noyer, J.L. & Hau, B. (2007). Microsatellite Diversity in Tetraploid *Gossypium* germplasm: Assembling a Highly Informative Genotyping Set of Cotton SSRs. *Molecular Breeding,* Vol. 19, No. 1, (January 2007), pp. 45-58, ISSN 1572-9788

Liu, D.; Guo, X.; Lin, Z.; Nie, Y. & Zhang, X. (2006). Genetic Diversity of Asian Cotton (*Gossypium arboreum L.*) in China Evaluated by Microsatellite Analysis. *Genetic Resources and Crop Evolution,* Vol. 53, No. 6, (September 2006), pp. 1145-1152, ISSN 1573-5109

Liu, S.; Cantrell, R.G.; McCarty, J.C.J. & Stewart, J.M. (2000). Simple Sequence Repeat-based Assessment of Genetic Diversity in Cotton Race Stock Accessions. *Crop Science*, Vol. 40, No. 5, (September 2000), pp. 1459-1469, ISSN 1435-0653

Liu, Z.; Zhang, Y.; Zhang, S.; Deng, K.; Dong, S. & Ren, Z. (2011). Genetic Diversity Analysis of Various Red Spider Mite-Resistant Upland Cotton Cultivars Based on RAPD. *African Journal of Biotechnology*, Vol. 10, No. 18, (May 2011), pp. 3515-3520, ISSN 1684-5315

Luconge, E.; Herselman, L. & Labuschagne, M.T. (2007). Genetic Diversity of Tanzanian Cotton (Gossypium hirsutum L.) Revealed by AFLP Analysis, *Proceedings of African Crop Science Conference*, Vol. 8, pp. 773-776, ISSN 1023-070X, El Minia, Egypt, October 27-31, 2007

Mahmood, Z.; Raheel, F.; Dasti, A.A.; Shahzadi, S.; Athar, M. & Qayyum, M. (2009). Genetic Diversity Analysis of the Species of Gossypium by using RAPD Markers. *African Journal of Biotechnology*, Vol. 8, No. 16, (August2009), pp. 3691-3697, ISSN 1684-5315

Mauer, F. M. (1954). Origin and Taxonomy of Cotton, In: *Cotton*, 383, Academy of Sciences of USSR, Tashkent, Uzbekistan (In Russian)

McCarty, J.C.Jr. & Jenkins, J.N. (1993). Registration of 79 Day-neutral Primitive Cotton Germplasm Lines. Crop Science. Vol. 33, No. 2, (March-April 1993), pp. 351, ISSN 1435-0653

McCarty, J.C.Jr.; Jenkins, J.N.; Parrott, W.L. & Greech, R.G. (1979). The Conversion of Photoperiodic Primitive Race Stocks of Cotton to Day-neutral Stocks. *Mississippi Agricultural & Forestry Experiment Station Research Report*, Vol. 4, No. 19, (July 1979), pp. 1-4, ISSN 0147-2186

Multani, D.S. & Lyon, B.R. (1995). Genetic Fingerprinting of Australian Cotton Cultivars With RAPD Markers. *Genome*, Vol. 38, No. 5, (October 2005), pp. 1005-1008, ISSN 1480-3321

Patel, J.P.; Fougat, R.S. & Jadeja, G.C. (2009). Genetic Diversity Analysis in some Elite Desi Cotton Cultivars of Gossypium herbaceum and G. arboreum and Genetic Purity Testing of Their Hybrids through Microsatellite Markers. *International Journal of Plant Sciences*, Vol. 4, No. 2, (May2009), pp. 464-470, ISSN 0973-1547

Paterson, A.; Boman, R.; Brown, S.; Chee, P.; Gannaway, J.; Gingle, A.; May, O. & Smith, C. (2004). Reducing the Genetic Vulnerability of Cotton. *Crop Science*, Vol. 44, No. 6, (November-December 2004), pp. 1900-1901, ISSN 1435-0653.

Percival, A.E.; Stewart, J.M. & Wendel, J.F. (1999). Taxonomy and Germplasm Resources, In: *Cotton: Origin, History, Technology and Production*, C.W. Smith and J.T. Cothren, (Ed), 33-63, ISBN 978-0-471-18045-6 John Wiley, New York.

Phillips, L.L. (1964). Segregation in New Allopolyploids of *Gossypium*. V. Multivalent Formation in New World x Asiatic and New World x wild American Hexaploid. *American Journal of Botany*, Vol. 51, No. 3, (March 1964), pp. 324-329, ISSN 0002-9122

Pillay, M. & Myers, G.O. (1999). Genetic Diversity in Cotton Assessed by Variation in Ribosomal RNA Genes and AFLP Markers. *Crop Science*, Vol. 39, No. 6, (November 1999), pp. 1881-1886, ISSN 1435-0653

Poland, J.A.; Bradbury, P.J.; Buckler, E.S. & Nelson, R.J. (2011). Genome-Wide Nested Association Mapping of Quantitative Resistance to Northern Leaf Blight in Maize. *Proceedings of the National Academy of Sciences of the USA*, Vol. 108, No. 17, (April 2011), pp. 6893-6898, ISSN 0027-8424

Rahman, M.; Yasmin, T.; Tabassam, N.; Ullah, I.; Asif, M. & Zafar, Y. (2008). Studying the Extent of Genetic Diversity among Gossypium arboreum L. Genotypes/Cultivars using DNA fingerprinting. *Genetic Resources and Crop Evolution*, Vol. 55, No. 3, (May2008), pp. 331-339, ISSN 0925-9864

Rana, M.K.; Singh, V.P. & Bhat, K.V. (2005). Assessment of Genetic Diversity in Upland Cotton (*Gossypium hirsutum L.)* Breeding Lines by Using Amplified Fragment Length Polymorphism (AFLP) Markers and Morphological Characteristics. *Genetic Resources and Crop Evolution*, Vol. 52, No. 8, (December 2005), pp. 989-997, ISSN 1573-5109

Reddy, U.K.; Rong, J.; Nimmakayala, P.; Vajja, G.; Rahman, M.A.; Yu, J.; Soliman, Kh.M.; Heller-Uszynska, K.; Kilian, A. & Paterson, A.H. (2011). Use of Diversity Arrays Technology Markers for Integration into a Cotton Reference Map and Anchoring to a Recombinant Inbred Line Map. *Genome*, Vol. 54. No. 5, (May 2011), pp. 349-359, ISSN 0831-2796

Roberts, P.A., Ulloa, M. (2010). Introgression of Root-knot Nematode Resistance into Tetraploid Cotton. *Crop Science*, Vol.50, No.3, (May 2010), pp. 940–951, ISSN 1435-0653

Robinson, A.F.; Bell, A.A.; Dighe, N.D.; Menz, M.A.; Nichols, R.L. & Stelly, D.M. (2007). Introgression of Resistance to Nematode *Rotylenchulus reniformis* into Upland

Cotton (*Gossypium hirsutum*) from *Gossypium longicalyx*. *Crop Science*, Vol. 47, No. 5, (September-October 2007), pp. 1865-1877, ISSN 0011-183X

Ruan, Y.L.; Llewellyn, D.J. and Furbank, R.T. (2003). Suppression of Sucrose Synthase Gene Expression Represses Cotton Fiber Cell Initiation, Elongation, and Seed Development. *The Plant Cell*, Vol. 15, No. 4, (April 2003), pp. 952-964, ISSN 1531-298X

Rungis, D.; Llewellyn, D.; Dennis, E.S. & Lyon, B.R. (2005). Simple Sequence Repeat (SSR) Markers Reveal Low Levels of Polymorphism Between Cotton (*Gossypium hirsutum L.*) Cultivars. *Australian Journal of Agricultural Research*, Vol. 56, No. 3, (March 23, 2005), pp. 301-307, ISSN 1836-5795

Sapkal, D.R.; Sutar, S.R.; Thakre, P.B.; Patil, B.R.; Paterson, A.H. & Waghmare, V.N. (2011). Genetic Diversity Analysis of Maintainer and Restorer Accessions in Upland Cotton (*Gossypium hirsutum* L.). *Journal of Plant Biochemistry and Biotechnology*, Vol.20, No.1, (January-June 2011), pp. 20-28, ISSN 0971-7811

Stelly, D.M.; Lacape, J.M.; Dessauw, D.E.G.A.; Kohel, R.; Mergeai, G.; Saha., S.; Sanamyan, M.; Abdurakhmonov, I.Y.; Zhang, T.; Wang, K.; Zhou, B. & Frelichowski, J. (2007). International Genetic, Cytogenetic and Germplasm Resources for Cotton Genomics and Genetic Improvement, *Proceedings World Cotton Research Conference-4*, P. 2154, ISBN, Lubbock, Texas, USA, September 10-14, 2007

Stich, B. & Melchinger, A.E. (2010). An Introduction to Association Mapping in Plants. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, Vol. 5, No. 039, (May 2010), ISSN 1749-8848

Sunilkumar, G.; Campbell, L.M.; Puckhaber, L.; Stipanovic, R.D. & Rathore K.S. (2006). Engineering Cottonseed for Use in Human Nutrition by Tissue-specific Reduction of Toxic Gossypol. *Proceedings of the National Academy of Sciences of the USA.* Vol. 103, No. 48, (November 2006), pp. 18054-18059, ISSN 0027-8424.

Tatineni, V.; Cantrell, R.G. & Davis, D.D. (1996). Genetic Diversity in Elite Cotton Germplasm Determined by Morphological Characteristics and RAPD. *Crop Science,* Vol. 36, No. 1, (January-February 1996), pp. 186-192, ISSN 1435-0653

Tian, F.; Bradbury, P.J.; Brown, P.J.; Hung, H.; Sun, Q.; Flint-Garcia, S.; Rocheford, T.R.; McMullen, M.D.; Holland, J.B. & Buckler, E.S. (2011). Genome-Wide Association Study of Leaf Architecture in the Maize Nested Association Mapping Population. *Nature Genetics*, Vol. 43, No. 2, (January 2011), pp. 159-162, ISSN 1546-1718

Tiwari, R.S. & Stewart, J.McD. (2008). Molecular Diversity and Determination of Hybridization Among the G-genome *Gossypium* Species, In: *Summaries of Arkansas Cotton Research* 2008, Research Series 573, (October 2008), D.M. Oosterhuis, (Ed.), pp. 30-33, ISSN 1942-160X

Ulloa, M.; Stewart, J.M.; Garcia-C, E.A.; Goday, A. S.; Gaytan-M, A. & Acosta, N.S. (2006). Cotton Genetic Resources in the Western States of Mexico: *in situ* Conservation Status and Germplasm Collection for *ex situ* Preservation. *Genetic Resources* and *Crop Evolution,* Vol. 53, No. 4, (June 2006), pp. 653-668, ISSN 1573-5109

Ulloa, M.; Brubaker, C.; & Chee, P. (2007). Cotton. In: *Genome Mapping & Molecular Breeding, Vol. 6: Technical Crops*. C. Kole (ed), pp.1-49. Springer, ISBN 9783540345374, Heidelberg, Berlin, New York, Tokyo

Ulloa, M.; Percy, R.; Zhang, J.; Hutmacher, R. B.; Wright S. D. & Davis., R. M. (2009). Registration of four Pima cotton germplasm lines having good levels of *Fusarium*

wilt race 4 resistance with moderate yields, and good fibers. *J. Plant Registr*. Vol. 3, No. 2, (May 2009), pp. 198-202, ISSN 1936-5209

Van Deynze, A.; Stoffel, K.; Lee, M.; Wilkins, T.A; Kozik, A.; Cantrell, R.G.; Yu, J. Z.; Kohel, R. J. & Stelly, D. M. (2009). Sampling nucleotide diversity in cotton. BMC Plant Biol., Vol. 9, (October 2009), pp.125, ISSN 1471-2229

Wallace, T.P.; Bowman, D.; Campbell, B.T.; Chee, P.; Gutierrez, O.A.; Kohel, R.J.; McCarty, J.; Myers, G.; Percy, R.; Robinson, F.; Smith,W.; Stelly, D.M.; Stewart, J.M.; Thaxton, P.; Ulloa, M. & Weaver, D.B. (2009). Status of the USA Cotton and Crop Vulnerability. *Genetic Resources and Crop Evolution*, Vol. 56, No. 4, (June 2009), pp. 507-532, ISSN 0925-9864

Wendel, J.F. (1995). Cotton, In: *Evolution of Crop Plants,* S. Simmonds and J. Smartt, (Ed), 358-366, 1st ed Longman, ISBN 978-0-582-08643-2, London

Wendel, J.F. & Albert, V.A. (1992). Phylogenetics of the Cotton Genus (*Gossypium*) - Character State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and Its Systematic and Biogeographic Implications. *Systematic Botany*, Vol. 17, No. 1, (January-March 1992), pp. 115-143, ISSN 1548-2324

Wendel, J.F. & Percy, R.G. (1990). Allozyme Diversity and Introgression in the Galapagos-Islands Endemic *Gossypium darwinii* and Its Relationship to Continental Gossypium *barbadense*. Biochemical Systematics and *Ecology,* Vol. 18, No. 7-8, (December 2002), pp. 517-528, ISSN 0305-1978

Wendel, J.F.; Brubaker, C.L. & Percival, A.E. (1992). Genetic Diversity in *Gossypium hirsutum* and the Origin of Upland Cotton. *American Journal of Botany*, Vol. 79, No. 11, (November 1992), pp. 1291-1310, ISSN 0002-9122

Wendel, J.F.; Rowley, R. & Stewart, J.M. (1994). Genetic Diversity in and Phylogenetic-relationships of Brazilian Endemic Cotton, *Gossypium mustelinum* (Malvaceae). *Plant Systematics and Evolution*, Vol. 192, No. 1-2, (March 1994), pp. 49-50, ISSN 1615-6110

Westengen, O.T.; Huaman, Z. & Heum, M. (2005). Genetic Diversity and Geographic Pattern in Early South American Cotton Domestication. *Theoretical and Applied Genetics*, Vol. 110, No. 2, (January 2005), pp. 392-402, ISSN 1432-2242

Wu, D.; Fang, X.; Ma, M.; Chen, J. & Zhu, S. (2010). Genetic Relationship and Diversity of the Germplasms in Gossypium barbadense L. from Four Different Countries Using SSR Markers. *Cotton Science,* Vol. 22, No. 2, (March 2010), pp. 104-109, ISSN 1002-7807

Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh, B.I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; Kresovich. S. & Buckler, E.S. (2006). A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness. *Nature Genetics*, Vol. 38, No. 2, (February 2006), pp. 203-208, ISSN 1546-1718

Zeng L.; Meredith, W.R.Jr.; Gutierrez, O.A. & Boykin, D.L. (2009). Identification of Associations Between SSR Markers and Fiber Traits in an Exotic Germplasm Derived From Multiple Crosses Among *Gossypium* Tetraploid Species. *Theoretical and Applied Genetics*, Vol. 119, No. 1, (June 2009), pp. 93-103, ISSN 1432-2242

Zhang, H.B.; Li, Y.; Wang, B & Chee, P.W. (2008). Recent Advances in Cotton Genomics. *International Journal of Plant Genomics,* Vol. 2008, No. 742304, (January 2008), pp. 1-20, ISSN 1687-5389

Zhang, J.; Lu, Y.; Cantrell, R.G. & Hughs, E. (2005a). Molecular Marker Diversity and Field Performance in Commercial Cotton Cultivars Evaluated in the Southwestern USA. *Crop Science,* Vol. 45, No. 4, (July 2005), pp. 1483-1490, ISSN 1435-0653

Zhang, J.F.; Lu, Y.Z. & Yu, S.X. (2005b). Cleaved AFLP (cAFLP), a Modified Amplified Fragment Length Polymorphism Analysis for Cotton. *Theoretical and Applied Genetics*, Vol. 111, No. 7, (November 2005), pp. 1385-1395, ISSN 1432-2242

Zhang, Y.; Wang, X.F.; Li, Z.K.; Zhang, G.Y. & Ma, Z.Y. (2011a). Assessing Genetic Diversity of Cotton Cultivars using Genomic and Newly Developed Expressed Sequence Tag-Derived Microsatellite Markers. *Genetics and Molecular Research*, Vol. 10, No. 3, (July 2011), pp. 1462-1470, ISSN 1676-5680

Zhang, M.; Zheng, X.; Song, S.; Zeng, Q.; Hou, L.; Li, D.; Zhao, J.; Wei, Y.; Li, X.; Luo, M.; Xiao, Y.; Luo, X.; Zhang, J.; Xiang, C. & Pei, Y. (2011b). Spatiotemporal Manipulation of Auxin Biosynthesis in Cotton Ovule Epidermal Cells Enhances Fiber Yield and Quality. *Nature Biotechnology*, Vol. 29, No. 5, (April 2011), pp. 453-458, ISSN 1546-1696

17

Exploring Statistical Tools in Measuring Genetic Diversity for Crop Improvement

C. O. Aremu
Department of Crop Production and Soil Science
Ladoke Akintola University of Technology, Ogbomoso, Oyo state
Landmark University, Omu-Aran
Nigeria

1. Introduction

Increase in global numerical population especially in developing nations has gradually led to food shortage and hence increase in poverty. Addressing and tackling the issue and causes of poverty in the developing nations is one major challenge to breeders (Fu and Somers 2009). The different theories of econometircs have identified the human and material resources traceable to poverty, but fail to identify the crop improvement techniques in addressing world food shortage (Baudoin and Mergeai 2001). Crop improvement techniques therefore remains a major concern to plant breeders (Akbar and Kamran, 2006; Aremu *et al,* 2007a). Several factors affect crop improvement for specific or general environment performance. Such factors include climate, weather, soil, edaphic and biological and more importantly crop genotype (Aremu, et al, 2007b). Crop genotypes are composed of different crop forms including inbred or pure lines hybrids, landraces, wildraces germplasm accessions, cultivars or varieties. These crop genotypes have wide and diverse origin and genetic background known as genetic diversity. Genetic diversity study is a major breakthrough in understanding intraspecie crop performance leading to crop improvement (Aremu, 2005). Knowledge of crop performance in genetic diverse population reveals the differences in the nature of genetic materials used.

Genetic diversity studies therefore, is a step wise process through which existing variations in the nature of individual or group of individual crop genotypes are identified using specific statistical method or combination of methods (Christini *et al.* 2009; Warburton and Crossa 2000; Aremu, 2005; Weir 1996). It is expected that the identified variations would form a pattern of genetic relationship useable in grouping genotypes.

Several researchers including breeders have employed different data source and type from diverse crops in their methods to study genetic diversity. Such data source include morphological and agronomic, pedigree, proximate or biochemical and molecular data (Aremu, *et al.,* 2007a in cowpea; Liu *et al*., 2000 in cotton; Mostafa *et al.,* 2011 in wheat; Adewale *et al*., 2010 in African Yam bean; Christine *et al*., 2009 in bentgrass.

The choice of statistical method to be used is dependent on the achievable objectives laid out in the studies. This chapter reveals the underlying importance of genetic diversity and

reviews useable statistical techniques for identifying and grouping genotypes for intraspecies crop improvement.

2. Need for germplasm resource in genetic diversity preservation

Crop genotypes sourced as germplasm accessions, landraces, breeding lines, wild species, have rich and variable genetic integrity explorable for breeding programmes. The first step of any meaningful breeding programme is to identify crop plants that exhibit exploitable variation for the trait(s) of interest. However, these genetic diverse crops are under threat. Continuous hybridization and crossing systems have reduced the genetic variations in cropping programmes and leave a dearth in harvesting and utilization of novel crop types with exploitable traits. Also, the continuous threat or loss of genetic diversity as a result of replacement of landraces, wild species and other primitive term of crop species by exotic high- yielding varieties remains an insurmountable problem to plant breeders. Another major source of loss of genetic diversity is by changes and or increase in population size, resulting in land use acts promoting deforestation, wars, industrialization, urbanization and other factors. According to Brown (1989), preservation of genetic diversity is possible when genetic or germplasm resource is realized as the most precious asset in conserving genetic diversity. Germplasm therefore is an essential resource for successful plant breeding. Certain areas of the world exhibit high level of genetic variability for crops (Vavilov, 1950). Falconer and Mackay (1996); Eivazi *et al;* (2007); reported that such areas are considered as regions or center of genetic diversity. Therefore genetic diversity in crop may be associated with the origin of the crop. This is supported by Christine *et al.* (2009), who reported genetic diversity to be associated with origin. Potter and Doyle, (1992) reported Tropical Africa to be the centre of diversity for African yam bean. Van Bueningen and Busch (1997), reported genetic diversity of wheat to be centered in North America. Ariyo and Odulaja (1991), found correlation between genetic diversity and eco-geographic background in okro. Some grouping methods in genetic diversity studies identified origin and geographical diversity not important in measuring genetic diversity. Nair *et al*. (1998) discovered diversity in sugarcane not to be associated with origin. Aremu *et al*. (2007a), discovered that center of origin is not a measure of genetic diversity in cowpea. If crop origin is somewhat not important in the measure of genetic diversity a resource centre is therefore needed to preserve and maintain the wide genetic sources exploitable in breeding programmes. Genetic relationship and diversity are useful for developing germplasm conservation strategies and utilization of crop genetic resources. The use of genetic diversity resource centre cannot be under estimated as earlier discussed.

3. Importance of genetic diversity studies

Study on genetic diversity is critical to success in plant breeding. It provides information about the quantum of genetic divergence and serves a platform for specific breeding objectives (Thompson et al, 1998). It identifies parental combinations exploitable to create segregating progenies with maximum genetic potential for further selection, as proven by Akoroda (1987), Weir, (1996), Liu *et al*.(2000); Dje *et al*.(2000), (Aremu *et al*, 2007b). Genetic diversity exposes the genetic variability in diverse populations and provides justification for introgression and ideotype breeding programmes to enhance crop performance. Mostafa *al et.* (2011), postulated that genetic diversity studies provides the understanding of genetic relationships among populations and hence directs assigning lines to specific heterogeneous

groups useable in identification of parents and hence choice selection for hybridization. Choice of parent has been identified to be the first basic step in meaningful breeding programme (Akoroda 1987); (Aremu et al. 2007a); (Islam 2004), (Rahim *et al,* 2010). Furthermore, the choice of parent selection in diversity studies is valuable because it is a means of creating useful variations in subsequent progenies.); Dje *et al.* (2000), discovered that the higher the genetic distance between parents, the higher the heterosis in the developed progenies. Hence the heterotic progenies can be further hybridized and selections based on transgressive segregation. Akbar and Kamran, (2006). exploited this parental selection technique in wheat breeding program through hybridization. Mostafa *et al*. (2011), investigated genetic distance among 36 winter wheat genotypes cultivated in different regions of Iran using principal component analysis and discovered five major groups in the genotypes to distantly related. Comprehensive and significant emphasis are made by researchers especially plant breeders on the analysis of genetic diversity in a number of field crops white and yellow yam, (Akoroda, 1987); cowpea, (Adewale and Aremu, 2010); African yam bean, (Baudoin and Mergeai 2001); Flax, (Mohammadi *et al.* 2010); wheat, (Mostafa *et al.* 2011) and several other crops.

The diversity studies on these crops at their respective primitive levels (Landrace, wildtype, accessions, lines *etc*) led to the development of their widely distributed cultivars and varieties with proven characteristics based on stability and adaptability of performance with consistent tolerance to adverse weather conditions and resistant to diseases around the world. Fu and Somers (2009) supported that the use of identified wheat parents resistant to environmental stress under different growing conditions has led to increased world wheat production. The early report of Mohammadi and Prasna (2003) revealed that appropriate parent selection for hybridization in maize using a definite diversity study technique, Bohn *et al* (1999), identified six groups of wheat land races in the Western Iran that can be grown in different geographical locations for improved yield. Martin *et al.,* (2008) discovered 42 cultivars of bentgrass in the mancet city and that only diversity studies would identify reliable and definite cultivar(s) with varietal purity and ensure protection of breeder and consumer rights. Understanding the inter and intra specie genetic relationships as provided by diversity studies has proven to increases hybrid vigor and reduce or avoid re-selection within existing germsplasm. It is worthy of note that existing cultivar populations have narrow genetic bases, hence need for creating variability within and among cultivars using genetic diversity methods.

4. Genetic diversity measurement tools

Genetic diverse populations arising from pure lines, accessions, landraces, wild or weed races are analyzed using a number of methods. Such method can be single or in combination of two or more methods. Franco *et al.* (2001) stressed the need for careful considerations to be made when measuring genetic diversity within and between crop populations in research. Such considerations include:

1. Use of multivariate data collected from morphological or agronomic traits. Such data may effectively display discrete, continuous, binomial ordinal etc. variables.
2. Use of multiple data sets arising from morphological, biochemical and DNA-based collections. The use of such multiple data sets in diversity study helps to reveal the adequacy in terms of strength and constraints in the choice of each of the data sets. The use of multiple data pose some puzzles including can analysis and result interpretation

be based on individual or combined data sets? And more worrisome is the puzzle on how to effectively combine the different data sets and still achieve meaningful result. To provide answers to these puzzles, Wrigley *et al.* (1982), studied phylogenetic relationships among triticeae species using individual and combined analysis of data sets consisting of morphological and DNA-based traits and discovered divergent results in the analysed individual and combined data. The discrepancies in the results may be attributed to the discrete nature of DNA-based data and the continuous variable nature of the morphological data. No wonder Hillis 1987; Chippindale and Wein (1994) suggested the assignment of specific numbers to both quantitative and qualitiative traits in morphological, biochemical and molecular data set. In view of this, Pedersen and Seberg (1998) advised that both individual and combined data sets can be analyzed in many possible and meaningful ways to draw conclusions on genetic divergence. In 1999 and 2001, Taba *et al.* and Franco et al., respectively utilized the modified Location Model (MLM) which combines all variables into one multinomial variable called "W" to classify maize accessions from the genetic resource centres of Latin America. Better still, this MLM can combine molecular and morphological data to classify data better than when individual data set is employed. Individual data from morphological, biochemical or molecular data set can be analyzed using one or a combination of techniques. These techniques shall be discussed.

3. Expected objective to be achieved. This dictates choice of statistical tool in measuring genetic distance and the level of clustering of the intragenic factors in use. Such objective(s) include to determine the quantum of variation and grouping such genotype based on genetic distance, identify action following parental selection. In essence, breeding focus determines applicable method in explaining the nature of genetic divergence.

Variations are recorded in the measurement of genetic diversity in genotype relationships based on genetic distances and grouping populations from individual genotypes such as accessions, lines, wild races etc. The recorded variations are primarily because of the differences in the nature of genetic materials. Therefore, the basis or genetic variance theories which identifies genotype relationships based on genetic distance estimating genetic diversity depends largely on statistical genetic variance theories which identifies genotype relationships based on genetic distance / variance.

5. The use of morphological data to measure genetic distance

Nei, (1973), first defined Genetic distance as the difference between two entities that can be described by allelic variation. This definition was later in 1987, modified to "extent of gene differences among populations that are measured using numerical values. Betterstill, in 1998, Beaumont *et al.,* provided a more comprehensive definition of genetic distance as any quantitative measure of genetic difference at either sequence or allele frequency level calculated between genotype individuals or populations.

The first early work of Anderson (1957), proposed the use of metrogliph and index-score to study the pattern of morphological variations in individual data set. In the early seventies (Singh and Chaudhary 1985) used this method to study morphological variation in green gram. This method uses a range of variations arising from trait such that extent of trait variation is determined by the length of rays on the glyph. The performance of a genotype is

adjudged by the value of the index score of that genotype. The score value determine the length of ray which may be small, medium or long Akoroda (1987); Ariyo and Odulaja (1991) and Van Bueningen and Busch (1997), extensively explored the use of metroglyph and index-score to morphological variations in yellow yam, Okro and wild rye accessions respectively.

Similar to metroglyph and the score index is Euclidian Distance (ED) measurement. According to Nei (1987), Euclidian distance measures similarity between two genotypes, populations or individuals using using statistical measures where two individuals *i* and *j*, having observations on morphological traits (*p)* denoted by $x_1, x_2, x_3, \ldots\ldots x_n$ and $y_1, y_2, \ldots\ldots y_n$ for *i* and *j* individuals respectively.

Metroglyph and index-score methods measures genetic distance by use of morphological traits. Euclidian distance measurements utilize both morphological and molecular based marker data sets. Smith *et al.* (1991), applied the following statistic to measure ED.

$$dij = \varepsilon[(T_{1(i)} - T_{2(i)}^2)/\sigma^2 T_{(i)}]^{1/2}.$$

Where T_1 and T_2 are the values of the ith trait for 1 lines and 2 and σ^2 T(i) is the variance for the ith trait over all the lines used. Much later, Weir (1996) developed a formula for calculating genetic distance to be.

$$d(I,j) = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots..(x_p - y_p)^2]^{1/2}$$

where i and j is the ED between two individuals lines having morphological traits (p)

$x_1, x_2 \ldots\ldots x_p$ is the traits for i individuals and
$y_1, y_2 \ldots\ldots x_p$ is the traits for j individuals

from here, the individual character distances are summed and then divided by the total number of characters scored in both individuals. ED measurement allows the use of both qualitative and quantitative data several workers identified genotype distances using ED. Van Bueningen and Busch (1997) in wheat, smith *et al*, 1987 in sorghum and Ajmone – Marsan (1998) in maize.

6. The use of molecular data in measuring genetic distances

The advent and explorations in molecular genetics led to a better definition of Euclidean distance by Beaumont *et al;* (1998) to mean a quantitative measure of genetic difference calculated between individuals, populations or species at DNA sequence level or allele frequency level.

Various genetic distance measurements are proposed for analyzing DNA-based data for the purpose of genetic diversity studies. Powel *et al.* (1996), identified different DNA-based marker techniques to include Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Restriction Fragment Length Polymorphic (RFLPs) and the most recent Simple Sequence Repeats (SSR) and Microsatellite (MT) of single nucleotide polymorphism (SNPs). The above nucleotide differences can be used effectively to run individual or combined data sets of morphological, biochemical or DNA based data. For DNA based data, where the amplification products are equated to alleles, the allele

frequencies can be calculated and the genetic distance between i and j individuals estimated as follows.

$$d(ij) = 1 < \left[\sum^{n} (X_{ai} - X_{aj}) \right]^{1/r}$$

Where X_{ai} is frequency of the allele a for individual I, and n is the number of alleles per loci; r is the constant based on the coefficient used. In its simple form, i.e $r = 1$, genetic distance can be calculated as:

$$d\int (ij) = 1/2 \left[\sum^{n} (X_{ai} - X_{aj}) \right]$$

Where r = 2, d(*i,j)* is referred to as Rogers (1972) measure of distance (RD), where

$$RD_{ij} = I/2[\ \Sigma(\ x_{ai}\text{-}x_{aj})\ ^2]^{1/2}$$

Where allele frequencies are to be calculated for some of the molecular markers, the data must first generate a binary matrix for statistical analysis. Binary data has been long and widely used before the advent of molecular marker data to measure genetic distance by Rogers (1972); Nei and Chesser (1983) coefficient and known as GD_{MR} and GD_{NL} respectively.

In the use of any given statistical formula to determine genetic diversity in molecular based data, one specific problem usually encountered is the failure of some genotypes to show amplification for some primer pairs. Robinson and Harris (1999) noted that lack of amplification may be due to "null alleles". Most often, it is difficult to ascribe lack of amplification to "null allele". It is therefore the reposed confidence of the researcher, that a "null allele" status of a genotype will not be considered as missing data during computation of genetic similarity- distance matrix so as to avoid gross error during result interpretation.

DNA based marker data have been successfully used to measure genetic distance in some crops (Pritchard *et al.* (2000) in pigeon pea; Beaumont *et al.* (1998) in wheat; Franco *et al.*, (2001) in maize; Dje *et al.* (2000) in Sorghum.

7. Grouping techniques in measuring genetic diversity

Genetic relationship among and with breeding materials can be identified and classified using multivariate grouping methods. The use of established multivariate statistical algorithms is important in classifying breeding materials from germplasm, accessions, lines, and other races into distinct and variable groups depending on genotype performance. Such groups can be resistant to diseases, earliness in maturity, reduced canopy drought resistant etc. The widely used techniques irrespective of the data source (morphological, biochemical and molecular marker data) are cluster analysis, Principal Component Analysis (PCA), Principal Coordinate Analysis (PCOA) Canonical Correlation and Multidimensional Scaling (MDS).

Cluster analysis presents patterns of relationships between genotypes and hierarchical mutually exclusive grouping such that similar descriptions are mathematically gathered

into same cluster (Hair *et al*. 1995); (Aremu 2005). Cluster analysis have five methods namely unweighted paired group method using centroids (UPGMA and UPGMC), Single Linkages (SLCA), Complete Linkage (CLCA) and Median Linkage (MLCA). UPGMA and UPAMC provide more accurate grouping information on breeding materials used in accordance with pedigrees and calculated results found most consistent with known heterotic groups than the other clusters (Aremu *et al.,* (2007a).

Principal components, canonical and multidimentional analyses are used to derive a 2-or 3-dimensonal scatter plot of individuals such that the geometrical distances among individual genotypes reflect the genetic distances among them. Wiley (1981), defined principal component as a reduced data form which clarify the relationship between breeding materials into interpretable fewer dimensions to form new variables. These new variables are visualized as different non correlating groups.

Principal components analysis first determines Eigen values which explain the amount of total variation displayed on the component axes. It is expected that the first 3 axes will explain a large sum of the variations captured by the genotypes. Cluster and principal component analysis can be jointly used to explain the variations in breeding materials in genetic diversity studies.

8. Conclusion

Genetic diversity studies is in no measure the first basic step in meaningful breeding programme and therefore require accurate and reliable means for estimation. Data sets sourced can morphological biochemical several workers successfully utilized various statistical tools in analysis diverse data sets and identified two major framework to really explain divergence in genotype performance. Genetic distance among and within individual data sets can be conveniently determined using specific tools while classificatory and cluster analysis require principal component and polymorphic sequence tools. Since each data set provide different molecular type of information, based marker data set is visualized to provide more reliable differentiate information on the genotypes. Analysis of data sets can be complex. Many software packages are available. There is still a need for a comprehensive and user-friendly software packages that would integrate different data set for analysis and generate reliable and useable information about genetic relationship. Equally important in genetic diversity studies is the need for a genetic resource centre. Studies should incorporate utilization of genetic diversity information in developing genetic resource centre accessible to breeders.

9. Acknowledgement

Many thanks to Ibirinde Olalekan for the secretariat assistance. I also appreciate Olayinka Olabode the Head of Department of Agronomy LAUTECH for the technical contributions given to this chapter.

10. References

Adewale, B.D.,Kehinde, O.B., Aremu, C. O. Popoola, and Dumet, J. 2010. Seed metrics for genetic and shape determinations in African yam bean [Fabaceae] (*Sphenostylis stenocarpa)* African Journal of Plant Science Vol. 4(4): 107-115

Ajmone-Marsan P., Carstiglioni, P., Fusari, F., Kuiper, M. and Motto, M. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor. Appl. Genet. 96:219-227.

Akbar, A.A. and Kamran, M. (2006). Relationship among yield components and selection criteria for yield improvement of Safflower – *Carthamustinctorious* L. J. Appl. Sci. 6: 2853-2855.

Akoroda, M.O. 1987. Principal component analysised metroglyph of variation among Nigerian yellow yam. Euphytica 32:565-573.

Aremu C.O. 2005. Diversity selection and genotypes Environment interaction in cowpea unpublished Ph.D Thesis. University of Agriculture, Abeokuta, Nigeria. P. 210.

Aremu, C.O., Adebayo M.A., and Adeniji O.T. 2008: Seasonal performance of cowpea in humid tropics using GGE biplot analysis world jour. of Biol.. Research. 1: (1) 8-13.

Aremu, C.O., Adebayo, M.A., Ariyo O.J., and Adewale B.D. 2007b. Classification of genetic diversity and choice of parents for hydridization in cowpea vigna unguiculata (L) walip for humid savanna ecology. African journal of biotechnology 6: (20) 2333-2339.

Aremu, C.O., Adebayo, M.A., Oyegunle, M. and Ariyo, J.O.2007a. TRhe relative discriminatory abilities measuring Genotype by environment interaction in soybean (Glycine max). Agricultural journ. 2 (2).: 210-215

Ariyo, O.J. and Odulaja, A. 1991. Numerical analysis of variation among accessions of Okra. Ann. Bot. 67:527-531.

Barrett, B.A. and Kidwell, K.K. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the pacific Northwest crop Sci. 38:1261-1271.

Baudoin, J. and Mergeai G. 2001. Yam bean sphenostylis stendcarpa. In: R.H. Raemaekers (ed) crop production in tropical Africa Directorate general for international (DGIC) Brussels, Belgium. Pp. 372 – 377.

Bearmont, M.A., Ibrahim, K.M., Boursot, P. and Bruqord, M.W. 1998. Measuring genetic distance. P. 315-325. In A. karp et al. (ed) molecular tools for screening biodiversity. Chapman and Hall, London.

Bohn, M.,H.F. Hutz and A.E.Melchinegr. 1999. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs ans SSRs and their used ion predicting progeny variance. Crop Sci. 39; 228-237.

Brown, A.H.D. 1989. The case for core collection. P. 135-156. In A.H.D England Brown et al. (ed). The use of plant genetic resources. Cambridge Univ. press Cambridge.

Brown-Guedira, G.L; Thompson J. Ajnelson, R.L. and Warburton M.L. 2000. Evaluation of genetic diversity of soybean introduction and North American ancestors using RAPD and SSR markers crop Sci. 40:815-823.

Bueninger, L.T. and Busch, Ritt. 1997. Genetic diversity among North American spring wheat cultivars. Analysis of the coefficient of parentage matrix. Crop Sc. 37:570-579.

Chippindale, P.T. and Weins J.I. 1994. Weighting portioning and combining characters in phylogenetic analyses. Syst. Biol. 43:273-287.

Christine, Joshua, H., William, A., Stacy, A. 2009. Genetic diversity of creeping bentgrass cultivars using SSR markers. Intern. Turfgrass Soc. Research Jour. 11.

Dje, Y, Hevretz, M., Letebure, C. and Vekemans, X. 2000. Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite marked theor. Appl. Gent. 100:918-925.

Eivazi, A.R, Naghavi, M.R, Hajheidari, M. Mohammadi, S.A, Majidi, S.A, Salakdeh, I. and Mardi, M. 2007. assessing wheat genetic diversity using quality traits, amplified fragment length polymorphism simple sequence repeats and proteome analysis Ann. Appl. Biol. 152:81-91.

Falconer, D.S. and Mackay T.F. 1996. Introduction to Quantitative Genetics,Longman, Harlow.

Franco, J., Crossa, J., Ribaot, M., Betran, J., Warburton, M. land Khairallah, M. 2001. A method for combinary molecular markers and phenotypic attributes for classifying plant genotype Theor-Appl. Gent. 103: 944-952.

Fu, Y. and Somers D. 2009. Genomo-wide reduction of genetic diversity in wheat breeding Crop Sci. 49:161-168.

Hair, J.R., Anderson, R.E., Tatham, R.L., and Black, W.C. 1995. Multivariatev data analysis with Readings. 4th edition, Prentice- Hall, Englewood Cliffs, NJ.

Hillis, D.M. 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Systems. 18:23-42.

Islam. M.R. 2004. Genetic diversity in irrigated rice pak. Jorn of Biol. Sci. 2:226-226.

Joshi, B.K, Mudwari, A., Bhatta, M.R, Ferrara, G.O. 2004. Genetic diversity in Nepalese wheat cultivars based on agro-morphological traits and coefficient of parentage. Nep Agric Res. J. 5:7-17.

Liu, S., Cantrell, R.G., Mccarty, J.C. and Stewart, M.D. 2000. simple sequence repeat based assessment of genetic diversity in cotton race accessions. Crop Sci. 40:1459-1469.

Martin, E., Cravero, V., Esposito A, Lozez F, Milanebi, L. and Cointry, E. 2008. Identification of markers linked to agronomic traits in globe artichoke. Aust. J. Crop Sci. 1:43-46.

Mohammadi A.A, Saeidi, G. and Arzuni, G. 2010. Genetic analysis of some agronomic traits in flax. Aust. J. Crop sci. 4:343-352.

Mostafa K., Mohammad, H. and Mohammad, M. 2011 genetic diversity of wheat genotype baspdon cluster and principal component analyses for breeding strategies Australian Jour. of Crop Sc. 5 (1): 17-24.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natil. Acad. Sci. (USA) 70. 3321-3323.

Nei, M. and Chesser, R.K. (1983). Estimation of fixation indices and gene diversities. Ann. Hum. Genet. 47: 253.-259

Pedersen, G. and Seberg, O. 1998. Moleculesvs Morphology. P. 359-365. In A. karp et al (ed) Molecular tools for screening Biodiversity.. Chapman and Hall, London.

Potter, D. and Doyle, J. J. 1992. Origin of African yam bean (*Sphenostylisstenocarpa*, Leguminosae): evidence from morphology, isozymes,chloroplast DNA and Linguistics. Eco. Bot. 46: 276-292.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J.Tinjey, S. and Rafalsky, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR markers for germplasm analysis. Mol.breed. 2: 225-238.

Rahim, M.A., Mia, A.A, Mahmud, F., Zeba, N. and Afrin, K. 2010. Genetic variability, character association and genetic divergence in murgbean platn Omic 3:1-6.

Rogers, J.S. 1972. Measures of genetic similarity and genetic distance studies in genetics. VII. Univ. Tex. Publ. 2713: 145-153.

Singh R.K. and Chaudhary B.D. 1985. Biometrical methods in quantitative genetic analysis. Kalyani publishers, New Delhi. India. P 38.

Smith, J.S., Paszkiewics, S. Smith,O.S.and Schaoffer, J. 1987. Electorphoretic, chromatographic and genetic techniques diversity among corn hybrids. P. 187-20. In Proc. Am.Seed Trade Assoc.Washington DC.

Smith, J.S., Smith, O.S., Boven, S.L., Tenburg, R.A.and Wall, S.J. 1991. The description and assessment of distances between inbred lines of maize III: A revised scheme for the testing of distinctiveness between inbred lines utilizing DNA RFLPs. Maydica, 36. 213-226.

Swofford D.L, Olsen, G.J. Wadell, P.J. and Hillis D.M. 1996. Phylogenetic inference P. 407-514. In D.M. Hillis et al (ed) Molecualr systematics. 2nd edition Sinaver Associates, Sunderland, M.A.

Taba, S., Diaz, J., Franco, J., Crossa, J.and Eberhart, S.A. 1999. A core subject of LAMP, from the Latin American maize project. CD-ROM, CIMMYT, Mexico, D.F., Mexico.

Thompson, J.A., Nelson, R.L. and Vodkin, L.O.. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Sci. 38: 1348-1355.

Van Bueningen, L.T. and Busch, R.H. 1997. Genetic diversity among North American spring wheat cultivars: I. Analysis of the coefficient of parentage matrix. Crop Sci. 37: 570-579.

Vavilov, N.I. 1950. The origin, variation, immunity and breeding of cultivated plants. Chronica botanica, B. chronica Botanica company, Waltham, Massachusetts. P 364.

Warburton, M. and Crossa, J. 2000. Data analysis in the CIMMYT. Applied Biotechnology Center for fingerprinting and Genetic Diversity Studies. CIMMYT, Mexico.

Weir, B.S. 1996. Intraspecific differentiation P. 385-403. in D.M. Hillis et al. (ed). Molecular systematics 2nd edition sunderlands M.A.

Wiley, E.O., 1981. Phylogenetics: The theory and practice of phylogenetics and systemic.John Wiley, New York.

Wrigley, C.W, Autran, J.C. and Bushuk, W. 1982. Identification of cereal varieties by gel electorphoresis of the grain proteins. Adv. Cereal. Sci. Techolog. 5:211-259.

18

Living on the Edge: Various Modes of Persistence at the Range Margins of Some Far Eastern Species

Elena Artyukova, Marina Kozyrenko, Olga Koren,
Alla Kholina, Olga Nakonechnaya and Yuri Zhuravlev
Institute of Biology and Soil Science, Far East Branch of Russian Academy of Sciences
Russia

1. Introduction

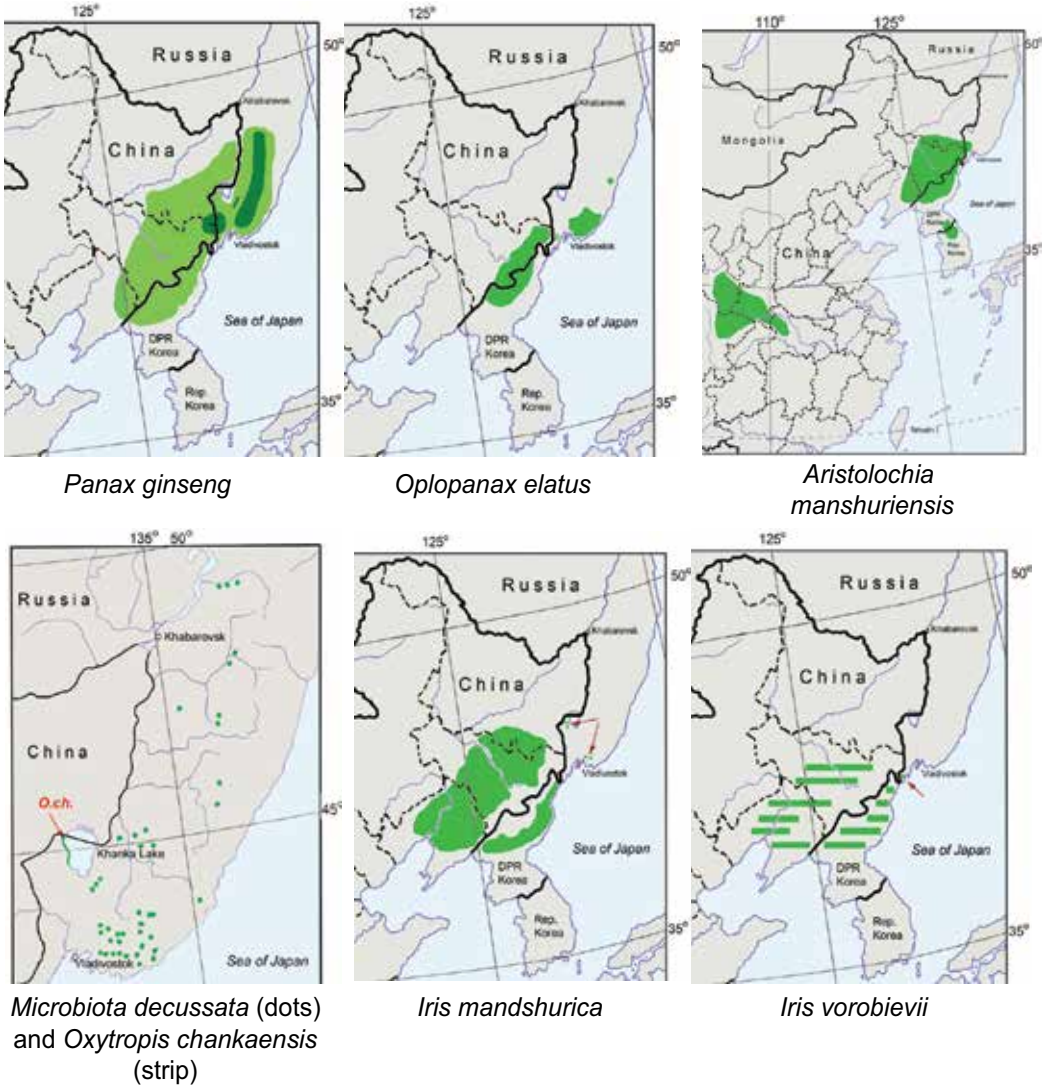
Present-day patterns of plant distribution have been formed under the influence of various biotic and abiotic factors. Plant distribution reflects the habitat preferences of species and the outcome of their competition as well as the complex evolutionary processes resulting in the specificity of mating systems, the genetic structure of different species and other aspects of species biology. Together, these factors determine the current ranges and distributions of plant species. At the edge of a species' range, the significance of particular interactions with the environment becomes more pronounced. However, our understanding of this class of interactions is limited. There is debate about whether these interactions represent a distinct and ordered set of related phenomena or whether they are unrelated and without order. Different approaches to this problem are needed in different situations. Understanding the processes of microevolution in species at the edges of their ranges is of great interest, particularly in view of the continuing decline in worldwide biodiversity and ongoing and future climate changes. When the area of a plant's habitat is sufficient, most populations exist in a relatively stable condition, and changes in their genetic structure follow slow processes, such as gene flow and genetic drift. However, in populations growing at the edge of their range, the rates of genetic processes can change dramatically. At the limits of the climatic and ecological tolerance of species, populations usually become smaller and more fragmented. These populations are generally less genetically diverse than those living at the center of the range because they exist in less favorable habitats and at lower densities, and consequently, they may be more prone to extirpation (Hampe & Petit, 2005; Vucetich & Waite, 2003). However, some species may have existed as groups of isolated populations for thousands of generations. The long-term survival and evolution of a species depends on the maintenance of sufficient genetic variability within and among populations. Patterns of population genetic diversity have been shown to be generally shaped by past climate-driven range dynamics, rather than solely by stochastic demographic and genetic processes (Hewitt, 2004). Given the enormous variety of plant life forms and their habitats as well as the complexity of their evolutionary histories, it is difficult to accept as a general rule that all marginal populations will exhibit lower genetic diversity than those from the center of a

species' range. Moreover, this rule would not hold true for all rare plants, which are associated with different causes underlying their rarity. A comparison of rare species belonging to different plant families and with different evolutionary histories may aid in inferring a number of scenarios under which a species may persist at the edge of its range. These scenarios may include common principles that do not correspond well to the "center-periphery" hypothesis.

2. The region and the species selected

The southern part of East Asia together with tropical Asia, is considered one of the centers of the origin and diversification of many plant taxa. The southern region of the Russian Far East (Primorsky Territory, Primorye) is located at the eastern edge of the Asian continent. The majority of this region is mountainous, with the Sikhote Alin Mountains extending throughout most of the area. The geographic location of the region accounts for the variety of its flora: it includes mountainous tundra areas, coniferous forests and coniferous-deciduous forests, and part of the lowlands surrounding Khanka Lake is occupied by forest-steppe. The flora of the region is unique and is characterized by a very complex mix of representatives of different kinds of vegetation. Unlike many other regions at the same latitude, most of this area was not glaciated during the Pleistocene; glaciers were limited to the highest peaks of Sikhote Alin, and the vegetation has undergone uninterrupted development since the Pliocene. The modern species complexes of Sikhote Alin have been formed during numerous migration processes under the influence of global climate change, the specific impact of the region's proximity to the ocean and marine transgressions. The region maintains a large number of rare and endangered species that originated in earlier epochs. Unfortunately, many of the previously abundant plants of the area have already become rare and are disappearing as a result of increasing anthropogenic pressure. The removal of rare and disappearing plants from their native habitats has led to the disruption of natural ecosystems and has significantly impoverished the biodiversity of the region.

Over 2,500 vascular plant species are represented in the flora of the southern part of the Russian Far East, and more than 340 species are listed in the Red Data Book of Primorsky Krai (2008) as endemic, rare or endangered. A large number of rare species are endemic, restricted to certain habitats, or their northern distribution limits are located in the region. Some of these species are relics of Tertiary flora with extremely limited ranges, whereas others exhibit range habitat preferences occupying tiny areas at specific locations. We have chosen a number of rare Far Eastern species (Figure) characterized by different life history traits for the present study. All of these species are listed in the Rare Plant Species of the Soviet Far East and Their Conservation (Kharkevich & Kachura, 1981), Red Data Book of Primorsky Krai (2008) and the Red Data Book of the Russian Federation (1988). The genetic diversity and population structure of each of these species have been studied using the following markers: allozymes, dominant DNA markers (random amplified polymorphic DNA, RAPD, and/or inter simple sequence repeat, ISSR) and sequences of noncoding regions of the chloroplast genome (cpDNA). Some life history traits and the main parameters related to genetic diversity and population structure in these species are presented in Table. Below, we provide a detailed description each of these species.



Panax ginseng | *Oplopanax elatus* | *Aristolochia manshuriensis*

Microbiota decussata (dots) and *Oxytropis chankaensis* (strip) | *Iris mandshurica* | *Iris vorobievii*

Fig. 1. Geographic ranges of rare Far Eastern species under study (according to Alexeeva, 2008; Huang et al., 2003; Kharkevich & Kachura, 1981; Shu, 2000; Xiang & Lowry, 2007; Zhu et al., 2010).

2.1 Ginseng (oriental ginseng), *Panax ginseng* C.A. Meyer

Ginseng, *Panax ginseng* C.A. Meyer, is a representative of the *Panax* L. genus related to an ancient family of angiosperms, the Araliaceae, members of which have been found in Cretaceous deposits. Most members of the Araliaceae are distributed in tropical and subtropical regions, with some species reaching the temperate zone. The *Panax* genus consists exclusively of herbaceous perennial plants, and all species in this genus are distinguished by the peculiar structure of an aboveground shoot that dies annually, whereas most members of the Araliaceae are trees or shrubs (Grushwitsky, 1961). The uniqueness of

Species*	Life-hystory traits					
	Life form	Life-span (years)	Clonality	Mating system	Pollination	Seed dispersal
Panax ginseng, E	PH	140	NC	S, A	I	G, A
Oplopanax elatus, V	S	300	C	S	I	G, A
Aristolochia manshuriensis, E	WL	50	NC	S	I	W, Wa
Microbiota decussata, V	S	250	C	S	W	G
Oxytropis chankaensis, V	PH	50	NC	S	I	W, Wa, G
Iris vorobievii, E	PH	≤7	PC	S	I	G, A
*Iris mandshurica,*V	PH	25	C	S	I	G, A

Species*	Genetic diversity parameters at population level								
	Allozymes			Dominant markers			Chloroplast DNA		
	P_{95}	*He*	F_{ST}	P_{95}	*He*	G_{ST}	$\pi \cdot 10^{-3}$	*h*	G_{ST}
Panax ginseng, E	7.6	0.022	0.204	4.0	0.013	0.249	–	–	–
Oplopanax elatus, V	25.0	0.113	–	23.5	0.088	0.293	–	–	–
Aristolochia manshuriensis, E	25.0	0.108	0.065	36.2	0.141	0.112	–	–	–
Microbiota decussata, V	–	–	–	45.1	0.249	0.352	0.603	0.954	0.090
Oxytropis chankaensis, V	37.1	0.294	0.025	66.9	0.290	0.135	0.480	0.703	0.146
Iris vorobievii, E	–	–	–	32.5	0.104	–	0.587	0.912	–
*Iris mandshurica,*V	–	–	–	31.3	0.108	–	0.285	0.733	–
Mean values for species with restricted ranges**	29.9	0.095	0.206	–	0.280	0.210	–	–	0.637a 0.165g

Table 1. Life-history traits and parameters of genetic variation in populations of rare Far Eastern species studied. *, Categories of rarity are given with the species name according to the Red Data Book of Primorsky Krai (2008): E, endangered; V, vulnerable. Life form: S, shrub; WL, woody liana; PH, perennial herb. Clonality: C, clonal; PC, poor clonal; NC, non-clonal. Mating system: S, sexual; A, apomixis. Pollination: W, wind; I, insect. Seed dispersal: W, wind; Wa, water; A, animal; G, gravity. Genetic diversity parameters at population level: P_{95}, percentage of polymorphic loci (95% criterion); *He*, expected genetic diversity; π, nucleotide diversity; *h,* haplotype diversity; F_{ST}, and G_{ST}, indices of genetic differentiation among populations; –, not determined. **, Mean values of genetic variation parameters in species with restricted ranges are cited from Gitzendanner & Soltis, 2000 (for allozyme data) and Nybom, 2004 (for dominant DNA marker data); G_{ST} based on chloroplast DNA data are cited from Petit et al. (2005): a, for angiosperms; g, for gymnosperms.

The genus *Panax* makes it difficult to ascertain its alliances with other genera of Araliaceae. The range of *Panax* is divided by the Pacific Ocean into two parts: an East Asian and a North American region (Grushwitsky, 1961). The intrageneric systematics of *Panax* has been revised as a result of the similarity of different *Panax* species habitus and the presence of intermediate forms that complicate species discrimination. Most *Panax* species are found in Eastern and Southeastern Asia, with the exception of *P. quinquefolium* L. and *P. trifolius* L.,

which inhabit North America. *P. ginseng* and *P. quinquefolium* have a chromosome number of 2n = 48, whereas most other *Panax* species exhibit 2n = 24.

In the past, *P. ginseng* was distributed throughout a wide territory in the forests of the Far East. It was estimated in the 1950s that in the first half of the 20th century, wild-growing ginseng plants could be found in large forestlands from 40° to 48° northern latitude and from 125° to 137° eastern longitude, covering approximately 500,000 square kilometers (Grushwitsky, 1961). At that time, the ginseng populations growing in Russia represented the northeastern boundary of the species' geographic range, whereas its main habitat area was located south and west of this territory, covering Heilongjiang, Jilin and Liaoning Provinces in China and the northern part of Korea. The natural range of ginseng is now drastically reduced (Figure). Wild-growing ginseng had disappeared completely from Korea and Liaoning Province in China by the 1930s and almost completely from Heilongjiang and Jilin Provinces in China by the 1990s (Zhuravlev & Kolyada, 1996). The distribution of ginseng has currently narrowed to a few patches in Russia and China (Zhuravlev & Kolyada, 1996; Zhuravlev et al., 2008). The largest patch of this species is located in the southern part of the Sikhote Alin mountain range; another population inhabits the southwest region of Primorye and Jilin and Heilongjiang Provinces (Changbai Mountains) in China; and a third population is located in the western part of Primorsky Territory (Figure, dark green). Reports of wild ginseng in China have become rare. In Russia, ginseng has been listed in the Red Book since 1975 as a federally threatened species (Red Data Book of the Russian Federation, 1988), and at present, Primorsky Territory in Russia is the only place in the world where natural ginseng populations exist, representing a remnant gene pool of wild-growing ginseng.

Ginseng is an herbaceous perennial species with an annually moribund shoot. It grows in certain special habitats and persists only in remote locations. It is very difficult to estimate the actual abundance of wild-growing ginseng in nature because of its ability to undergo lengthy dormancy, which may last from one to several dozen years; therefore, the recorded number of vegetative ginseng plants does not correspond to their true abundance. I.V. Grushwitsky investigated natural ginseng habitats in the 1950s and wrote that this plant did not exhibit a tendency toward extinction for the most part, despite its relic origin and low rate of regeneration (Grushwitsky, 1961). However, he also noted that human activity has a destructive influence on natural ginseng populations, reducing them to potentially threatened levels. In addition, logging, fires and the shallowing of rivers hamper the restoration of ginseng populations. Poaching and overexploitation of ginseng resources appear to be the most important and most obvious reasons for the reduction and exhaustion of ginseng populations today (Zhuravlev & Kolyada, 1996).

Ginseng reproduces exclusively through seeds and is nearly incapable of vegetatively reproducing (Grushwitsky, 1961; Zhuravlev & Kolyada, 1996). This species is characterized by a mixed mating system and the ability to produce seeds via autogamy, outcrossing or agamospermy without pollination (Koren et al., 1998). Presumably, self-pollination has prevailed in natural ginseng populations because of their low plant density. The occurrence of outcrossing, which is carried out by insect pollen transfer, cannot be excluded in natural environments, but this appears to be very rare and likely does not play a significant role in ginseng pollination. *P. ginseng* has also been shown to be a facultative apomict with a type of agamospermy resembling diplospory (Koren et al., 1998; Zhuravlev et al., 2008).

Studies on the genetic variability of *P. ginseng* using different molecular markers (allozymes, RAPD, ISSR) have detected very low levels of genetic polymorphism (Koren et al., 2003; Zhuravlev et al., 2008; Reunova et al., 2010a). Only 3 allozyme loci among 39 studied and only one of 74 RAPD loci were found to be polymorphic. These low levels (Table) do not differ significantly from estimates obtained with ISSR markers (P_{ISSR} = 9.3%, He_{ISSR} = 0.014, Reunova et al., 2010a). These data are also in agreement with the results of genetic studies on cultivated ginseng sampled from China and Korea (e.g., Kim & Choi, 2003). Thus, *P. ginseng* is characterized by a lower level of genetic variation than the average values found for rare endemic species (Table).

An analysis of population subdivisions based on allozymes and dominant DNA markers showed a low level of differentiation of natural *P. ginseng* populations and a decrease in the total genetic diversity in the Sikhote Alin population (Zhuravlev et al., 2008). The great majority (up to 95%) of the genetic variability in this species is concentrated within populations, whereas only 4.1% of the total variation was found to be distributed among 8 sub-populations (estimated by allozymes, Zhuravlev et al., 2008). Approximately 25% of the variation is distributed between populations (estimated with DNA markers, Reunova et al., 2010a). No correlation between geographic and genetic distances was found for the investigated populations using allozyme and DNA markers.

Extremely low levels of genetic variation are usually found in endemic or relic plant species with narrow ranges which can often be connected to a species' life history and/or evolutionary events such as selection or genetic drift. The low genetic diversity of *P. ginseng* populations indicates that this species has experienced a severe genetic bottleneck. In particular, the lack of variation in the Sikhote Alin population may be a result of a founder effect because of the lack of a refugium during the last Pleistocene-Holocene cooling. This hypothesis is supported by the results of an analysis of the genetic relationships of extant ginseng populations on the basis of allozymes and DNA markers (Zhuravlev et al., 2008; Zhuravlev et al, 2010; Reunova et al., 2010a). Without other available natural populations to analyze, we can only assume that the center of the genetic richness and, possibly, the center of the origin of ginseng was located southwest of its extant natural populations, potentially in a place where there is great industrial activity in modern China.

Another reason for the low genetic diversity found in *P. ginseng* may be adaptive selection that has occurred in response to climate cooling and during the expansion of the species northward from its southern refugia. Some peculiarities of ginseng biology (such as its underdeveloped embryos and aboveground germination) indicate that this species evolved during a warm climate as a representative of the ancient thermophilic flora. A number of *P. ginseng* traits, especially related to its mating system, indicate recent adaptations (Koren et al., 1998; Zhuravlev et al., 2008).

Past evolutionary events have resulted in extant ginseng populations being characterized by very low genetic variation. However, this low genetic diversity did not prevent the species from surviving across a wide territory until recently. Despite its weak competitiveness, *P. ginseng* possesses some degree of ecological flexibility, certain adaptations to unfavorable environments and the ability to ensure seed reproduction via different pathways. Among these properties, increased individual longevity may be the main mechanism underlying the long-term survival of the species under unstable conditions. Discoveries of 100-year-old ginseng plants are occasionally recorded, and cases of 300-year-old plants are well known

(Zhuravlev & Kolyada, 1996). In addition, the life spans of individual plants may lengthen in accordance with their ability to undergo long-term dormancy (in the form of so-called "dormant roots"). Because ginseng maintains a generative stage starting from an age of 3-5 years and lasting until death, annually producing up to 100 seeds per plant, a single individual can maintain a population's size for hundreds of years, even if only a small proportion of its offspring survives and reaches generative age. Moreover, the seeds of one generation can germinate over a period of several years because to mature, the underdeveloped embryo requires an alternation of warm and cold periods that can take from one to several years.

At the same time, limited opportunities for recombination (autogamy, agamospermy) can result in further reducing the genetic diversity of ginseng. Moreover, because all three extant ginseng populations occupy the northeastern margin of its former distribution area, their gene pools are not enriched by gene flow from extirpated central populations. With the continuing reduction of natural ginseng populations as a result of human activity, this species may become extinct in the wild very soon.

2.2 Japanese devil's-club, *Oplopanax elatus* (Nakai) Nakai

Oplopanax elatus is another member of the ancient family Araliaceae. This species is a deciduous shrub with a spiny stem approximately 1 m in height and large palmately compound leaves. The species' distribution area is quite limited. In Russia, *O. elatus* occurs in the southern Primorye, where its range is represented by several isolated populations associated with the main mountain peaks of the southern Sikhote Alin. Outside Russia, *O. elatus* grows on the northern Korean Peninsula (Kurentsova, 1968; Zhuravlev & Kolyada, 1996) and in Jilin Province in China (Xiang & Lowry, 2007) (Figure). *O. elatus* is a valuable medicinal plant; its effects are similar to those of ginseng, and it is authorized for medical use (Kurentsova, 1968; Zhuravlev & Kolyada, 1996; Schreter, 1975). This species is particularly vulnerable as a result of intensive harvesting and habitat disturbance due to fires and logging.

Oplopanax Miq. is a small genus that includes three species and demonstrates a classical Eastern Asian and North American disjunct distributional pattern. *O. horridus* occurs on the Pacific coast of the United States and Canada as well as around Lake Superior. *O. japonicus* is endemic to the Japanese Islands (Hokkaido, Honshu and Shikoku). This disjunct distribution pattern is observed for many plant genera and is explained by the existence of the land bridges between Eurasia and North America and between the Japanese Archipelago and the mainland during the Pliocene-Pleistocene. Based on the internal transcribed spacer sequences of nuclear ribosomal DNA (ITS rDNA), phylogenetic analysis confirmed the origin of the three *Oplopanax* species from a common East Asian ancestor. The closely related species *O. horridus* and *O. elatus* form a sister pair, and both are closely related to *O. japonicus* (Artyukova et al., 2005). An ancestral form that gave rise to *O. horridus* and *O. elatus* could have survived in the coastal zone during global cooling and the strengthening of climate continentality during the Early Pliocene and subsequently spread along the Pacific coast and across the Bering Bridge into North America.

In southern Primorye, *O. elatus* is mainly confined to the orotemperate belt in the altitudinal range of ca. 800 to 1,500 m above sea level and is a common species in the understory of fir-spruce forests (Kurentsova, 1968) occurring in moist, well-drained ecosystems. This species,

which is clearly representative of the thermophilic and hydrophilous Turgai flora, has found specific refugia in the understory of fir-spruce forests, which favors its growth. Only under these conditions does the species occur at lower altitudes in coniferous–deciduous forests. However, its growth in mixed forests is hindered not only by low humidity, but also by competition from shrubs and grasses, which are abundant in the understory of mixed forests in Primorye. In spruce forests, the shrub and herb layers are poorly developed, and *O. elatus* has an opportunity to achieve a wider distribution (Kurentsova, 1968). In these habitats, the species is usually very abundant and sustainable. However, it is unable to withstand competition from heliophilous species, which rapidly fill gaps after disturbances to primary stands, such as natural or artificial fires or logging.

O. elatus mostly reproduces vegetatively through lateral branches that form adventitious roots when in contact with the soil. This species grows clonally, and individual clones can include 20 or more shoots, which can remain connected by decumbent stems to the parental plant for extended periods of time (Kurentsova, 1968). The life span of a single shoot is up to 40 years, and the overall life span of an individual from germination to the death of all parts of the clone can reach 300 years (Zhuravlev & Kolyada, 1996). Propagation of *O. elatus* through seeds is impeded; the seed set on individual plants is high, but most seeds exhibit underdeveloped embryos. Delayed embryo development is an ancient feature of many Araliaceae. A small number of *Oplopanax* seeds germinate in the second year, and most of the seedlings die.

An analysis of genetic diversity using dominant molecular DNA markers (RAPD and ISSR) allowed the levels of intra- and interpopulation variability to be evaluated in three geographically isolated *O. elatus* populations (Reunova et al., 2010b). The level of genetic diversity in *O. elatus* is much less (Table) than that in *Kalopanax septemlobus* (*P* = 59.2%, *H*e = 0.119; Huh et al., 2005) or *Dendropanax arboreus* (*P* = 70.2%, *H*e = 0.253; Figueroa-Esquivel et al., 2010), which are Araliaceae species with more continuous ranges that have been studied using dominant markers.

The *O. elatus* sample from Mt. Litovka (Livadiiskii Range, southern Sikhote Alin) with the highest level of RAPD variability (*P* = 29.41%, *H*e = 0.110; Reunova et al., 2010b) is characterized by a level of genetic diversity similar to that detected based on allozyme data (*P* = 25.0%, *H*o = 0.131, *H*e = 0.113; Kholina et al., 2010). Allozyme analysis has shown a slight excess of heterozygotes in this population. This may be attributable in part to vegetative reproduction of pre-existing heterozygous genotypes and/or to selection favoring heterozygotes that are better able to adapt to new environments. The low polymorphism and allelic diversity in *O. elatus,* along with a rather high level of observed heterozygosity (*H*o), may be indirect evidence of the effect of genetic drift, which reduces allelic diversity and, consequently, the proportion of polymorphic loci. It seems likely that the source from which the studied *O. elatus* populations were established was heterozygous plants, especially if they were characterized by increased viability. In addition, an excess of heterozygotes and high values of *H*o may be explained by mutations that could have arisen in long-lived clones and been maintained by vegetative reproduction (similar to the serpentine endemic *Calystegia collina*; Wolf et al., 2000). Given the mainly vegetative mode of reproduction of *O. elatus* and the total life span of an individual accession lasting up to 300 years, accumulation of mutations maintained by vegetative reproduction cannot be ruled out.

Thus, based on different nuclear DNA marker data, the levels of genetic variation in the populations located near the northern edge of the *O. elatus* range appear to be comparable with the mean values reported for rare plant species (Table). In addition to its small distribution range and ecological specificity, the low level of genetic variation detected in *O. elatus* may be determined by the species' history. Reductions in polymorphism levels are caused by population bottlenecks resulting from dramatic decreases in population size. Bottleneck events have been proven to be the most probable reason for low polymorphism levels in relic Araliaceae species such as *P. ginseng* (Koren et al., 2003) and *Dendropanax morbifera* (Kim et al., 2006). As mentioned above, *O. elatus* inhabited deciduous Turgai forests. During periods of glaciation, broadleaved forests experienced dramatic range contractions and were forced to retreat southward, and the zone of high mountain vegetation was shifted to lower elevations. It has been suggested that coastal areas and the temporarily emerged continental shelves of the Sea of Japan connecting the Japanese Archipelago to the continent may have served as refugia for coniferous and broadleaved mixed forests (Sakaguchi et al., 2010). During its isolation in refugia and its subsequent expansion, *O. elatus* might have adapted to life in the understory of the fir-spruce forests. The possibility cannot be excluded that contractions in the species' population size following re-establishment from just a few founders could have happened more than once and may be occurring today, when these processes are associated with intense human activity (logging, fires and plant harvesting).

An analysis of the clonal structure of the population from Mt. Litovka indicated that in a sample of 29 accessions, there were a total of 22 multilocus genotypes. Only four of these genotypes were found more than once (from two to four times). The genotypic diversity (G/N) was determined to be 0.76, indicating that 76% of the individuals exhibited unique genotypes. Simpson's diversity index, D, is equal to 0.97, whereas the mean D value for 21 species of clonal plants has been 0.62 (Ellstrand & Roose, 1987). The existence of a great number of different genotypes along with low genetic variation has been reported for a number of species (Watkinson & Powell, 1993; Xie et al., 2005 and references therein). This phenomenon may be explained by the sporadic seed propagation in these species. In general, even a small number of individuals resulting from sexual reproduction is sufficient to make a population genotypically variable. Our findings imply that in the population of *O. elatus*, sexual reproduction succeeds periodically and results in the maintenance of a certain level of genotypic diversity. In addition, it has been suggested that events such as a presumptive origin from heterozygous founders, the accumulation of mutations and the retention of the changes via vegetative reproduction could promote the high level of genotypic variation observed.

The rarity of *O. elatus* is largely a result of this species' ecological specificity, dependence on humidity and low competitive ability. Anthropogenic influences (plant harvesting and habitat destruction) threaten the existence of this species. The flexibility of its reproductive system, combining different modes of reproduction, allows this species to renew heterozygous genotypes by clonal growth and to contribute additional variability resources through sporadic seed reproduction. At the same time, some features of the species' biology (the long life span of a single clone, overlapping generations and the ability to cross-pollinate) also help to maintain a certain level of polymorphism. The existing level of genetic diversity can be crucial to preventing the negative consequences associated with a small number of isolated populations and genetic drift.

2.3 Manchurian birthwort, *Aristolochia manshuriensis* Komarov

Manchurian birthwort, *Aristolochia manshuriensis* Kom., is a relic woody liana that belongs to the ancient angiosperm family Aristolochiaceae. This species is endemic to the Manchurian floristic region (Kitagawa, 1979) and occurs in the montane mixed forests of China and Korea (Figure). In the southern part of the Russian Far East, the species reaches the northern boundary of its range. *A. manshuriensis* prefers specific habitats in river floodplains at a certain altitude above sea level, along chutes and in steep slope foots, especially those with northern slope aspects, avoiding sunlit habitats associated with soil overheating. All Russian *A. manshuriensis* populations are located in the valleys of just three rivers and their tributaries. These populations are fragmented and separated by ridges. Within populations, plants grow at an uneven density, forming patches separated by 0.5–4 km from each other. The growth and expansion of natural *A. manshuriensis* populations are suppressed, and their natural regeneration is very poor (Kurentsova, 1968). At least three of four extant *A. manshuriensis* populations are now located in disturbed habitats and are experiencing strong anthropogenic pressure.

A. manshuriensis exhibits no means of vegetative reproduction, possesses a poor rooting ability and requires a long period of root formation (Shulgina, 1955). Seed reproduction in this species is usually successful, but spontaneous fruit set of only 2% has been documented. The seed set on a fruit is rather high because all mature fruits contain approximately 100 fully viable seeds. A histological analysis showed normal development of flower structures, a high level of pollen-grain fertility (97%) and a large number of pollen grains in the anthers of this species (Nakonechnaya et al., 2005).

Its seeds are adapted to spread by water and by wind (Nechaev & Nakonechnaya, 2009). These seeds are characterized by underdeveloped embryos, but seed germination is usually successful after winter dormancy or brief artificial stratification (Adams et al., 2005). Under favorable conditions, seedlings develop to the generative stage over 10-12 years, reaching a height of 15 m or more (with support) by this age. A long life span for a single *A. manshuriensis* plant is estimated to be more than 40 years. Under unfavorable conditions, plants can be kept in a virgin state for many years without flowering.

Similar to most Aristolochiaceae species, *A. manshuriensis* is an evolutionarily outcrossing plant exhibiting special adaptations of its flower structures (gynostemium) to ensure cross-pollination by insects. Autogamy and geitonogamy are possible, but pollinators are required to perform self-pollination as well as cross-pollination (Nakonechnaya et al, 2008). This species is characterized by long-term and abundant blossoming as well as a long life span of individual flowers, which increases the possibility of cross- and self-pollination. However, the presence of suitable insect pollinators during *A. manshuriensis* blooming is a precondition for successful seed reproduction in this plant. The flower structure of *A. manshuriensis* is matched to insect pollinators with a certain body size and thorax structure to allow the transfer of a sufficient number of pollen grains to the stigma. Among the visitors to *A. manshuriensis* flowers, flies of the genus *Pegoplata* (Anthomyiidae) are the only possible pollinators (Nakonechnaya et al., 2008). These flies have many other substrates available for breeding and larval feeding, which may be a cause of their rare visitations to *A. manshuriensis* flowers and, as a result, the rare fertilization and low fruit set of this species. Given that many Aristolochiaceae species are associated with specific insect pollinators (Nakonechnaya et al., 2008 and references therein), it may be assumed specific pollinators of

marginal populations of *A. manshuriensis* have been lost as a result of shifts in the areas occupied by flora and fauna during the Pleistocene cooling. However, the high seed productivity of each mature fruit indicates the high potential fecundity of the species.

The mean level of genetic variability in *A. manshuriensis* is rather moderate based on allozymes (Nakonechnaya et al., 2007) and RAPDs (Table). The relatively high level of heterozygosity in *A. manshuriensis* populations compared with the average values for rare endemic species (Table) may be explained by outcrossing. Indeed, an excess of heterozygotes has been revealed within *A. manshuriensis* populations, despite their small population sizes, and no inbreeding has been found at the species level (Koren et al., 2009).

The relatively low level of population differentiation identified with molecular markers may be indicative of gene flow among populations in the present or recent past. Another possible reason for the low differentiation observed may be the greater degree of habitat integrity and the closer connections between populations that existed in the past (Koren et al., 2009). This explanation is supported by the low mean genetic distances between populations (D_N = 0.0096) and the value of gene flow between populations (Nm = 3.96) calculated based on G_{ST}, which shows that the isolation of the populations studied is incomplete.

Within two of the most disturbed populations, a statistically significant subdivision is shown with both allozyme and RAPD markers. Moreover, an effect of a genetic bottleneck is revealed in three populations that have undergone strong anthropogenic pressure (Koren et al., 2009). These findings indicate a reduction in gene flow between subpopulations in recent years as result of the intensive disturbance of their habitats.

Thus, the relic populations of *A. manshuriensis* exhibit moderate levels of genetic diversity. The low level of differentiation between the populations is probably connected with the evolutionary history of the species and the interaction of various factors, such as migration and selection. Despite the fact that all of the extant *A. manshuriensis* populations in the Russian part of the range are isolated from one another by ridges, they all grow in the valleys of rivers with their headwaters in the Borisovskoe Plateau and may have a common origin. It is likely that fragmentation of the *A. manshuriensis* populations occurred recently during the increase of human activities and the forest destruction that took place in the 20th century (Koren et al., 2009). The ongoing degradation of natural populations seems not to be associated with depletion of the gene pool and inbreeding depression. Because of the small size and fragmentation of these populations and their isolation from each other and from the species distribution center, genetic drift makes a significant contribution to the decrease in this species' genetic variability. In addition, anthropogenic influences, such as fires and uncontrolled harvesting of plants for their medicinal value (Bulgakov & Zhuravlev, 1989), also play a role in the contraction of the *A. manshuriensis* population size. Seed reproduction through outcrossing seems to be the only means for Manchurian birthwort to maintain a sufficient level of genetic variation. The preference of this species for specific ecotopes, its poor potential to undergo vegetative reproduction and its rare fertilization as well as its weak competitive ability in the virgin stage do not allow it to expand beyond the boundaries of its existing stands.

2.4 Russian arborvitae (Siberian Cypress, Russian Cypress), *Microbiota decussata* Komarov

The perennial evergreen decumbent coniferous shrub *Microbiota decussata* Kom. (Cupressaceae) is known from the southern part of the Russian Far East, where it occurs on

some mountain peaks in the Sikhote Alin Mountains. The species is strongly restricted in its distribution, with its natural range stretching from 43˚ 00′ N to 48˚ 50′ N, from the subalpine zone of southeastern Primorsky Territory to the mountains in the Anyuy River basin in Khabarovsk Territory (Figure).

M. decussata is the sole species in the genus *Microbiota*, which is the only Cupressaceae genus endemic to the Sikhote Alin Mountains, and it is considered to represent one of the plant species that existed in this mountain system before the mountain-valley glaciers developed (Shlotgauer, 2011). The Sikhote Alin palynofloras contain pollen of Cupressaceae (Pavlyutkin et al., 2005 and references therein), but the cupressaceous pollen grains could not be distinguished beyond the family level. The fossil species *Cupressinoxylon microbiotoides* Blokhina from the Eocene/Oligocene deposits of Yuri Island of the Kurils is considered to have represented a putative species of the genus *Microbiota* (Blokhina, 1988). Of the extant Cupressaceae species, *M. decussata* is most closely related to *Platycladus orientalis* based on morphological data. The only fossil record of *M. decussata* was found in the Pavlovka lignite field (the southern Primorye) dated to the Pliocene or the late Eocene through late Oligocene (Pavlyutkin et al., 2005 and reference therein). In the Pavlovka deposits, wood of *M. decussata* was found together with fossilized wood of *P. orientalis* (Bondarenko, 2006). The presence of both species in the Pavlovka lignite field indicates their co-occurrence in the Oligocene/Pliocene plant associations of Sikhote Alin. Currently, *Platycladus orientalis* is a common tree in China and is widely cultivated elsewhere in Asia, eastward to Korea and Japan, southward to northern India and westward to northern Iran, whereas *M. decussata* is not found outside Russia. Its range is restricted to the Sikhote Alin, where the species exhibits a strongly disjointed distribution (Figure). Phylogenetic studies based on ITS rDNA and chloroplast markers (Gadek et al., 2000; Little et al., 2004) have confirmed the genetic affinity of the two taxa, which are placed in a sister clade to *Calocedrus*.

The heliophilous Russian arborvitae shrubs inhabit steep stony slopes and scree fields at altitudes from ca. 300 to 1,700 m above sea level (mainly at or above the timberline) under climatic conditions that impede the growth of forest vegetation. The xerophytic species *M. decussata* is a poor soil-tolerant plant that is resistant to a wide range of temperatures. As it grows under very harsh climatic and edaphic conditions, *M. decussata* is a pioneer in the colonization of cold stone deserts and participates in soil formation processes (Urusov, 1979). This low shrub with creeper and ascending branches often forms dense (crown density of 1.0) monodominant stands (Kurentzova, 1968; Krestov & Verkholat, 2003). In the southern part of its range, populations of *M. decussata* are more sustainable than those of another subalpine shrub, *Pinus pumila*. At the northern boundaries of its range, in the Anyuy and Chor River basins, the viability of *M. decussata* populations appears to be lower as a result of competitive inhibition by boreal species. Under these circumstances, fruiting of Russian arborvitae is rarely observed (Melnikova & Machinov, 2004), and the species is replaced by *P. pumila* (Kurentzova, 1968; Urusov, 1988).

M. decussata is an anemophilous monoecious plant that propagates sexually and by layering. The plant begins bearing fruits at 14–17 years, and the maximum life span of individuals has been determined to be approximately 250–300 years. The female cones of this species contain only a single 2 mm long naked seed. The seeds are able to disperse over a very short distance, dropping near the parent plant, though sporadic dispersal by animals cannot be ruled out. The seeds retain germination capacity for a long period but germinate almost

exclusively after fires (Urusov, 1979). Outside the area occupied by their parent population, seedlings and juvenile plants are extremely rare (Kurentzova, 1968). *M. decussata* grows very slowly (3–7 cm a year) and can extend into new free habitats through the slow creeping of stems (up to 3–5 m in length) that produce roots at their nodes (Urusov, 1988).

M. decussata exhibits a high level of nuclear genome variation (Table), despite the restricted and fragmented range and geographical isolation of its populations (Artyukova & Kozyrenko, 2009). The gene diversity within populations from the southern and middle parts of the range of this species is slightly higher than that in the northern population from the Chor River basin, which is in close proximity to its northern range limits. The lower level of gene diversity in the northern population might be caused by severe temperature conditions interfering with sexual reproduction; thus, species dispersal occurs mainly by layering. The highest level of genetic diversity is found in the population from the central Sikhote Alin. The lack of population-specific RAPD markers and the similar high levels of diversity retained in the populations could be caused by a common gene pool and ancient polymorphism. The genetic differences among all pairs of populations (separated by 10–400 km) fit an isolation-by-distance model. Overall, based on RAPD markers, the level of nuclear genome variation in endemic *M. decussata* (Table) is comparable with that found in some other Cupressaceae species with fragmented or restricted ranges (e.g., Hwang et al., 2001; Allnutt et al., 2003; Hao et al., 2006).

Based on sequence data from noncoding cpDNA regions, a considerable level of haplotype diversity and a low level of nucleotide diversity have been revealed in *M. decussata,* which is similar to what has been found in some other woody species, including two species of Cupressaceae, *Cunninghamia konishii* and *C. lanceolata* (Hwang et al., 2003). A common haplotype for all populations has not been found, though three haplotypes of the central Sikhote Alin population are also present in southern and/or northern populations. The large number of unique, closely related haplotypes within each population may suggest that the distribution area of *M. decussata* was fragmented a long time ago by the extirpation of populations in the adjacent territory.

In contrast to nuclear DNA, there is no significant isolation-by-distance effect observed in the plastid genome in this species, but cpDNA data related to differentiation show a nonrandom geographical distribution of haplotypes. Differentiation in *M. decussata* appears to be associated with historical events and the complex mountain topography of its range. The results of nested clade analysis and coalescent simulation data provide evidence of species expansion (Artyukova et al., 2009). The presence of the same substitutions and shared haplotypes in populations from opposite ends of the range indicate that ancestral populations of this species might have formerly exhibited a contiguous distribution range. The highest gene diversity in both genomes and the presence of shared haplotypes in the population from the central Sikhote Alin, which coincides in position with the assumed site of this species' origin at the watershed of the ancient Ussuri and Partizanskaya River basins (Urusov, 1979), may indicate its ancient origin. The fossil data (see above) confirm the occurrence of *M. decussata* in the Oligocene/Pliocene plant associations of Sikhote Alin near the southern limits of the current species range.

Unique species life-history traits have ensured the survival and range expansion of *M. decussata*, while most Arcto-Tertiary species, including species of Cupressaceae, either

shifted their ranges south (e.g., *Platycladus orientalis*) or vanished (e.g., putative common ancestor of *M. decussata* and *P. orientalis*). The long-term persistence of *M. decussata* in the territory that has been ice-free during glaciations as well as traits such as a long life span and pre-reproductive phase, the long-term survival of its seeds in soil seed banks and spreading by layering, seem to enable the retention of historically established levels of gene diversity in fragments of the ancestral populations of this species. Climatic and landscape changes at the Pleistocene–Holocene boundary caused the timberline to rise, and *M. decussata*, like most contemporary montane plants, retreated toward higher elevations. The enrichment of plant communities of the Sikhote Alin Mountains with cool temperate and boreal species also resulted in the shifting of *M. decussata* stands to ecotopes with severe climatic and soil conditions (e.g., steep, stony slopes and scree fields) and in the splitting of the large ancestral population, forming its disjunct present-day distribution.

The modern populations of *M. decussata* have the ability to survive the ongoing climate changes and global warming because of the physiological and ecological range of tolerance and life history traits of this species. Upward shifts of the timberline forcing *M. decussata* to migrate up to mountain summits with taluses (Urusov, 1988; Krestov & Verkholat, 2003) may lead to further contraction of its remnant populations on mountains at middle elevations. However, because this species can survive under harsh environmental conditions (e.g., poor soil, high solar radiation, extreme winds), which are unsuitable for other species, it could survive on the highest mountain peaks.

2.5 *Oxytropis chankaensis* Jurtzev

Oxytropis chankaensis Jurtz. (synonym *O. hailarensis* subsp. *chankaensis* (Jurtzev) Kitag.) is a perennial herb with a narrow habitat range that is restricted to the west shore of Khanka Lake (Kharkevich & Kachura, 1981; Pavlova, 1989; Yurtsev, 1964), which is the largest lake in Northeast Asia (Figure). *O. chankaensis* plants occur only in sandy habitats on a narrow strip along the Khanka Lake shoreline, forming separate populations numbering approximately 80 to 500 individuals (Kholina & Kholin, 2006). The genus *Oxytropis* DC is comprised of approximately 450 species occurring predominantly in the mountains of Asia. The high ecotopic diversity and the mosaic of conditions in mountain ecosystems lead to the occurrence of neighbouring populations of *Oxytropis* species with different ecological requirements and contribute to the enhanced speciation and interspecific hybridization in the genus. These processes explain the current controversy within the taxonomy of this genus (Malyshev, 2007, 2008).

A legume species originally found on the shoreline of Khanka Lake was first described as a distinct endemic species, *Oxytropis chankaensis*, based on the observation of definite morphological differences compared with its congeners (Yurtsev, 1964). In the International Legume Database (ILDIS), *O. chankaensis* is considered to be a subspecies of *O. hailarensis* Kitag., the species that occurs in China and Mongolia (Bisby et al., 2009). In the Flora of China (Zhu et al., 2010), both species are regarded as synonyms of *O. oxyphylla* (Pall) DC, a type species of section *Baicalia,* subgenus *Oxytropis* (Malyshev, 2007). However, *O. chankaensis* and *O. oxyphylla* are clearly distinguished by morphological features and by their ploidy levels: *O. oxyphylla* is diploid (2n = 16; Zhu et al., 2010), whereas *O. chankaensis* is tetraploid (2n = 32; Probatova et al., 2008a). Phylogenetic analyses of the ITS rDNA and

three noncoding regions of chloroplast DNA strongly confirm that *O. chankaensis* and *O. oxyphylla* are distinct species (Artyukova et al., 2011a).

O. chankaensis is the only representative of section *Baicalia* in Primorye (Pavlova, 1989). It is an outcrossing species that is pollinated by bumblebees, like most *Oxytropis* species (Yurtsev & Zhukova, 1968), with pollen potentially being dispersed over long distances. *O. chankaensis* is characterized by a long flowering period (from the third week of May until mid-August) and a high number of flowers per plant (up to 400, an average of 40 inflorescences with 4–14 flowers per plant) as well as high pollen fertility (95.7 ± 1.4%), which contribute to successful pollination and fertilization. The plants of this species exhibit high fecundity, with fruits containing up to 20 seeds, and an individual plant produces approximately 4,000 seeds (Kholina et al., 2003). The first fruits mature by the end of May, and fruiting lasts until September. Mature spherical pods can be dispersed by wind and water over long distances beyond the limits of local populations, while some seeds from dehiscent pods are gravity-dispersed over only a short distance from the maternal plant to form the soil seed bank. During the vegetative season, seedlings emerge from the seeds of the first fruits, and some of the seeds begin to germinate in the following year, after a winter dormancy period.

Ontogenic features of *O. chankaensis* (Kholina & Kholin, 2006), such as its long life span, overlapping generations, the multiplicity of its development, its early transition to the generative state and the long period of the generative state, are the most important characteristics of this species for maintaining its population numbers and preserving genetic heterogeneity. The juvenile and immature stages are the most vulnerable stages of the species' life cycle.

The characteristics of the reproductive biology of *O. chankaensis*, such as the normal structure and function of the reproductive organs, the high fertility of pollen, the considerable duration of the flowering period and the long life of the flower, result in reliable pollination and high seed production. This species is characterized by a combination of different modes of dissemination, hardseededness and long-term maintenance of germination (over 10 years). Seed dispersal over short and long distances by wind and water promotes intraspecific genetic structure and the homogenization of populations through gene exchange. The longevity of the seeds leads to their accumulation in the seed bank in the soil, and these seeds replenish the gene pool of a population when they germinate. The high fecundity of *O. chankaensis,* together with mechanisms that support recombination (the predominance of cross-pollination) and the exchange of genes (via pollen and seeds), provides reliable renewal of this species *in vivo*.

An allozyme analysis showed that *O. chankaensis* is an autotetraploid that arose through the fusion of nonreduced gametes in the course of multiple crosses between genetically different plants (Kholina et al., 2004). The recurrent polyploidy events in the evolutionary history of this species are confirmed by the presence of several chloroplast DNA haplotypes in each population (Artyukova et al., 2011b). In addition, the levels of genetic diversity in *O. chankaensis* populations revealed using allozymes (Kholina et al., 2009) and RAPDs are high compared with the average values for rare endemic species (Table).

The most striking feature of the *O chankaensis* plastid genome, which is maternally inherited in Fabaceae, is the unexpectedly high cpDNA haplotype diversity (Table) for a species with

an extremely narrow geographic range. At the same time, its nucleotide diversity is low. Most (≥90%) of the genetic variability of nuclear markers and of the chloroplast markers is distributed within populations. Unlike most angiosperms (Petit et al., 2005), the level of cpDNA subdivision in this species does not differ from the levels of differentiation of the nuclear genome based on RAPD markers (Table). In addition, correlation between geographic and genetic distances is absent for the plastid and nuclear genomes. The low population partitioning (cohesive genetic system) and lack of phylogeographic patterning may be attributed to both recent fragmentation of a once continuous population and extensive (past/modern) gene flow via pollen and seeds, which prevents the accumulation of genetic differences.

Despite the slight divergence observed, each population possesses a unique part of the species gene pool. The biological traits and reproductive features of this species as well as the existence of tetrasomic inheritance and recurrent tetraploidy events in its evolutionary history contribute significantly to the maintenance of genetic diversity.

Thus, the narrowly endemic species *O. chankaensis* exhibits adaptive mechanisms that enable it not only to successfully renew its populations in the coastal zone, which are exposed to frequent flooding and other adverse factors, but also to maintain the high level of recombination responsible for the survival of the species in a changing environment. Apparently, the rarity of this species is a result of its high habitat specificity; it lives only on the sandy shores of a large lake where there is intense insolation and high air humidity. Fluctuations in its population size as a result of lake-level oscillations can result in reduced numbers at some localities, but in these cases, the species' high productivity and reserves of genetic variability, which enable adaptive responses of species, help to restore the populations. The threat of total destruction of the species is raised by human-induced habitat destruction.

2.6 *Iris mandshurica* Maximowicz and *I. vorobievii* N.S. Pavlova

Iris L. is a Northern Hemisphere genus of flowering plants composed of approximately 280 valid species. As they are mostly open land plants, *Iris* species are adapted for living in a wide range of habitats from cold and montane regions to grassy slopes, steppe meadowlands, arid and marsh areas and riverbanks, though there are no truly sylvestral plants among this group. Irises are outcrossing species exhibiting flowers with specialized structures to insure cross-pollination; their fruits (dry capsule) contain numerous seeds that can disperse through a variety of mechanisms, such as barochory, autochory, anemochory, hydrochory, myrmecochory and zoochory. In addition, these perennial plants reproduce asexually through bulbs or rhizomes that form dense or loose colonies (tufts) and grow in size over 20 years. Many *Iris* species inhabiting Russia are at the border of their geographical ranges and occur in small, isolated populations. In the southern part of the Russian Far East, there are 11 *Iris* species, mainly belonging to the subgenus *Limniris,* section *Limniris* (Pavlova, 1987). Only two species, *Iris mandshurica* Maxim. and *I. vorobievii* N.S. Pavlova, are representatives of a small section of dwarf irises, *Psammiris,* of the subgenus *Iris.* Psammirises usually have yellow flowers with a yellow beard in the center of the outer three perianth segments ("falls") and seeds with a white appendage (loosely called aril) indicative of possible dispersal by ants.

Psammirises are mainly Asian species, and only *Iris arenaria*, which is considered a synonym of *I. humilis*, is widely distributed in southern Europe (Alexeeva, 2008). *I. mandshurica* and *I. vorobievii*, which are found in the south of Primorye, are very similar to *I. humilis*, but they differ from each other and from *I. humilis* in some morphological features (Pavlova, 1987; 2006; Alexeeva & Mironova, 2007; Bezdeleva et al., 2010) and in their chromosome numbers (Probatova et al., 2008b; Shu, 2000). The distinctiveness of these species has also been confirmed by the use of molecular DNA markers (Kozyrenko et al., 2009).

In Russia, small and disjunct populations of *I. mandshurica* occur rarely (Figure, arrows) on dry grassy slopes, stone hills and in the steppe meadow. *I. vorobievii* is found in the only locality in the extreme south of Primorye (Figure, arrow), where it grows on open grassy slopes of hills and in the meadows of fluvial terraces as low-density, isolated patches (Pavlova, 2006; Alekseeva, 2008). Outside Russia, *I. mandshurica* mainly occurs at altitudes of ca. 400–800 m in northeast China (Heilongjiang, Jilin and Liaoning Provinces) and in northern Korea (Figure; Shu, 2000), whereas there are no available data on the range of *I. vorobievii*, though the occurrence of the species in adjacent regions in China and Korea (Figure, strips) cannot be ruled out.

Like all *Iris* species, *I. mandshurica* and *I. vorobievii* reproduce sexually and asexually through rhizomes. The seeds of both species are gravity-dispersed to only a short distance from a maternal plant, though secondary dispersal by ants cannot be excluded entirely. *I. mandshurica* exhibits a thick, shortened rhizome growing almost horizontally and forming loose turf. The vertical, stout rhizome of *I. vorobievii* is very short (1 cm in length) and presents only a few lateral buds; the rhizome grows at one end while the old part of it dies, and the rhizome does not reach great lengths (Bezdeleva et al., 2010). These features indicate that the species is a short-lived perennial (Alexeeva, 2008), living no more than 5–7 years. In contrast to most *Iris* species, *I. vorobievii* is difficult to cultivate.

Both psammirises studied exhibit similar levels of nuclear and plastid DNA diversity (Table). Based on RAPD markers, the genetic diversity in populations of *I. mandshurica* and *I. vorobievii* corresponds with that in natural populations of the rare and endangered European steppe plant *Iris aphylla* (P = 30.6%, He = 0.097; Wróblewska & Bzosko, 2006) but is lower than that in the widespread species *I. humilis* (P = 48.1%, He = 0.168; Kozyrenko et al., 2009). Notably, the genetic diversity value in *I. humilis* is in accord with the average value for plants with mixed breeding systems (He = 0.18; Nybom, 2004), whereas in *I. mandshurica* and *I. vorobievii*, these values are significantly lower (Table). This may indicate the prevalence of vegetative propagation over propagation through seeds in the populations studied.

I. mandshurica and *I. vorobievii* show considerable levels of haplotype diversity, along with low levels of nucleotide diversity in the chloroplast genome, which is transmitted through seeds in *Iris* species (Cruzan et al., 1993). Low nucleotide diversity and high levels of cpDNA haplotype diversity have been found in some widespread (e.g., *I. humilis*, Kozyrenko et al., 2009) and endemic perennial herb species (e.g., *Aconitum gimnandrum*, Wang et al., 2009; and *Oxytropis chankaensis*, Artyukova et al., 2011b). However, for most angiosperms, including *Iris* species (e.g., Cornman & Arnold, 2007), populations are often fixed for single cpDNA haplotypes, and polymorphic populations possessing different haplotypes occur in potential contact zones of different maternal lines or at sites of long-

term persistence. In contrast, the populations of *I. vorobievii* and *I. mandshurica* contain 13 and 6 haplotypes, respectively, which may result from historical gene flow, retention of ancestral polymorphisms that accumulated over a long period in more continuous ancient populations, or a putative origin from several founders. Based on cpDNA haplotypes, demographic event analyses show that populations of both species have undergone bottleneck events and expansion in the past (Kozyrenko et al., 2009).

Apparently, the main reason for the rarity of *I. mandshurica* and especially of *I. vorobievii* is the scarcity of suitable habitats for these psammirises in Primorye. As *I. vorobievii* and *I. mandshurica* represent components of steppe vegetation and grow in specific edaphic conditions, they are members of several steppe communities that are likely remnants of previously more widespread steppe vegetation (Krestov & Verkholat, 2003). The contractions of such relic communities result from climate and natural community changes (natural succession) and, in recent years, from anthropogenic habitat destruction. Their isolation from the main part of the species' ranges, limited seed dispersal and poor vegetative reproduction make these psammirises particularly vulnerable and may lead the species to extinction.

3. Conclusion

It is usually assumed that a certain level of genetic diversity is necessary for the long-term prosperity of a species. Indeed, different genotypes confer different levels of resistance to various environmental stresses, and consequently, the greater the diversity of the genotypes in a population, the more effective its ability to withstand unfavorable conditions will be. Therefore, the existence of a low level of genetic diversity is often considered to represent a crucial stage for the survival of a species or even a sign of its extinction. However, despite several attempts to determine the average levels of genetic variation in different categories of plants (e.g., Gitzendanner & Soltis, 2000; Nybom, 2004), it is still unknown what level of polymorphism should be considered critical for the existence of a certain species. Thus, some rare endemic species exhibit higher levels of genetic diversity than the average values found for this category of plants, such as *Oxytropis chankaensis* (Table). At the same time, very low levels, or even an absence of genetic variability has been found in other narrowly endemic plants (e.g., *Bensoniella oregona*, Soltis et al., 1992). It is impossible to predict the fate of a species on the basis of its genetic diversity alone. Here, we have attempted to address some species existing under extreme conditions in the context of a variety of environmental, biological, evolutionary and other influences.

All of the species described above are represented by small, fragmented, marginal populations. Each of these species is characterized by a particular level of genetic diversity, ranging from very low (*Panax ginseng*) to high (*Oxytropis chankaensis*). Despite these differences, all of these plants exhibit weak competitiveness, and none of them can be considered prosperous. These species are not dominant in their respective plant communities, and they inhabit specific (often quite narrow) ecological niches where their existence is maintained more or less successfully for a long period. Without the effects of human activity, these species could probably exist in this state indefinitely. What are the mechanisms that ensure the long-term existence of these species in small, isolated populations at the limit of their climatic and environmental tolerance?

For ginseng, longevity appears to be most important factor for its survival. Indeed, a lifetime of up to several hundred years is unusual for an herbaceous plant that is incapable of vegetative reproduction. Ginseng maintains the ability to produce seeds throughout its life span, and it appears that as a result of apomixes, its seed production is not particularly dependent on environmental conditions or the availability of pollinators. A large number of fully viable seeds and an extended period of embryo maturation allow revival of populations of this species, even from a small number of individuals over the course of many years.

Similar to ginseng, *Aristolochia manshuriensis* only exhibits seed-based reproduction. The life span of this species is not as long as that of other woody plants and is near the average for woody vines. This species has a very poor ability to reproduce vegetatively, and there are major limitations on its seed reproduction as a result of adaptation to pollination only by certain insects. Though adaptation to specific pollinators prevents inbreeding, pollination decreases if pollinators are absent, and the resulting fruit development is low. In this case, high seed production (a large number of seeds per one fertilization event) and adaptability to seed transfer by wind and water seems to be of primary importance for the survival of the species. Even a single successful fertilization every several years could guarantee population restoration if conditions are suitable for seed germination. Under unfavorable conditions, plantlets of this species can exist in a juvenile state for a long time, which allows a population to survive until the re-establishment of suitable conditions.

The ability for vegetative propagation is a pathway for the survival of some species (*Oplopanax elatus*, irises and *Microbiota decussata*). This pathway also allows for the rehabilitation of a species, even after a significant reduction in its population, and it effectively allows clonal colonies to reoccupy their habitats, where suitable conditions for the species are associated with an absence of competitors. The flexibility of the reproductive system of *O. elatus*, combined with its different modes of reproduction allows it to renew heterozygous genotypes by clonal growth and to contribute additional variability resources through sporadic seed reproduction. At the same time, some features of the biology of this species (the long life of a single clone, overlapping generations and the ability to cross-pollinate) also help to maintain a certain level of polymorphism.

For *I. mandshurica* and *I. vorobievii,* vegetative reproduction (even if poor) seems to be the only way to survive under unfavorable conditions. Rare reproduction through seeds allows for the maintenance of a certain level of genetic diversity in its populations. However, the spread of both species beyond the borders of their existing populations is unlikely as a result of the scarcity of suitable habitats. Poor vegetative reproduction and limited seed dispersal as well as isolation from the main part of the species range (*I. mandshurica)* or its occurrence in a single locality (*I. vorobievii)* make populations of both species vulnerable. In the case of *I. vorobievii*, this vulnerability may lead to rapid species extinction.

The pattern of genetic diversity and the structure of populations in the endemic species *M. decussata* are congruent with the leading edge model of colonization. It could be proposed that extant southern populations represent the putative species range center. Populations that expanded southward during dry and cool periods at the Oligocene/Miocene boundary have become completely extirpated following climate changes in the Quaternary. The distribution area of *M. decussata* became fragmented quite some time ago through the displacement of the species toward mountain peaks and the extinction of stands in the

adjoining territories. The remnant *M. decussata* populations have the potential for survival under ongoing climate changes and global warming because of this species' physiological and ecological range of tolerance, reproduction by layering and other life history traits.

In the case of endemic or rare species, one must distinguish between relics left by the extinction of related populations and newly evolved taxa. All relics that survived the repeated periods of Pleistocene climate cooling have apparently experienced a severe genetic bottleneck, not only as a result of genetic drift associated with the reduction of populations, but also because of selection acting in a rapidly changing climate. Because of this selection, the fittest individuals have survived. Among the representatives of the ancient tropical floras (*P. ginseng, O. elatus, A. manshuriensis*) only a small number of genotypes are likely to have possessed such fitness. By acquiring a mechanism for survival in harsh environments, these species have lost much of their genetic diversity.

In the case of the narrowly endemic species *O. chankaensis*, the situation is different. This species possesses adaptive mechanisms that enable it not only to successfully renew its populations in the coastal zone, which are exposed to frequent flooding and other adverse factors, but also to maintain the high level of recombination responsible for the survival of the species in a changing environment. This is a relatively young species, and its biological characteristics may promote its prosperity and wide distribution; the only obstacle to this is its high habitat specificity; it only inhabits the sandy shore of a large lake where there is intense insolation and high air humidity. However, there are no such habitats nearby, and the small number of plants in some populations makes this species vulnerable.

The patterns observed for various rare species often do not fully correspond to the general idea of survival at the edge of their range. Plant species exhibit tremendous variation in life history traits that may help them survive in harsh or changeable environments. The cause of a plant's rarity depends on the effects of different historical, biological and genetic factors. In all cases, different compensatory mechanisms, such as increased longevity and fertility, the formation of soil seed banks and vegetative reproduction, are involved. The adaptations of these species are not always successful because the historically established balance between reproduction and dispersal can be disturbed. However, in the absence of destructive human activities, many rare species could exist for an indefinite time.

4. Acknowledgements

We are grateful to Dr. E.V. Sundukova and Dr. V.N. Makarkin for help with the manuscript preparation. This work was supported by the Grant "Molecular and Cell Biology" of the Russian Academy of Sciences (No 09-I-P22-03), by the Grant "Biological Diversity" of the Russian Academy of Sciences (No 09-I-P23-06), by the Grant "Biological Resources of Russia: Estimation and Fundamental Basis for Monitoring" (No 09-I-OBN-02), by the Grant No 11-III-B-06-094 of FEBRAS and by Grant “Leading Schools of Thought” from the President of Russian Federation.

5. References

Adams, C.A., Baskin, J.M. & Baskin, C.C. (2005). Trait Stasis Versus Adaptation in Disjunct Relict Species: Evolutionary Changes in Seed Dormancy–Breaking and

Germination Requirements in a Subclade of *Aristolochia* Subgenus *Siphisia* (Piperales). *Seed Sci. Res.*, Vol.15, No.2, (June 2005), pp. 161–173, ISSN 0960-2585

Alexeeva, N.B. (2008). Genus *Iris* L. (Iridaceae) in the Russia. *Turczaninowia*, Vol.11, No.2, (June 2008), pp. 5–68, ISSN 1560-7259

Alexeeva, N.B. & Mironova, L.N. (2007). Critical Notes on Some Species of the Genus *Iris* (Iridaceae) in Siberia and the Far East. *Bot. Zhurn.*, Vol.92, No. 6, (June 2007), pp. 916–925, ISSN 0006-8136

Allnutt, T.R., Newton, A.C., Premoli, A. & Lara, A. (2003). Genetic Variation in the Threatened South American Conifer *Pilgerodendron uviferum* (Cupressaceae), Detected Using RAPD Markers. *Biol. Conserv.*, Vol.114, No.2, (December 2003), pp. 245–253, ISSN 0006-3207

Artyukova, E.V., Gontcharov, A.A., Kozyrenko, M.M., Reunova, G.D. & Zhuravlev, Yu.N. (2005). Phylogenetic Relationships of the Far Eastern Araliaceae Inferred from ITS Sequences of Nuclear rDNA. *Russ. J. Genet.*, Vol.41, No.6, (June 2005), pp. 649–658, ISSN 1022-7954

Artyukova, E.V. & Kozyrenko, M.M. (2009). Genetic Diversity in the Endemic of Sikhote Alin, *Microbiota decussata* Kom. (Cupressaceae), Inferred from Nuclear and Chloroplast Genome Data, *Proceedings of the III International Conference "Mountain ecosystems and their components, Soil and Vegetation of Mountain Territories"*, pp. 119–125, ISBN 978-5-87317-581-9, Association of scientific publications of KMK, Moscow, Russia, August 24–29, 2009

Artyukova, E.V., Kozyrenko, M.M., Gorovoy, P.G. & Zhuravlev, Yu.N. (2009). Plastid DNA Variation in Highly Fragmented Populations of *Microbiota decussata* Kom. (Cupressaceae), an Endemic to Sikhote Alin Mountains. *Genetica*, Vol.137, No.2, (November 2009), pp. 201–212, ISSN 0016-6707

Artyukova, E.V., Kozyrenko, M.M., Kholina, A.B. & Aistova, E.V. (2011a). Genetic Relationships of *Oxytropis chankaensis* Jurtz. and *O. oxyphylla* (Pall.) DC. (Fabaceae) Inferred from Analyses of Nuclear and Chloroplast Genomes, *Proceedings of International symposium "Modern achievements in population, evolutionary, and ecological genetics"*, p. 10, ISBN 978-5-7442-1512-5, Vladivostok, Russia, June 19–24, 2011

Artyukova, E.V., Kozyrenko, M.M., Kholina, A.B. & Zhuravlev, Yu.N. (2011b). High Chloroplast Haplotype Diversity in the Endemic Legume *Oxytropis chankaensis* May Result from Independent Polyploidization Events. *Genetica*, Vol.139, No.2, (February 2011), pp. 221–232, ISSN 0016-6707

Bezdeleva, T.A., Mironova, L.N. & Dudkin, R.V. (2010). Features of Structural Organization of Some Species of the Genus *Iris*. *Bulletin of the Botanical Garden Institute FEB RAS*. No.5, pp. 21–25, ISSN 2222-5579

Bisby, F.A., Roskov, Y.R., Orrell, T.M., Nicolson, D., Paglinawan, L.E., Bailly, N., Kirk, P.M., Bourgoin, T. & Baillargeon, G. (Eds.). (2009). *Species 2000 & ITIS Catalogue of Life: 2009 Annual checklist.* Available from http://www.catalogueoflife.org/annual-checklist/2009/

Blokhina, N.I. (1988). Fossil Woods, In: *Cretaceous–Palaeogene of the Lesser Kuril Islands. New Data on Palaeontology and Geological History*, V.A. Krasilov, N.I. Blokhina, V.S. Markevitch & M.J. Serova (Eds.), 26–40, Far Eastern Branch of the Academy of Sciences of the USSR, Vladivostok, Russia

Bondarenko, O.V. (2006). Fossil Woods from the Pliocene of Southern Primorye. *PhD Thesis*, Institute of Biology and Soil Science, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, Russia

Bulgakov, V.P. & Zhuravlev, Yu.N. (1989). Generation of *Aristolochia manshuriensis* Kom. Callus Tissue Cultures. *Rastitel'nye Resursy*, Vol.25, No.2, (May 1989), pp. 266–270, ISSN 0033-9946

Cornman, R.S. & Arnold, M.L. (2007). Phylogeography of *Iris missouriensis* (Iridaceae) Based on Nuclear and Chloroplast Markers. *Mol. Ecol.*, Vol.16, No.21, (November 2007), pp. 4585–4598, ISSN 0962-1083

Cruzan, M.B., Arnold, M.L., Carney, S.E. & Wollenberg, K.R. (1993). cpDNA Inheritance in Interspecific Crosses and Evolutionary Inference in Louisiana Irises. *Am. J. Bot.*, Vol.80, No.3, (March 1993), pp. 344–350, ISSN 0002-9122

Ellstrand, N.C. & Roose, M.L. (1987). Patterns of Genotypic Diversity in Clonal Plant Species. *Am. J. Bot.*, Vol.74, No.1, (January 1987), pp. 123–131, ISSN 0002-9122

Figueroa-Esquivel, E.M., Puebla-Olivares, F., Eguiarte, L.E. & Núñez-Farfán J. (2010). Genetic Sructure of a Bird-Dispersed Tropical Tree (*Dendropanax arboreus*) in a Fragmented Landscape in Mexico. *Revista Mexicana de Biodiversidad,* Vol.81, No.3, (November 2010), pp. 789–800, ISSN 1870-3453

Gadek, P.A., Alpers, D.L., Heslewood, M.M. & Quinn, C.J. (2000). Relationships Within Cupressaceae Sensu Lato: a Combined Morphological and Molecular Approach. *Am. J. Bot.*, Vol.87, No.7, (July 2000), pp. 1044–1057, ISSN 0002-9122

Gitzendanner, M.A. & Soltis, P.S. (2000). Patterns of Genetic Variation in Rare and Widespread Plant Cogeners. *Am. J. Bot.*, Vol.87, No.6, (June 2000), pp. 783–792, ISSN 0002-9122

Grushwitsky, I.V. (1961). *Ginseng: the Aspects of Biology*, Nauka, Leningrad, Russia

Hampe, A. & Petit, R.J. (2005). Conserving Biodiversity under Climate Change: the Rear Edge Matters. *Ecol. Lett.*, Vol.8, No.5, (May 2005), pp. 461–467, ISSN 1461-0248

Hao, B., Li, W., Linchun, M., Li, Y., Rui, Z., Mingxia, T. & Weikai, B. (2006). A Study of Conservation Genetics in *Cupressus chengiana*, an Endangered Endemic of China, Using ISSR Markers. *Biochem. Genet.*, Vol.44, No.1-2, (February 2006), pp. 31–45, ISSN 0006-2928

Hewitt, G.M. (2004). Genetic Consequences of Climatic Oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.,* Vol. 359, No.1442, (February 2004), pp. 183–195, ISSN 1471-2970

Huang, S., Kelly, L.M. & Gilbert, M.G. (2003). *Aristolochia* Linnaeus., In: *Flora of China* (Ulmaceae through Basellaceae), 19.12.2003, Z.-Y. Wu, P.H. Raven & D.Y. Hong, (Eds.), Vol.5, , pp. 258–269, Available from http://www.efloras.org

Huh, M.K., Jung, S.D., Moon, H.K., Kim, S.-H. & Sung J.S. (2005). Comparison of Genetic Diversity and Population Structure of *Kalopanax pictus* (Araliaceae) and its Thornless Variant Using RAPD. *Korean Journal of Medicinal Crop Science*, Vol.13, No.2, (March 2005), pp. 69–74, ISSN 1225-9306

Hwang, S.-Y., Lin, H.W., Kuo, Y.S. & Lin, T.P. (2001). RAPD Variation in Relation to Population Differentiation of *Chamaecyparis formosensis* and *Chamaecyparis taiwanensis*. *Bot. Bull. Acad. Sin.,* Vol.42, No.3, (July 2001), pp. 173–179, ISSN 1817-406X

Hwang, S.-Y., Lin, T.P., Ma, C.-S. & Lin, C.-L. (2003). Postglacial Population Growth of *Cunninghamia konishii* (Cupressaceae) Inferred from Phylogeographical and Mismatch Analysis of Chloroplast DNA Variation. *Mol. Ecol.*, Vol.12, No.10, (October 2003), pp. 2689–2695, ISSN 0962-1083

Kim, C. & Choi, H.-K. (2003). Genetic Diversity and Relationship in Korean Ginseng (*Panax schinseng*) Based on RAPD Analysis. *Korean J. Genet.,* Vol.25, No.3, (September 2003), pp. 181-188, ISSN 0254–5934

Kim, S.H., Jang, Y.S., Han, J.G., Chung, H.G., Chung, H.G., Lee, S.W. & Cho, K.J. (2006). Genetic Variation and Population Structure of *Dendropanax morbifera* (Araliaceae) in Korea. *Silvae Genet*., Vol.55, No.1, (January 2006), pp. 7–13, ISSN 0037-5349

Kitagawa, M. (1979). *Neo-Lineamenta Florae Manshuricae*, J. Cramer, ISBN 3-7682-1113-4, Hirchberg, Germany

Kharkevich, S.S. & Kachura, N.N. (1981). *Rare Plant Species of the Soviet Far East and Their Conservation,* Nauka, Moscow, Russia

Kholina, A.B., Markelova, O.V. & Kholin, S.K. (2003). Population Structure and Reproduction Biology of the Rare Endemic Species *Oxytropis chankaensis* Jurtz., *Proceedings of the XI Congress of the Russian Botanical Society "Botanical Researches in Asian Russia"*, pp. 369–370, ISBN 5-93957-063-1, Novosibirsk–Barnaul, Russia, August 18–22, 2003

Kholina, A.B., Koren, O.G. & Zhuravlev, Yu.N. (2004). High Polymorphism and Autotetraploid Origin of the Rare Endemic Species *Oxytropis chankaensis* Jurtz. (Fabaceae) Inferred from Allozyme Data. *Russ. J. Genet.,* Vol.40, No.4, (April 2004), pp. 393–400, ISSN 1022-7954

Kholina, A.B., Koren, O.G. & Zhuravlev, Yu.N. (2009). Genetic Structure and Differentiation of Populations of the Tetraploid Species *Oxytropis chankaensis* (Fabaceae). *Russ. J. Genet.,* Vol.45, No.1, (January 2009), pp. 70–80, ISSN 1022-7954

Kholina, A.B. & Kholin, S.K. (2006). Population Age Structure of Rare Plant *Oxytropis chankaensis*, *Proceedings of the 2 International Conference "Problems of Preservation of Wetlands of International Meaning: Khanka Lake"*, pp. 26–35, ISBN 5-91162-002-2, Spassk-Dalny, Russia, June 10-11, 2006

Kholina, A.B., Nakonechnaya, O.V., Koren, O.G., & Zhuravlev, Yu.N. (2010). Genetic Variation of *Oplopanax elatus* (Nakai) Nakai (Araliaceae). *Russ. J. Genet.*, Vol.46, No.5, (May 2010), pp. 555–561, ISSN 1022-7954

Koren, O.G., Potenko, V.V. & Zhuravlev, Yu.N. (2003). Inheritance and Variation of Allozymes in *Panax ginseng* C.A. Meyer (Araliaceae). *Int. J. Plant Sci.*, Vol.164, No.1, (January 2003), pp. 189–195, ISSN 1058-5893

Koren, O.G., Krylach, T.Yu., Zaytseva, Yu.A. & Zhuravlev, Yu.N. (1998). Floral Biology and Embryology of *Panax ginseng* C.A. Meyer, *Proceedings of the 1st European Ginseng Congress,* pp. 221–231, S. Imhof., Marburg, Germany, December 6–11, 1998

Koren, O.G., Nakonechnaya, O.V. & Zhuravlev, Yu.N. (2009). Genetic Structure of Natural Populations of the Relict Species *Aristolochia manshuriensis* (Aristolochiaceae) in Disturbed and Intact Habitats. *Russ. J. Genet.*, Vol. 45, No.6, pp. 678–684, ISSN 1022-7954

Kozyrenko, M.M., Artyukova, E.V. & Zhuravlev Yu.N. (2009). Independent Species Status of *Iris vorobievii* N.S. Pavlova, *Iris mandshurica* Maxim., and *Iris humilis* Georgi

(Iridaceae): Evidence from the Nuclear and Chloroplast Genomes. *Russ. J. Genet.,* Vol.45, No.11, (November 2009), pp. 1394–1402, ISSN 1022-7954

Krestov, P.V. & Verkholat, V.P. (2003). *Rare Plant Communities of Amur Region.* Institute of Biology and Soil Sciences, ISSN 5744213406, Vladivostok, Russia

Kurentsova, G.E. (1968). *Relic Plants of Primorye*, Nauka, Leningrad, Russia

Little, D.P., Schwarzbach, A.E., Adams, R.P. & Hsieh, C.-F. (2004). The Circumscription and Phylogenetic Relationships of *Callitropsis* and the Newly Described Genus *Xanthocyparis* (Cupressaceae). *Am. J. Bot.*, Vol.91, No.11, (November 2004), pp. 1872–1881, ISSN 0002-9122

Malyshev, L.I. (2007). Phenetics in the Section Verticillares of the Genus *Oxytropis* DC. (Fabaceae). *Bot. Zhurn.*, Vol.92, No.6, (June 2007), pp. 793–807, ISSN 0006-8136

Malyshev, L.I. (2008). Diversity of the genus *Oxytropis* in Asian Russia. *Turczaninowia*, Vol.11, No.3, (December 2008), pp. 5–141, ISSN 1560-7259

Melnikova, A.B. & Machinov, A.N. (2004). On the Record of *Microbiota decussata* (Cupressaceae) at Unusually Low Altitude. *Bot. Zhurn.*, Vol.89, No.9, (September 2004), pp. 1470–14720, ISSN 0006-8136

Nakonechnaya, O.V., Koren, O.G., Nesterova, S.V., Sidorenko, V.S., Kholina, A.B., Batygina, T.B. & Zhuravlev, Yu.N. (2005). Elements of Reproductive Biology of *Aristolochia manshuriensis* Kom. (Aristolochiaceae) in the Conditions of Introduction. *Rastitel'nye Resursy*, Vol.41, No.3, (August-September 2008), pp. 14–25, ISSN 0033-9946

Nakonechnaya, O.V., Koren, O.G. & Zhuravlev, Yu.N. (2007). Allozyme Variation of the Relict Plant *Aristolochia manshuriensis* Kom. (Aristolochiaseae). *Russ. J. Genet.*, Vol.43, No.2, (February 2007), pp. 156–164, ISSN 1022-7954

Nakonechnaya, O.V., Sidorenko, V.S., Koren, O.G., Nesterova, S.V. & Zhuravlev, Yu.N. (2008). Specific features of pollination in the Manchurian birthwort, *Aristolochia manshuriensis. Biol. Bull.* Vol.35, No.5, (October 2008), pp. 459–465, ISSN 1062-3590

Nechaev, V.A. & Nakonechnaya, O.V. (2009). Structure of Fruits and Seeds and Ways of Dissemination of Two Species of the Genus *Aristolochia* L. in Primorsky Krai. *Biol. Bull.*, Vol.36, No.4, (August 2009), pp. 393–396, ISSN 1062-3590

Nybom, H. (2004). Comparison of Different Nuclear DNA Markers for Estimating Intraspecific Genetic Diversity in Plants. *Mol. Ecol.,* Vol.13, No.5, (May 2004), pp. 1143–1155, ISSN 0962-1083

Pavlova, N.S. (1987). Family Iridaceae, In: *The vascular plants of the Soviet Far East,* Vol.2, S.S. Kharkevich, (Ed.), pp. 414–426, Nauka, Leningrad, Russia

Pavlova, N.S. (1989). Family Fabaceae. In: *The vascular plants of the Soviet Far East,* Vol.4, S.S. Kharkevich, (Ed.), pp. 191–339, Nauka, ISBN 5-02-026577-2, Leningrad, Russia

Pavlova, N.S. (2006). Family Iridaceae. *Iris* L., In: *Flora of the Russian Far East, Additions and Modifications to "Vascular Plants of the Soviet Far East",* A.E. Kozhevnikov, N.S. Probatova, (Eds.), pp. 277–279, Dal'nauka, ISBN 5-8044-0534-9, Vladivostok, Russia

Pavlyutkin, B.I., Petrenko, T.I. & Chekryzhov, I.Y. (2005). The Problems of Stratigraphy of the Tertiary Deposits of the Pavlovka Coal-field (Primorye). *Tikhookeanskaya Geologiya*, Vol.24, No.6, (November 2005), pp. 59–76, ISSN 0207-4028

Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. & Vendramin, G.G. (2005). Comparative Organization of Chloroplast, Mitochondrial and Nuclear Diversity in Plant Populations. *Mol. Ecol.,* Vol.14, No.3, (March 2005), pp. 689–701, ISSN 0962-1083

Probatova, N.S., Seledets, V.P. & Rudyka, E.G. (2008a). *Oxytropis chankaensis*. In: *Taxon* Marhold K. IAPT/IOPB chromosome data 5. Vol.57, No.2, (May 2008), pp. 560, E20, ISSN 0040-0262

Probatova, N.S., Rudyka, E.G., Seledets, V.P. & Nechaev, V.A. (2008b). *Iris vorobievii* In: *Taxon* Marhold K. IAPT/IOPB chromosome data 6. Vol.57, No.4, (November 2008), pp. 1270, E8, ISSN 0040-0262

Red Data Book, Primorsky Krai: Plants. (2008). AVK Apelsin, ISBN 978-5-98137-017-5, Vladivostok, Russia

Red Data Book of the Russian Federation: Plants. (1988). Rosagropromizdat, ISBN 5-260-00254-7, Moscow, Russia

Reunova, G.D., Kats, I.L., Muzarok, T.I. & Zhuravlev, Yu.N. (2010a). Polymorphism of RAPD, ISSR and AFLP Markers of the *Panax ginseng* C. A. Meyer (Araliaceae) Genome. *Russ. J. Genet.,* Vol.46, No.8, (August 2010), pp. 938–947, ISSN 1022-7954

Reunova, G.D., Kats, I.L., & Zhuravlev, Yu.N. (2010b). Genetic Variation of *Oplopanax elatus* (Araliaceae) Populations Estimated Using DNA Molecular Markers. *Doklady Biological Sciences.* Vol.433, No.1, (August 2010) pp. 252–256, ISSN 0012-4966

Sakaguchi, S., Sakurai, S., Yamasaki, M. & Isagi, Y. (2010). How Did the Exposed Seafloor Function in Postglacial Northward Range Expansion of *Kalopanax septemlobus*? Evidence from Ecological Niche Modeling. *Ecol. Res.*, Vol.25, No.6, (July 2010), pp. 1183–1195, ISSN 0912-3814

Schreter, A.I (1975). Medicinal Flora of the Soviet Far East. Medicine, Moscow, Russia

Shlotgauer, S.D. (2011). Characteristics of the Alpine vegetation of the Sikhote-Alin Range Illustrated by Ko Mountain (Khabarovsk Krai). *Contemp. Probl. Ecol.,* Vol.4, No.2, (April 2011), pp. 159–163, ISSN 1995-4255

Shu, Y.W. (2000). *Iris* Linnaeus. In: *Flora of China (Flagellariaceae through Marantaceae)*, n.d, Z.-Y. Wu, P.H. Raven & D.Y. Hong, (Eds.), Vol.24, pp. 297–313, Available from http://www.efloras.org

Shulgina, V.V. (1955). Woody Vines and Their Culture in Leningrad, In: *Introduction of Plants and Green Building*, S.Ya. Sokolov, (Ed.), pp. 157–194, Publishing House of Academy of Sciences, Moscow-Leningrad, Russia

Soltis, P.S., Soltis, D.E. & Tucker, T.L. (1992). Allozyme Variability Is Absent in the Narrow Endemic *Bensoniella oregona* (Saxifragaceae). *Conserv. Biol.*, Vol.6, No.1, (March 1992), pp. 131–134, ISSN 0888-8892

Urusov, V.M. (1979). Ecological and biological peculiarities of *Microbiota decassata* Kom. (Cupressaceae). *Bot. Zhurn.*, Vol.64, No.3, (March 2004), pp. 362–376, ISSN 0006-8136

Urusov, V.M. (1988). *Genesis of Vegetation and Conservancy in the Far East*. Far Eastern Branch of the Academy of Sciences of the USSR, Vladivostok, Russia

Vucetich, J.A. & Waite, T.A. (2003). Spatial Patterns of Demography and Genetic Processes Across the Species' range: Null hypotheses for landscape conservation genetics. *Conserv. Genetics*, Vol.4, No.5, (September 2003), pp. 639–645, ISSN 1572-9737

Wang, L., Abbott, R.J., Zheng, W., Chen, P., Wang, Y. & Liu, J. (2009). History and Evolution of Alpine Plants Endemic to the Qinghai-Tibetan Plateu: *Aconitum gimnandrum* (Ranunculaceae). *Mol. Ecol.*, Vol.18, No.4, (February 2009), pp. 709–721, ISSN 0962-1083

Watkinson, A.R. & Powell, J.C. (1993). Seedling Recruitment and the Maintenance of Clonal Diversity in Plant Population – a Computer Simulation of *Ranunculus repens*. *J. Ecol*., Vol.81, No.4, (December 1993), pp. 707–717, ISSN 00220477

Wolf, A.T., Howe, R.W. & Hamrick, J.L. (2000). Genetic Diversity and Population Structure of the Serpentine Endemic *Calystegia collina* (Convolvulaceae) in Northern California. *Am. J. Bot.*, Vol.87, No.8, (August 2000), pp. 1138–1146, ISSN 0002-9122

Wróblewska, A. & Bzosko, E. (2006). The Genetic Structure of the Steppe Plant *Iris aphylla* L. at the Northern Limit of its Geographical Range. *Bot. J. Linn. Soc*., Vol.152, No.2, (October 2006), pp. 245–255, ISSN 0024-4074

Xiang, Q.B. & Lowry, P.P. (2007). Araliaceae, In: *Flora of China*, 1.07.2007, Z.-Y. Wu, P.H. Raven, D.Y. Hong, (Eds.), Vol.13, pp. 435–491, Available from http://www.efloras.org

Xie, G.W., Wang, D.L., Yuan, Y.M. & Ge X.-J. (2005). Population Genetic Structure of *Monimopetalum chinense* (Celastraceae), an Endangered Endemic Species of Eastern China. *Ann. Bot.-London.,* Vol.95, No.5, (April 2005), pp. 773–777, ISSN 0305-7364

Yurtsev, B.A. (1964). Conspectus of the System of Section *Baicalia* Bge., Genus *Oxytropis* DC. In: *News in the systematics of higher plants*. I.A. Linchevskii, (Ed.), pp. 191–218, Nauka, Moscow-Leningrad, Russia

Yurtsev, B.A. & Zhukova, P.G. (1968). Polyploid Series and the Taxonomy (on the Basis of the Analysis of Some Groups of Arctic Legumes). *Bot. Zhurn*., Vol.53, No.11, (November 1968), pp. 1531–1542, ISSN 0006-8136

Zhu, X.Y., Welsh, S.L. & Ohashi, H. (2010). *Oxytropis*, In: *Flora of China (Fabaceae)*, 16.02.2010, Z.-Y. Wu, P.H. Raven & D.Y. Hong, (Eds.), Vol.10, pp. 453–500, Available from http://www.efloras.org

Zhuravlev, Yu.N. & Kolyada, A.S. (1996). *Araliaceae: Ginseng and Others*, Dalnauka Press, ISBN 5-7442-0867-4, Vladivostok, Russia

Zhuravlev, Yu.N., Koren, O.G., Reunova, G.D., Muzarok, T.I., Gorpenchenko, T.Yu., Kats, I.L. & Khrolenko, Yu.A. (2008). *Panax ginseng* Natural Populations: Their Past, Current State and Perspectives. *Acta Pharmacol. Sin*., Vol.29, No.9, (September 2008), pp. 1127-1136, ISSN 1671–4083

Zhuravlev, Yu.N., Reunova, G.D., Kats, I.L., Muzarok, T.I. & Bondar, A.A. (2010). Molecular Variation of Wild *Panax ginseng* C.A. Meyer (Araliaceae) by AFLP Markers. *Chin. Med.*, Vol.5, No.21, (June 2010), ISSN1749-8546 (Electronic), Available from http://www.cmjournal.org/

19

Founder Placement and Gene Dispersal Affect Population Growth and Genetic Diversity in Restoration Plantings of American Chestnut

Yamini Kashimshetty[1], Melanie Simkins[2], Stephan Pelikan[3] and Steven H. Rogstad[1]
[1]*Department of Biological Sciences*
[2]*Department of Environmental Studies*
[3]*Department of Mathematical Sciences*
University of Cincinnati, OH
USA

1. Introduction

The American chestnut, *Castanea dentata* [Marsh.] Borkh (Fagaceae), was an abundant canopy tree inhabiting the mixed mesophytic forests of eastern North America. The species was struck by a fungal pathogen (*Cryphonectria parasitica* [Murrill] Barr*)* introduced from East Asia in the late 1800's on imported Asian chestnut material with the consequence that billions of trees have been destroyed (Barakat *et al.,* 2009; Pierson *et al.,* 2007; Elliot & Swank 2008; Jacobs 2007; Stilwell *et al.,*2003; Paillet, 2002; Huang *et al.,* 1998; Russel, 1987). The near elimination of this once important species has had widespread effects on the ecological functioning of eastern North American forests, and has also had a severe impact on economic forest extraction practices (e.g., strong workable lumber; chestnuts as food and forage). *Castanea dentata* has escaped complete elimination from its native range by persisting as occasional sprouts from the root collar of trees damaged by the blight; these sprouts rarely reach full sexual maturity (Jacobs, 2007; Stilwell *et al.,*2003; Paillet, 2002). Efforts are underway to restore this former keystone species and prevent extinction, reestablishing its ecological and economic roles in its natural habitat. Breeding of blight-resistant strains is being attempted (e.g., by the American Chestnut Foundation, TACF) through a series of initial hybridizations between *C. dentata* and the Chinese chestnut *(C. mollisima* Blume), followed by a series of backcrosses with the American chestnut always selecting for blight-resistance, to develop strains that are predominantly American chestnut in genotypic constitution but which retain Chinese blight-resistance genes. For example, some of the most recent blight-resistant strains ready for re-introduction are calculated to be genetically 94% American chestnut and 6% Chinese chestnut (Jacbos, 2007; Diskin *et al.,* 2006).

The conservation of endangered plant species often involves restoring these species back to their natural habitats and/or *ex situ* rescue plantings (Merritt & Dixon, 2011), and the American chestnut is no exception. Considerable resources are being expended in generating blight-resistant strains of *C. dentata*, thus necessitating optimization of restoration programs which are usually labor-intensive and expensive. Costs of such restoration efforts

include propagule generation and collection, storage, treatment, site preparation, planting, protecting, provisioning, travelling to and from the introduction site, monitoring, and future manipulations of individuals (Rogstad & Pelikan, 2011).

One approach to optimizing restoration of the American chestnut is to use computer programs to model the population growth and genetic effects of restoring plant populations in different ways. While restoration programs have been widely undertaken for a number of plant species, we lack the tools to analyze how factors like founder number and geometry of placement within a restoration habitat interact with varying founder or species life history characteristics, and whether these factors impact resultant population growth rates and preservation of genetic diversity. While ecological edge effects have been documented as being evident at the borders and edges of restoration preserves due to altered micro-environments in terms of wind-speeds, light availability and organismal composition among other factors (Primack, 2010), less is known about whether placement of founders at varying distances from the preserve edge impacts the population dynamics and genetic diversity measures of establishing populations. Further, can differing founder placement patterns interact with other life-history characteristics such as pollen and offspring dispersal distances to bring about different population growth rates and genetic diversity levels? Exploring these potential demographic and genetic edge effects in the field with *C. dentata* would not be feasible due to the costs associated with carrying out such experiments on a wide-scale to ensure statistical reliability of data, especially since blight-resistant individuals are expensive and time-consuming to produce, and thus must be used judiciously. In this case modeling virtual populations through computer simulations represents a more tractable alternative to field experiments, potentially providing valuable insight for restoration managers on how best to re-introduce the American chestnut into preserves.

In this study, the computer program NEWGARDEN (Rogstad & Pelikan, 2011) was used to model blight-resistant American chestnut population restoration in a virtual preserve to explore the population growth and genetic effects of placing founders at different distances from preserve borders under differing patterns of gene dispersal. Our null hypothesis is: Varying American chestnut founder placement at various diagonal distances from the preserve border, while altering offspring and pollen dispersal distances, will have no effect on population growth rates or retention of founding genetic diversity. We used comparative trials to examine the degree to which some patterns of introduction might be preferable over others.

2. Methods

We used NEWGARDEN to model the population growth and genetic diversity of newly establishing chestnut populations. This program simulates natural population development based on a set of user-specified initial input conditions (Rogstad & Pelikan, 2011). One set of input conditions constitutes a "trial" and input between trials may vary as to founder number and geometric patterning, preserve characteristics, and with regard to various life-history characteristics (see below). For each age (bout of mating with mates chosen randomly as conditioned by the input) of the developing population, NEWGARDEN provides output statistics (for each new cohort and for the entire population; only the latter are given here) concerning the population size, the number of founding alleles retained, heterozygosity (observed and expected), and F_{it} (a measure reflecting deviation from Hardy-Weinberg expectations due to subdivision and/or inbreeding; hereafter referred to as F

value). For each set of input trial conditions, the user can specify the number of replicate runs of those conditions to be used to calculate a mean and standard deviation for each of the listed output statistics. The user can thus compare whether means of statistics differ for populations of contrasting trials that differ in one or more input conditions.

2.1 Input parameters held constant

The input parameters common to all experimental runs are described following their order in the NEWGARDEN input file. These constant input parameters are based on population development data obtained from the only known chestnut restoration stand which is naturally developing at West Salem, Wisconsin (for details, see Pierson *et al.,* 2007; Rogstad & Pelikan, 2011).

2.1.1 The preserve grid system

Populations for all trials develop on a Cartesian grid system defined by the user. Individuals can only establish on grid points, the spacing of which represents the average minimum distance that can exist between reproductive trees. In our simulations, the grid system represents a restoration preserve (a 5 km x 5 km square), comprising a 1000 x 1000 grid point system with 5 m between grid points. This simple model represents an approximate average minimum distance spacing needed between two reproductive trees based on a mature chestnut community at the carrying capacity.

2.1.2 Loci and allelic variation

The ideal, non-inbred source population for the founders was set to have 30 loci per founder, each with 100 alleles of equal frequency (frequency = 0.01 per allele). NEWGARDEN randomly draws two alleles for each locus from the source population when creating a founder's genotype. In each new generation, every individual was censused for its allelic status at each of the 30 loci to generate the genetic diversity output statistics.

2.1.3 Mating system

All individuals are bisexual with 100% outcrossing.

2.1.4 Offspring production

The age-specific reproduction rate (r) specified in all input files (based on population development at West Salem) was held constant as shown in Table 1. In the intermediate years not defined, the rate of offspring production is linearly interpolated between the bounding age values. An established offspring does not reach reproductive age until its eighth year (Table 1). Offspring production is distributed across eligible reproducing individuals according to a Poisson distribution. As in natural settings, reproductive individuals may create multiple offspring, but all of these may not establish into saplings. Offspring counted in a generation's reproduction rate are potential recruits for establishment and growth at a grid point. The number of such potential recruits does not indicate the exact size of the newly established cohort for any particular age in one bout of mating. See mortality for ways in which potential offspring "die" or fail to establish in NEWGARDEN.

2.1.5 Age-specific pollen rate

For established individuals, the relative rate of serving as an eligible pollen donor that contributes pollen to a given mating is conditioned by its age according to Table 1. Pollen production is proportional relative to the highest value in the input. For the ages that are not specified by the input values, the pollen rate is linearly interpolated. Once NEWGARDEN selects an age-class and distance class (see below) for the pollen donor, one donor is chosen at random for a particular mating from the pool of eligible potential donors.

2.1.6 Mortality

Individuals can "die" in a number of different ways during a trial run. Death means that the individual is removed from the grid system and from further population processes and analyses. If the offspring is distributed outside of the defined preserve grid system, it automatically dies. An offspring will die in the event that it lands on a grid point that is already occupied by an individual that will survive to the next generation. When an ovule isn't pollinated since an eligible pollen donor is not within the range of the ovule producer, this counts as a reproductive event followed by immediate death of the potential offspring. If two or more offspring land at the same grid point, one is randomly chosen as the survivor with others deleted. Death of individuals is also age-dependent as specifiable by the user. For all trials here, age-specific probabilities of dying are given in Table 1. Ages without a specified risk of death are linearly interpolated between bounding values. In our trials, founders reach age 114, so there is a probability that several founders are still alive by the end of the trial run.

Age	Relative rate of reproduction	Relative rate of pollen provisioning	Percent mortality
0	0	0	0.3
1	0	0	0.2
4	0	0	0.1
5	0	0	
6	0	0.2	
7	0		0.05
8	0.02		
10	0.04	0.5	0.02
12	0.08		
17	0.2		
25		0.95	
70	1.3		
75	1.3	1.2	
115	1.3	1.16	0.02

Table 1. Input Parameters Held Constant. User specified age-specific input conditions for reproductive rate, pollen provisioning, and mortality used in all trials. Values not given were interpolated between the bounding values. See text for more details.

2.1.7 Initial population

Aging of the population begins with the founders and it will continue to age until the specified number of generations is completed. 169 founders, this being approximately the minimum number of founders needed to capture > 95% of the source population alleles (Lawrence *et al*., 1995; Chakraborty, 1993), of initial age 13 were used with founders located every other grid point in a 13 x 13 individual square (26 x 26 grid points). This simulates a 10 m distance between founders in a planting square 130 m on each side. This square of founders remained constant across all trials, although it was moved to different distances from the preserve borders.

2.1.8 Generations

Beginning with the founders, 100 bouts of reproduction (101 generations or ages including the founding generation) were conducted for each set of trial conditions. Output results in Figures 1 to 3 are reported for the entire population at each age. For the pollen and offspring dispersal distance comparison trials (Figures 4 to 6) only the results at the terminal age 100 are given.

2.1.9 Replicate runs

The output values reported are mean values calculated from 40 replicate runs for each set of trial input conditions. When trials are said to differ, we mean that reported generation mean values differ significantly (p value ≤ 0.05) unless otherwise stated. This convention is used rather than reporting each standard deviation and t-test results for all trial comparisons.

2.2 Variable input parameters

2.2.1 Region and founder placement

In the first series of comparative trials, the square of 169 founders was placed at increasing diagonal distances from the preserve border beginning with the founders at the lower left corner (trial a). Seed and pollen dispersal distances were held constant at 25% to or from each distance frame (see below) in this initial series. In a subsequent series of trials, the constant-sized founder square was likewise moved diagonally toward the center of the preserve while offspring and pollen dispersal distances differed in various combinations (Table 2).

2.2.2 Offspring dispersal distribution

Offspring dispersal on the grid system is based on a nested series of limiting frames called "distance frames". When an offspring is dispersed to a particular distance frame based on given probabilities, the lower and upper values of each frame define the limits of dispersal. For a given dispersal event, one point within the selected distance frame is chosen at random. The limits of the four frames used in our trials and their varying respective dispersal probabilities across trials are described in Table 2. In the "Basic Conditions" trials, offspring establishment was evenly divided to the four distance frames with 25% probability of dispersal to each frame. Modifications to the "basic" trial conditions to create alternate comparative trials are detailed in Table 2. We use the summary phrase "Offspring

less distant" to indicate restricted offspring dispersal (compared to the "Basic Conditions" trials) within 65 m of the parent (13 grid points), whereas in "Offspring least distant" trials, the majority of offspring dispersal was within 30 m (6 grid points) of the parent plant. The summary phrase "Offspring more distant" is used for the trials in which the greatest percentage of offspring were dispersed between distances 110 m to 1500m (22 to 300 grid points) from the parent plant (Nathan *et al.*, 2008; Jansen *et al.*, 2008).

2.2.3 Pollen transport distances

The "Basic Conditions" and "Offspring less/least/more distant" trials had equal probability of pollen dispersal from within each of the four distance frames (Table 2). Trial E has 90% of pollen dispersal limited to from within a 30 m frame of the producing individual. Trial F has 90% of pollen dispersed from 110 m to 1500 m (22 to 300 grid points) from the parent plant.

2.3 Output

Based on population characteristics developing from the initial input specifications, NEWGARDEN provides means for four statistical measures reported here: population size, mean number of founding alleles retained, observed heterozygosity, and F_{it} (or F value) calculated as:

$$F_{it} = 1 - H_{ob} / H_{ex} \quad (1)$$

where H_{ob} equals observed heterozygosity (based on actual counts of heterozygous loci across all loci for all the individuals in the population) and H_{ex} equals the Hardy-Weinberg expected heterozygosity based on the allele frequencies across all loci in the entire population. In general, F_{it} increases as inbreeding and/or subdivision increase in the population and $F_{it} = 0$ in the absence of inbreeding. For the first series of trials (Figures 1 through 3) mean output values are reported for the total number of individuals in the population at the end of each age. For the second series of trials (Figures 4 through 6), values are given only for the total population after 100 bouts of mating.

3. Results

Under the "basic" conditions of offspring and pollen dispersal (25% to or from each frame respectively; see Table 2), mean population size increased with increasing diagonal distance of the standard square of 169 founders from the preserve corner (Figure 1). At age 100, founders situated at the corner (trial a) had the lowest mean population size (2,018 individuals), while the equally highest mean population sizes (approximately 7,000) were attained by populations with founders inset by 300, 400 and 500 grid points from the corner (trials d, e, and f, respectively). Even an inset distance of 100 grid points (ca. 500 m; trial b) increased the mean population size by 148%, compared to founders placed at the corner. The percent gain in population increase per unit inset distance declines as the inset distances increases beyond 100 grid points, only increasing by approximately 21% when inset distances increases from 100 to 200 grid points, and approximately 13% when inset distances increase from 200 to 300 grid points. Beyond 300 grid points the rise in mean population size was not statistically different.

		Frame dimensions (in grid points)							
		0-5		6-12		13-22		22-301	
		%Pollen	% Offspring	%Pollen	% Offspring	%Pollen	% Offspring	%Pollen	% Offspring
Trial Description	Trial	from	to	from	to	from	to	from	to
Basic conditions	A	25	25	25	25	25	25	25	25
Offspring less distant	B	25	60	25	30	25	4	25	6
Offspring least distant	C	25	90	25	5	25	3	25	2
Offspring more distant	D	25	2	25	3	25	15	25	80
Pollen less distant, Offspring more distant	E	90	2	5	3	3	15	2	80
Pollen more distant, Offspring least distant	F	2	90	3	5	15	3	80	2

Table 2. Percent offspring and pollen dispersed to or from each distance frame for NEWGARDEN trials described further in the text. Trials were otherwise identical.

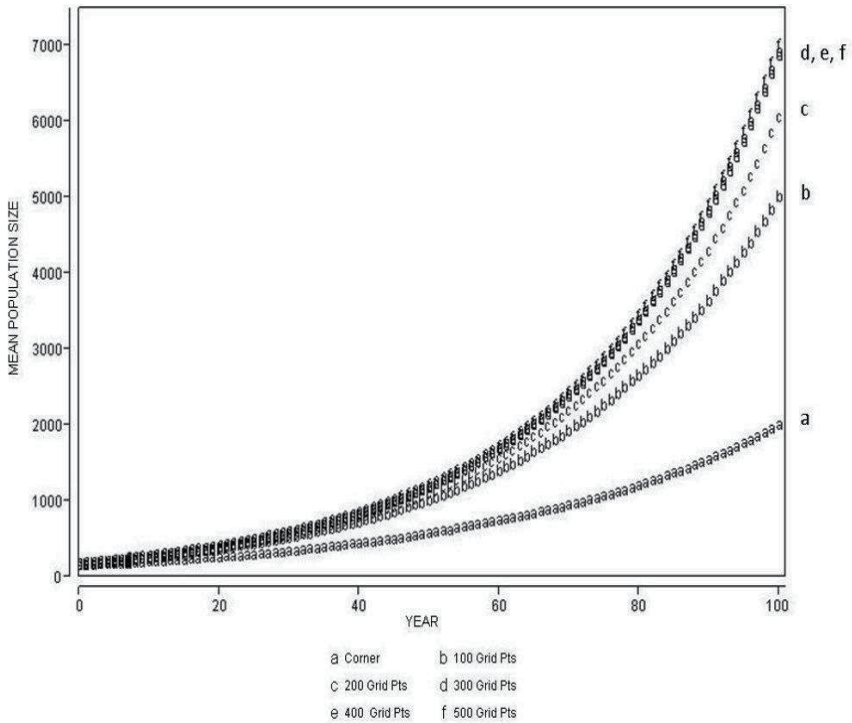


Fig. 1. Mean population sizes for each age of American chestnut populations founded at differing diagonal inset distances. These comparative trials span 101 generations at the basic offspring and pollen dispersal conditions (25% to and from each frame respectively). See text for more details.

Figure 2 shows the mean number of founding alleles retained across 101 generations for the trials depicted in Figure 1. In the source population, 3000 alleles each at frequency of 0.01, were available for the founding population. Drawing 169 founders, approximately 2900 alleles were present at founding (97% of the source population alleles; Figure 2). For example, imagine a restoration project where 1,690 trees are planted with 90% attrition prior to reproduction (e.g., Primack & Miao, 1992). On average such a population would have approximately 97% of the alleles in the original source population. NEWGARDEN can be used to estimate founding population sizes needed given the effects of attrition, to provide target numbers of individuals for establishing new populations or supplemention as needed.

For all populations, the number of alleles retained declined through generations, with populations founded at greater inset distances retaining higher numbers of alleles (Figure 2). The 169 founders were all aged 13 at the beginning of the trials, and given the mortality rate of 2% per generation (Table 1), approximately 22 founders should remain in populations after 100 bouts of mating. This suggests that a significant proportion of the total founding alleles remaining in the population at age 100 are carried by the descendents of founders that have died.

After 100 bouts of mating, there was approximately a 6% decline in number of founding alleles retained for populations inset at least 100 grid points or more (trials b, c, d, e and f), compared to the 9.6% decline seen for populations situated at the corner (trial a). At year 100, placing the founders 100 or more grid units inward from the corner produced a 3.8% (at the least; trials b through e) increase in alleles retained compared to placing the founders at the corner (trial a). The differences between various inset distances from 100 to 500 grid points (trials b through f) were not statistically different (although notice the trend). Variations of inset distances under basic dispersal conditions did not have significant or biologically meaningful effects on observed heterozygosity levels (ranging from 0.97 to 0.99) across generations, at basic dispersal conditions (data not shown).

There was little population subdivision and inbreeding observed across generations under basic dispersal conditions as founders were placed at greater inset distances (Figure 3). At year 100, F values increased slightly across generations, with the highest F values (almost 0.02) being reached by founders inset the furthest (by 500 grid points; trial f). F values tended to be lower for decreasing inset distances (e.g., trials a and b) at year 100. Although all trials are increasing in F, only by increasing the number of generations could it be determined the degree to which F values might become important (e.g., approach 0.05) in future generations.

Next, we ran a series of comparative trials varying not only in the inset distance of the square of 169 founders, but also altering the offspring and pollen dispersal conditions relative to the basic conditions (25% to or from each frame respectively) used in the trials just reviewed in Figures 1 through 3. We conducted this next trial series to investigate how differing types of gene dispersal might interact with founder distance from a border to affect population growth and genetic diversity.

Figure 4 shows population growth results (at population age 100 only) from these trials when both inset distances and dispersal conditions were varied. Trends observed here included increasing mean population sizes as inset distances increased from the corner to

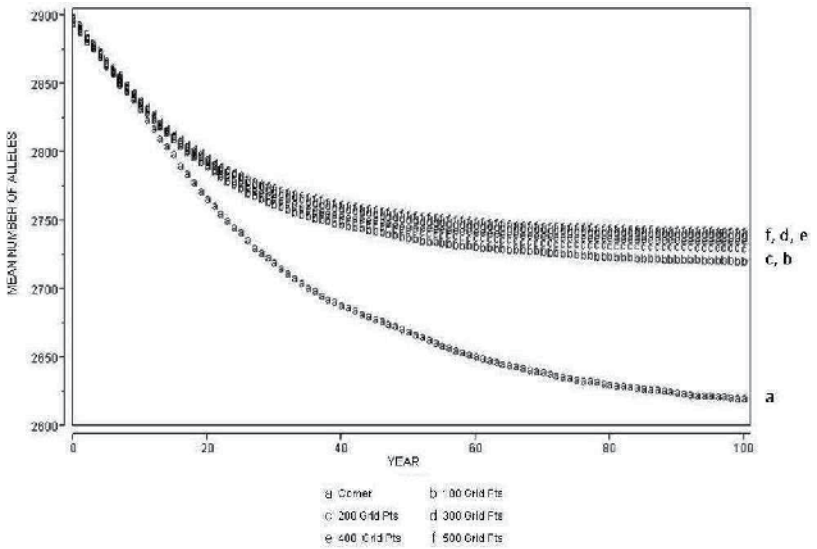


Fig. 2. Mean number of founding alleles retained across 101 generations for trial populations differing only in the distance to which founders were inset from a preserve border (population sizes shown in Figure 1). For all of these trials, the "basic" conditions of offspring and pollen dispersal distances were used (25% to or from each frame respectively). See text for more details.

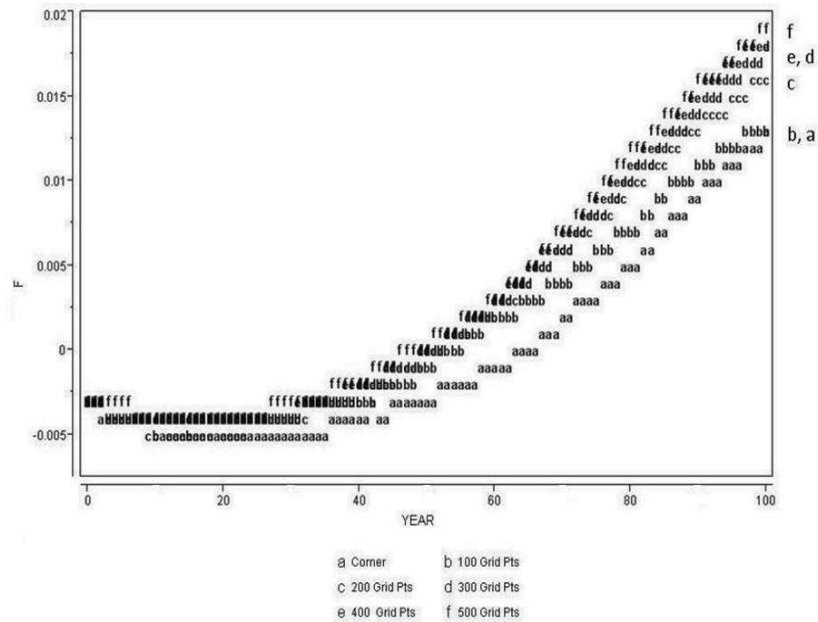


Fig. 3. F values across 101 generations for trials differing only in founder inset distance. These data are from the same trials depicted in Figures 1 and 2. "Basic" offspring and pollen dispersal conditions were used in each trial (25% to or from each frame respectively). See text for more details.

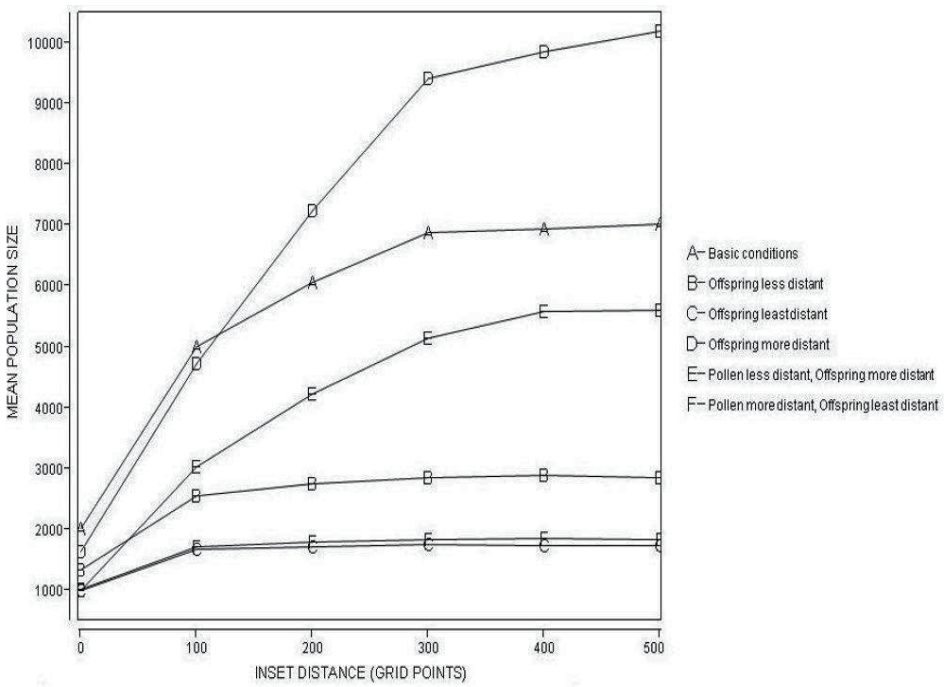


Fig. 4. Mean population size differences for comparative trial populations differing as to both founder inset distances (x-axis) and with greater or lower offspring and/or pollen dispersal distances (indicated by different letters A through F) relative to the "basic" dispersal conditions (see Table 2). Data points depict population size for each trial at age 100 only. In the trial condition summaries to the right of the graph, if a dispersule type (i.e., offspring versus pollen) is not mentioned, that dispersule disperses according to the basic conditions.

500 grid points for population D, up to 400 grid points (E), up to 300 grid points (A), and up to 100 grid points (B, C, and F). For populations B, C and F, increasing founder inset distance to 100 grid points from the corner raised mean population size significantly. However beyond 100 grid points, population sizes remained more or less constant across further increase in inset distances. If actual dispersal patterns match those in trials B, C, or F, a restoration manager will not gain higher rates of population growth by planting founders beyond 100 m into the preserve. Relative to the "basic" dispersal conditions trials just discussed in Figures 1 through 3 (trials marked A in Figure 4), altering offspring and pollen dispersal distances caused various but pronounced differences in rates of population growth. Trial D, with only the offspring dispersing to greater distances, was the only trial that exceeded trial A ("basic" conditions) at founder inset distances greater than 100 grid points. Trial D showed the greatest overall population growth compared to all trials at those inset distances.

For each set of trial conditions (A through F) the greatest allele loss occurred when founders were placed at the preserve corner (Figure 5). More alleles were retained under otherwise

constant trial conditions when founders were inset only 100 grid points although even further inset distances caused no further major increases in allele retention for trial conditions A, B, C, and F, while inset distances of 300 grid points were needed to maximize allele retention for trial conditions D and E. Considering all of these trial conditions, the greatest difference in the mean number of founding alleles retained is between trial E at the corner and trial D at 400 grid points (a 10.9% difference).

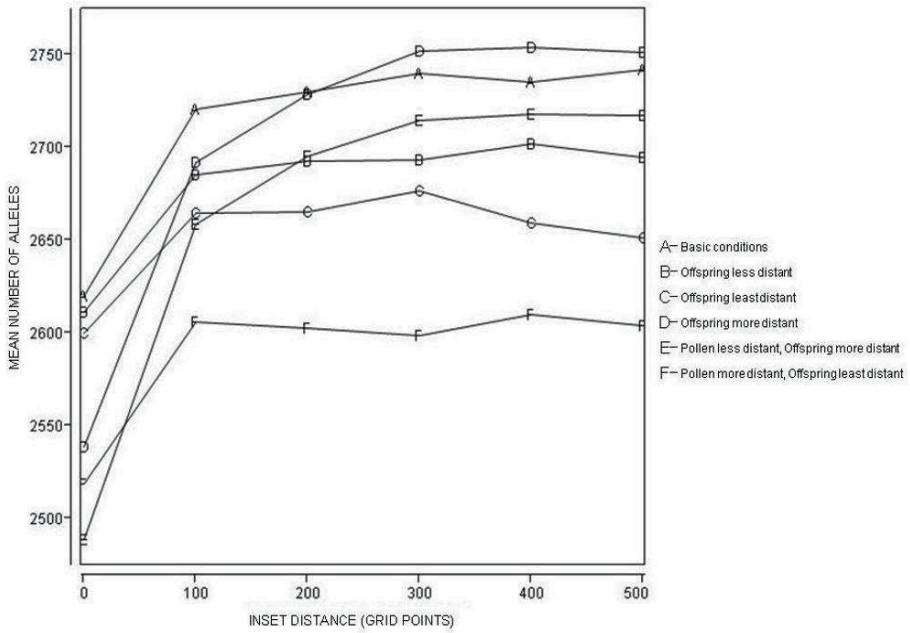


Fig. 5. Change in mean number of founding alleles retained in NEWGARDEN populations after 100 bouts of mating for trials depicted in Figure 4. Trial conditions differed with regard to both founder inset distance and offspring and pollen dispersal distances as indicated (see summaries for A through F to the right of the graph). Connected data points show alleles retained under one set of trial dispersal conditions when founders were placed at diagonally increasing distances from the preserve corner (x-axis). See text for more details.

Observed heterozygosity values for all trials after 100 mating bouts did not vary by a large amount (ranging from 0.967 to 0.980), showing only a 1.3% difference between the highest and lowest values. At generation 100 there was little population subdivision and inbreeding seen across trials with differing founder inset distances and various dispersal conditions (Figure 6). The highest F values were seen for populations under basic dispersal conditions (trial conditions A), followed by populations with offspring being dispersed less distantly (trial conditions B and C). Populations tended to have slightly greater F values at inset distances beyond 100 grid points, except for trial D, where offspring were more distantly dispersed (80% to the last frame) and trial F, where pollen was more distantly dispersed and offspring least distantly dispersed (80% from the last frame and 90% to the first frame, respectively).

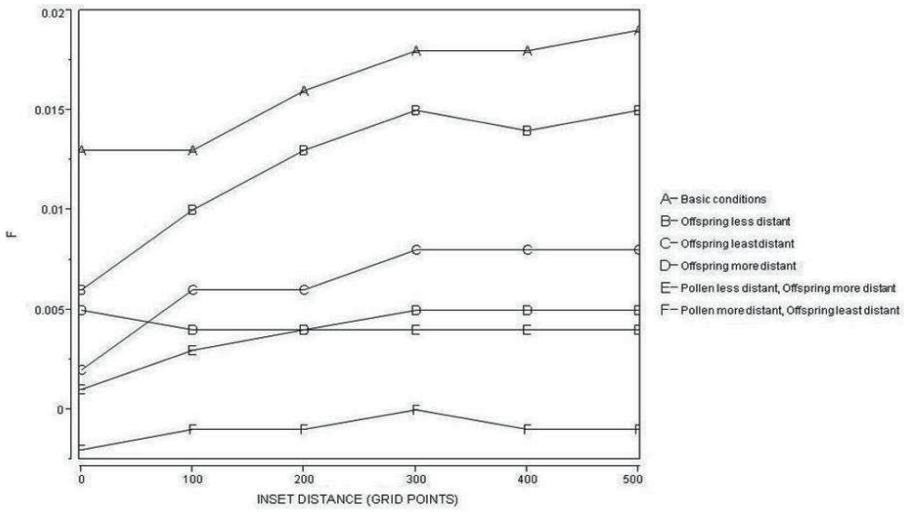


Fig. 6. Changes in F values in populations after 100 bouts of mating (the same populations shown in Figures 4 and 5). Trial conditions differed with regard to both founder inset distance and offspring and pollen dispersal distances as indicated (see summaries for A through F to the right of the graph). Connected data points show F values under one set of trial dispersal conditions when founders were placed at diagonally increasing distances from the preserve corner (x-axis). See text for more details.

4. Discussion

4.1 Practical applications

The null hypothesis that placing founders at differing distances from the edge of a restoration preserve will have no effect on subsequent population growth or genetic diversity was falsified according to results involving the first series trials with the basic dispersal conditions only (Figures 1 through 3). These results confirm that populations can manifest different degrees of population growth and genetic edge effects (Rogstad & Pelikan, 2011) depending on the distance of founders from preserve borders. In terms of population development, increases in growth rates were correlated with increasing distances of founder placement from a border up to a point: with the "basic" gene dispersal distances, planting founders 300 grid points (1500 m) into the preserve is as good as planting them 500 grid points (2500 m) into a 5 km x 5 km or larger preserve. Increasing founder placement up to 300 grid points increasingly allows founders a greater number of viable grid points to establish their offspring thereby reducing mortality levels. As for genetic diversity, the greatest loss of alleles occurred when founders were placed less than 100 grid units from the border (trial a, Figure 2). However, even at the corner, losses were not considerably greater than placing the founders 100 grid points into the preserve (corner trial a retained 90.4% while trial b retained 93.9% of the founding alleles). A preserve manager not interested in maximizing population growth but who only wants to maximize genetic diversity retention could thus plant the founders 100 grid points (500 m) into a 5 km x 5 km or larger preserve. Beyond that, planting further into the preserve does not significantly increase the number of founding alleles conserved. Differences in losses of

genetic diversity as indicated by changes in observed heterozygosity or F values were not pronounced at age 100 among these trials, and thus are not of major concern to a restoration manager through the initial establishment period modeled.

The second series of trials in which offspring and pollen movement distances were varied in populations with founders placed at increasing distances from the border (Figures 4 through 6) demonstrate that population and genetic edge effects are further affected by variations in gene dispersal distances. Compared to the basic conditions used in Figures 1 through 3, population growth was increasingly reduced (Figure 4) when pollen dispersal was more restricted and offspring dispersal was more distant (trial E), when offspring dispersal was less distant but pollen dispersal matched the basic conditions (trial B), when pollen dispersal was more distant and offspring dispersal less distant (trial F), and when offspring dispersal was most restricted but pollen dispersal matched the basic conditions (trial C). Only when offspring establishment was more distant but pollen dispersal matched the basic conditions (trial D) did population growth exceed that exhibited when the basic dispersal condition applied (trial A). These results suggest that in restoration projects with conditions similar to those simulated here, large gains in population growth might be promoted by ensuring that more offspring establish at greater distances than are occurring naturally.

Differences in founding allele retention among this second series of trials were also evident (Figure 5), with the greatest retention in alleles being under the basic and offspring more distant dispersal conditions (trials A and D respectively). For trial conditions A, B, C, and F, there was a sharp increase in allele retention by moving founders from the corner inwards 100 m, after which further gains were not as pronounced to various degrees for different trial conditions at increasing distances. Under trial conditions E and D, placing founders at 300 grid points would be preferable since significant gains in allele retention were not had beyond that distance. Across all of these trials, the greatest difference in allele retention was 9.6% (between trial E founders at the corner and trial D founders inset by 400 grid points). Inbreeding and subdivision appear to attain the highest values under the basic dispersal conditions with founders inset into the preserve (Figure 6), although none of the values are yet approaching pronounced levels. Causes driving the minor differences seen in F values among trials are not always readily interpretable (e.g., note trials E versus F versus C).

Overall, the relationships among trial conditions, inset distance, population growth, genetic diversity retention, and inbreeding/subdivision can be complex and are not necessarily intuitive. Given this complexity, NEWGARDEN modeling can be used to suggest restoration management strategies. As noted earlier, restoration programs and management need to be as cost-efficient as possible. By understanding interactions between life-history characteristics of *C. dentata* and inset distance for founder establishment, program managers can estimate the best methods to minimize such costs. For example, results of this study suggest that planting and stewarding a limited number of founders (thousands at one location are not needed) at least 1500 m into a preserve and promoting successful offspring establishment beyond that which is occurring naturally, rather than expending that effort promoting establishment within the developing stand, is less costly and potentially more successful in terms of ease of population establishment and growth, maintenance, as well as retention of genetic diversity and avoidance of inbreeding and subdivision compared to other options. Managers may have to make cost-benefit decisions such as which is more important: saving funds by planting near a border, maintaining genetic diversity at the

sacrifice of maximizing population growth, or planting further into a preserve at greater expense. The results here stress the need for accurate knowledge of realized gene dispersal attributes (and of other life history attributes) in any modeling-derived restoration management planning. Lacking such information, restoration managers should preferably take the most conservative approach to restoration using as accurate estimations of life history characteristics as possible, in this case, planting the founders at least 1500 m from a border.

4.2 Evolutionary implications

Natural populations establish with varying numbers of founders (often low numbers), degrees of isolation, and geometry of founding. Interpopulation and interspecific life history variation compound the complexity of establishment events as indicated in trials here. The results above suggest that these factors can interact differentially among newly founded populations, and that even seemingly slight differences in initial conditions (e.g., just 500 m difference in founder placement) may have significant effects on the future trajectories of population growth and genetic diversity measures. NEWGARDEN can be used to explore this variation in theoretical and existing situations.

4.3 Further considerations

We emphasize that the results reported here were generated with a subset of the potential input conditions including constant input parameter values designed to simulate population development in the only known restoration population of chestnuts, the West Salem stand (Rogstad & Pelikan, 2011). However, conditions at other restoration sites will surely differ (e.g., density of competitors and rates-distances of establishment, age-specific offspring and pollen production, age-specific mortality rates, gene dispersal distances, etc.). Further, other species will often have drastically different conditions from those modeled here (e.g., short lived perennials or annuals, different densities, different gene dispersal patterns). Finally, the ecological niche will vary across species in ways that might affect geometric patterning (e.g., the section of a forest the species naturally inhabits: one species that would typically be found living on the outer regions of a forest would have different life history characteristics than a species that would inhabit the center part of that forest). NEWGARDEN can be used to explore such intra- and interspecific differences in the conditions of population establishment and development.

Previous studies (Rogstad & Pelikan, 2011) have indicated that differences in the geometric placement of founders (e.g., spacing between individuals; arranged in lines versus squares; founder subdivision into a series of smaller squares) can affect rates of population growth and genetic diversity maintenance. When planting a founding population, the simplest pattern is to plant trees in straight lines to make rectangular or square founder areas. In our experiments, the founders were established in a 13 x 13 square of individuals to be easily manipulated for each trial of varying inset distances. This square shape only allowed for one offspring to establish within the founding square between founding parents, which may not allow for higher rates of population growth in the first few years of reproduction if the offspring have limited dispersal. One question not addressed in our trials was whether or not varying geometry of founder establishment patterns (e.g., completely random founder locations, X-shaped lines, various straight lines, circles, or hollow squares) would affect

population growth and genetic diversity. Different establishment patterns may allow for a greater number of offspring to land in unoccupied grid points than in the square pattern used in the above trials, and possibly reduce the loss of genetic diversity. It is also possible that the geometric pattern of the founder establishment interacting with the inset distance would affect the overall population growth and maintenance of genetic diversity. Trial comparisons where geometry is varied in combination with distance from edges and dispersal distances are underway.

For the above trials, 169 founders were used based on previous studies (Lawrence *et al*., 1995; Chakraborty, 1993) that suggest a minimum of approximately 172 founders is needed to capture the greatest majority of alleles from the source population while also minimizing reintroduction costs in restoration projects. NEWGARDEN could be used to further investigate what would be the minimal number of founders needed to minimize allele loss over generations under a range of conditions.

Since it is not possible to investigate differences in all input conditions *a priori*, NEWGARDEN modeling can also be used *a posteriori* to generate simulated populations reflecting conditions and population growth outcomes for stands that have already been established. Such simulations can then be used to evaluate the degree to which population growth and genetic diversity preservation could be improved by supplementations or manipulations of actual populations (e.g., planting more individuals, altering gene exchange distances via increased seed distribution, reducing mortality rates at certain life stages, etc.).

If the restoration program includes harvesting offspring to redistribute within the preserve or at new locations, where should the offspring be collected from? Samples could be taken in the immediate area surrounding the entire founding population. However, this might not be optimal due to travel distance into the preserve. It would be better to collect offspring closer to one edge of the preserve, minimizing travel distance. However, perhaps the farther in one direction the offspring are from the founders, the greater the risk of loss of genetic diversity and increased offspring subdivision. One area for further study would be to analyze cohort (not population) growth, heterozygosity and F values, and loss of genetic diversity in subregions that are at varying distances from the founders. This would allow restoration managers to determine how close to an edge they could collect offspring that would still retain the greatest genetic diversity for new plantings. The results of such experiments would also provide information on the erosion of genetic diversity over space and time in entire populations and the degree to which erosion might be localized.

5. Acknowledgements

We thank the Departments of Biological Sciences, Environmental Studies, and Mathematical Sciences at the University of Cincinnati. We are grateful for funding from the Women In Science and Engineering Program, also at the University of Cincinnai. Specifically, we thank C. Daley, U. Ghia, and anonymous reviewers for their contributions.

6. References

Barakat, A., DiLoreto, D., Zhang, Y., Smith, C., Baier, K., Powell, W., Wheeler, N., Sederoff, R., & Carlson, J. (2009). Comparison of the transcriptomes of American chestnut

(*Castanea dentata)* and Chinese chestnut (*Castanea mollisima)* in response to the chestnut blight infection. *BMC Plant Biology,* Vol. 2009, No. 9, (May 2009), 513-958, 1471-2229

Chakraborty, R. (1993). A class of population genetic questions formulated as the generalized occupancy problem. *Genetics*, Vol. 134, No. 3, (July 1993), 953-958,1061- 4036

Diskin, M., Steiner, C., & Hebard, F. (2006). Recovery of American chestnut characteristics following hybridization and backcross breeding to restore blight-ravaged *Castanea dentata. Forest Ecology and Management,* Vol. 223, No. 1-3, (March 2006), 439-447, 0378-1127

Elliot, K., & Swank, T. (2008). Long-term changes in forest composition and diversity following early logging (1919-1923) and the decline of the American chestnut (*Castanea dentata). Plant Ecology,* Vol. 197, No. 2, (September 2007), 155-172, 1573-5052

Huang, H., Dane, F., & Kubisiak, T. (1998). Allozyme and RAPD analysis of the genetic diversity and geographic variation in wild populations of the American chestnut (Fagaceae). *American Journal of Botany*, Vol. 85, No. 7, (July 1998), 1013-1021, 1537-2197

Jacobs, D. (2007). Towards development of silvical strategies for forest restoration of American chestnut (*Castanea dentata)* using blight resistant hybrids. *Biological Conservation,* Vol. 137, No. 4, (July 2007), 497-506, 0006-3207

Jansen, P., Bongers, F., & Van der Meer, P. (2008). Is farther seed dispersal better? Spatial patterns of offspring mortality in three rainforest tree species with different dispersal abilities. *Ecography*, Vol. 31, No. 1, (February 2008), 43-52, 09067590

Lawrence, M., Marshall, D., Davies, P. (1995). Genetics of genetic conservation. I. Sample size when collecting germplasm. *Euphytica*, Vol. 84, No. 2, (June 1995), 89-99, 00142336

Merritt, D., & Dixon, K. (2011). Restoration Seed Banks – A Matter of scale. *Science*, Vol. 332, No. 6028, (April 2011), 424-425, 1095-9203

Nathan, R., Schurr, F., Spiegel, O., Steinitz, O., Trakhtenbort, A., & Tsoar, A. (2008). Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution,* Vol. 23, No. 11, (November 2008), 638-647, 0169 -5347.

Paillet, F. (2002). Chestnut: history and ecology of a transformed species. *Journal of Biogeography,* Vol. 29, No. 10-11, (2002), 1517-1530, 1365-2699

Pierson, S., Keiffer, C., McCarthy, B., & Rogstad, S. (2007). Limited reintroduction does not always lead to rapid loss of genetic diversity: An example from the American chestnut (*Castanea dentata:* Fagaceae). *Restoration Ecology,* Vol. 15, No. 3, (September 2007), 420-429, 1526-100X

Primack, R. (May 1, 2010). *Essentials of Conservation Biology* (5th edition), Sinauer Associates, 978-0-87893-640-3, Sunderland MA

Primack, R., & Miao, S. (1992). Dispersal can limit local plant distribution.*Conservation Biology,* Vol. 6, No. 4. (December 1992), 513–519, 0888-8892

Rogstad, S., & Pelikan, S., (September 2, 2011). *Genetic Diversity in Establishing Plant Populations: Founder Number and Geometry* (1st edition), Science Publishers, 157808721X, New Hampshire.

Russel, E. (1987). Pre-blight distribution of *Castanea dentata* (Marsh.) Borkh. Bulletin *of the Torrey Botanical Club,* Vol. 114, No. 2, (April 1987), 183-190, 0040-9618

Stilwell, K., Wilbur, H., Werth, C., & Taylor, D. (2003). Heterozygote advantage in the American Chestnut, *Castanea dentata* (Fagaceae). *American Journal of Botany*, Vol. 90, No. 2, (February 2003), 207-213, 1537-2197

20

Genetic Structure and Diversity of Brazilian Tree Species from Forest Fragments and Riparian Woods

Danielle Cristina Gregorio da Silva et al.*
Universidade Estadual do Norte do Paraná
Brazil

1. Introduction

Historical patterns of human occupation in Brazilian Neotropical Region, featured by deforestation for urbanization, economic exploitation and agriculture, have changed the Atlantic Forest landscape to a collection of fragments. Nowadays this biome is characterized for being highly fragmented but still possessing one of the highest rates of species diversity in the world.

Understanding the genetic structure of populations that occur in forest remnants is fundamentally necessary to establish efficient strategies for the re-composition, management, and conservation programs. For such, it is necessary not only to understand the genetic diversity of a species, but also, how this diversity is distributed within and between forest populations. Notably, a considerable part, if not the majority, of Brazilian Atlantic Forest fragments are linked to rivers or streams, once the policy applied in Brazil regarding conservation in agricultural areas favours the maintenance of legal reserves in proximity of water sources. The vegetation of river margins are subjected to flooding, a strong limiting factor which can lead to local adaptation. These ecological and landscape characteristics may have important outcomes to the genetic diversity of tree populations.

In this chapter, we assembled information from review and research papers of impact in this study area intending to raise knowledge to assist conservation initiatives of Brazilian Atlantic Forest fragments and riparian woods. We plan to broach fundamental concepts regarding the historical fragmentation process in Brazilian Neotropical Region, the effects of fragmentation upon the genetic diversity of forest remnants, and the local adaptation to seasonally variable river levels. This discussion is not intended to be a summary of the existing literature in the theme, but to address important information concerning genetic diversity of neotropical tree species, focusing in the results of eleven years of research on species frequently used in reforestation of legal reserves in Southern Brazil.

* Mayra Costa da Cruz Gallo de Carvalho[1], Cristiano Medri[2], Moacyr Eurípedes Medri[3], Claudete de Fátima Ruas[3], Eduardo Augusto Ruas[3], Paulo Maurício Ruas[3]
[1]*Empresa Brasileira de Pesquisa Agropecuária, Brazil*
[2]*Universidade Estadual do Norte do Paraná, Brazil*
[3]*Universidade Estadual de Londrina, Brazil*

2. The Neotropical floristic kingdom

The word *Neotropic* (from the Greek *neos* = "new") refers to the tropical region of the American continent (Antonelli & Sanmartín, 2011), or "New World" – a term coined by Peter Martyr d' Anghiera in 1493, shortly after Christopher Columbus's first voyage to the Americas (O'Gorman, 1972). As currently defined (Schultz, 2005), the Neotropical kingdom extends from central Mexico, in the north, to southern Brazil, in the south, including Central America, the Caribbean islands and most of South America.

In the Neotropics, equatorial and tropical climates predominate with low climatic seasonality, when compared with kingdoms with cold and temperate climates. Precipitation and annual mean temperatures are generally high, but there is great regional variation (Antonelli & Sanmartín, 2011). Before human colonization, the rain forest of Amazonia accounted for about one third of the entire South American continent. There are, however, several other terrestrial biomes in the Neotropics that are noteworthy for their size and ecological importance, such as the Cerrado and the Atlantic forest of eastern Brazil (Antonelli & Sanmartín, 2011).

The outstanding species richness found today in the Neotropics has remained elusive in our understanding of the evolution of life on Earth (Antonelli & Sanmartín, 2011). Comprising around 90,000–110,000 species of seed plants, the Neotropics alone harbours about 37% of the world's species, more than tropical Africa (30,000–35,000 spp.) and tropical Asia and Oceania combined (40,000–82,000 spp.) (Govaerts, 2001; Thomas, 1999). Sanmartín and Antonelli punctuated the factors that can explain this high species richness (Antonelli & Sanmartín 2011 and references therein). They are related 1) with the geographical position of Central and South America, resulting largely in tropical and equatorial climates; 2) edaphic heterogeneity; 3) biotic interactions that promote speciation mechanisms; 4) relatively stable environments over time, resulting in very ancient ecosystems that conserve niches; 5) adaptive radiation favoured by the great ability of dispersal of flowering plants; 6) geographic isolation (only 3.0 million years ago, South America became connected to North and Central America by the Isthmus of Panama); 7) climatic fluctuations of Pleistocene, leading to formation of refuges for many isolated populations and allopatric speciation; 8) uplift of the Andean cordillera, occurred largely in the last 25 million years; and 9) the profound change in the vast hydrological systems, especially in the Amazon region.

During millions of years there was a synergism between gradual and slow climate changes and speciation, giving time for natural selection and other evolutionary tolls to play their role. The final result of this long process of genesis and evolution of the Neotropical region, is that most of the Neotropical countries, such as Brazil, Colombia, Ecuador and Peru, are on the higher positions of any ranking of species richness. Brazil, in particular, occupies the first position in such rankings and is therefore considered the most Megadiverse country.

Neotropical kingdom is divided into Caribbean, Guayana Highlands, Amazonian, Brazilian and Andean region (Takhtajan, 1986). Brazilian region is, in turn, subdivided into the provinces of the Caatinga, the highlands of central Brazil, Chaco, Atlantic, and Paraná (Takhtajan, 1986). In the Brazilian territory, major biomes are spread over these provinces, such as the Caatinga, Cerrado, Atlantic Forest, Pantanal and Pampas. Even in the case of a predominantly tropical region, where, typically, there is less thermal seasonality and spatial climate variations than in temperate regions, there are important environmental variations

that create significant intra-regional heterogeneity. This fact can be explained by: 1) large latitudinal variation found, the northern boundary of the region is located around 3 ° S, and the south boundary, around 33 ° S latitude; 2) great altitudinal variation, with elevations ranging from sea level to mountainous regions that can reach approximately 2900 meters of altitude; 3) wide edaphic variation, the result of different soil genesis processes over time and space and 4) continentality, which determines a lower overall level of humidity and higher thermic amplitude, the most distant of the Atlantic Ocean. This great heterogeneity, combined with the inherent environmental characteristics of tropical regions that favor the development of life, and added to the climate and geological history of the entire Neotropic, which favored isolation and speciation, have provided to these biomes, in terms of floristic and physiognomy, a high species richness, high endemism and great structural complexity.

According to the Brazilian Institute of Geography and Statistics - IBGE (2004), Atlantic Forest biome constitutes the extra-Amazonian forest large set of South America, formed by rainforests (dense, open and mixed) and seasonal (deciduous and semideciduous). It comprises an environmental complex that includes mountain ranges, plateaus, valleys and plains of the entire eastern Brazilian Atlantic continental band. In southern and southeastern Brazil, it expands westward, reaching the borders of Paraguay and Argentina, also advancing on the southern highlands of Brazil, reaching the state of Rio Grande do Sul.

3. The history of fragmentation of the Atlantic forest biome

The Atlantic forest is the most uncharacterized Brazilian biome (IBGE, 2004). Since the beginning of European colonization, from 1500, several economic cycles of exploration occurred, generating successive impacts in a growing area (Dean, 1996). Brazil has the most diverse flora of the planet and also had its name inspired in a tree, Brazil-wood (*Caesalpinia echinata* Lam), a typical species of coastal forests of the southeast. The first cycle of the ancient Portuguese colony was the exploitation of this species, highly valued for producing a resin which conferred a reddish color to fabrics. Later, with the largest reserves of this wood already exhausted, other exploitation or agricultural cycles began, such as the sugar cane, gold, and ultimately, coffee. They all contributed strongly to clearance and degradation of new areas (Dean, 1996). From the late 19th century and throughout the 20th century, new development and national integration projects have come and settled a consistent process of industrialization and urbanization exactly in the area originally occupied by the Atlantic Forest. Nowadays, these urban areas have the highest population densities and lead the economic activities in the country (IBGE, 2004). An estimated 112 million people live in this area, which accounts for 61% of the population of Brazil (IBGE, 2007). The current results are the almost complete loss of primary forests and a continuous process of destruction of existing remnants, which place the Atlantic Forest biome in unworthy position in the world: as one of the most endangered ecosystems (SOS Mata Atlantica and Instituto Nacional de Pesquisas Espaciais (INPE), 2011). In sum, predatory economic cycles and projects of development and national integration have led, throughout the biome area, to the expansion of the agricultural frontier, the establishment of industrial activities, mining and power generation and intense and disorganized urbanization process, causing the destruction of, approximately, 92% of the original vegetation, which had, in 1500, 1,315,460 square kilometers (Fundação SOS Mata Atlântica and INPE, 2011). The remaining 8% are highly fragmented, separated by a matrix that includes pastures, crops,

water reservoirs, industrial plants, mining and urban areas, especially on the margins of water bodies and mountain areas with highly tilted ground. Even so, one should not underestimate the importance of these fragments. For example, in terms of species richness, occur, even today, in the Atlantic Forest biome, about 20.000 species of vascular plants, of which 6.000 are endemic (Fundação SOS Mata Atlântica and INPE, 2011).

If on one hand the historical processes of economic development led to a predatory pattern of destruction with the formation of relictual fragments of sizes and levels of isolation and different anthropic impact, on the other, the Brazilian environmental legislation, considered one of the most advanced of the planet, through his last Forest Code, established in 1965 and improved in recent decades, defined areas of permanent preservation and legal reserve (Medeiros et al., 2004). The first has the function of conservation of ecosystem services, encompassing riparian forests, river headwaters, hills tops and mountains, hillsides with slopes greater than 45 °, "restingas" and mangroves. The second has the goal of biodiversity conservation and must have, at least, 20% of the property area in the southern and southeastern Brazil, reaching 80% of the area in the Amazon region. Thus, the historical process that led to the intense fragmentation of the Atlantic Forest biome has suffered, in recent decades, the influence of an environmental legislation more effective for conservation. The result is that most of the remaining fragments can be found in permanent preservation areas and legal reserves, especially in riparian areas and steep slopes with unstable soils.

At this time, Brazilian society is burned with a heated debate, in the Brazilian National Congress, on the proposed changes to the Forest Code. In one side, the sectors of society and policymakers linked to the economically strong Brazilian agricultural sector require the flexibility of code, aiming to reduce the need for recovery and conservation of areas of permanent preservation and legal reserve, in order to result in the release of more land for agriculture. In the other side, the Brazilian Society for Science Progress, urban sectors of society and policymakers linked to the environmental movement strongly oppose the proposed changes. There is a concern that, if these changes are approved, the fragmentation process of natural areas and the destruction of relictual fragments will be intensified, which would result in ecological and evolutionary consequences for populations of many species present in these natural areas.

4. Genetic diversity of tree species

4.1 Fundamental concepts

The genetic variability contained in plant species may occur in distinct levels: 1) species within ecosystems; 2) populations within species; and 3) individuals within populations of a species. The genetic structure of a species can be defined as the distribution of the genetic variability within and between its populations as a direct result of the combination among mutation, migration, selection and genetic drift. Also, many tree species harbor effective mechanisms that allow the dispersion of alleles, enabling them to maintain high levels of genetic variability within their populations (Hamrick et al., 1979; Hamrick, 1983; Loveless and Hamrick, 1987). Studying several natural tropical tree populations, Hamrick (1983), concluded that the genetic variability within natural populations is directly linked to their mating system, pollen dispersion syndrome, seed propagation and also by their effective

population size. These factors are also related to the geographic distribution and the type of community in which such species naturally occur.

As stated above, understanding the genetic structure of populations that occur in forest remnants is fundamentally necessary to establish adequate criteria in to which these populations will play a role in the re-composition of degraded ecosystems (Kageyama, 1987). For such, it is necessary not only to understand the genetic diversity of a species, but also, how this variability is distributed within and between their populations present in the disturbed areas. To such intention many statistical tools have since been developed to measure, qualify and partition this genetic variability.

In 1951, Sewal Wright established one of the main components of the distribution of genetic diversity in natural populations: the partition of the endogamy coefficient within and among populations. In his method Wright was able to determine how the endogamy coefficient determine not only the level of crossing between close related individuals within a population but also how it can be related to the differentiation of multiple populations and the overall adaptability of a metapopulation. Also he was able to demonstrate how this partition could be directly linked to the matting system present within each species. His method partitioned the components of the endogamy coefficient f into three distinct coefficients: $F_{IS,}$ which is mainly used to measure the degree of crossing between closely related individuals within a population; $F_{ST,}$ which can be considered an estimative of the endogamy level among populations. Although such concept might be strange, Wright was able to determine that this endogamy between populations was equivalent to determining the genetic relatedness among these populations; and F_{IT}, which represent the endogamy level present in the whole metapopulation and correspond to the overall adaptability measured for the sum of populations. We can also consider that F_{IS} represents the endogamy level related to the reproductive system present within a species, F_{ST} to be the endogamy level due to the partition of the population into subpopulations, and F_{IT} the endogamy level related to the reproductive system and the subdivision of all subpopulations.

This concept of partitioning the f statistic developed by Wright is highly important when we consider the genetic study of the fragmented population as it gives an overview of how the genetic variability is distributed within and among the subpopulations of a species in a determined area and permits us not only to infer the level of fragmentation within a metapopulation but also main type of reproduction present in a species.

When considering the genetic pattern of natural populations, we need also to know specific patterns of genetic richness within each population to answer a wide range of questions like: degree of conservation, percentage of variation within each specific population, differences in genetic diversity and degree of heterozygosity. This being said, we have to consider other important measures of population genetics.

The percentage of polymorphic loci is one of the simplest measurements of the genetic variation that can be used to evaluate the genetic variability present in a population. As the name says, it shows the number of polymorphic loci present in a population in relation to all the amplified loci obtained with the DNA-based Markers.

Another form to evaluate the genetic variability present within a population is to obtain information about its genetic diversity. Initially geneticist borrowed the method of

calculating such diversity from our friends Ecologists. The Shannon and Weaver (1949) index (*H*) was employed in ecology to measure the diversity of species within a given area and later was adapted to measure the genetic diversity within the studied populations. Later on, a more specific index was developed to measure this genetic diversity eliminating some of the bias that was generated when adapting the Shannon index for the evaluation of the genetic data. The Nei´s genetic diversity (1973) was developed as a specific way to measure the population genetic diversity using data obtained specifically with the DNA-based markers. This index was able to measure more accurately the degree of genetic variation within each population and presented different considerations when analyzing the data obtained by dominant (there is no way to differentiate the recessive alleles) codominant (all alleles are differentiated, and each pair of primers is considered to amplify only one molecular locus) markers. For dominant (*H*) data Nei´s genetic diversity is analyzed in terms of within population gene diversity (*Hs*) and the total gene diversity present in the pool of populations studied (*Ht*). But as said above, for codominant marker more detailed information can be given by this statistical index, dividing this statistic into observed (*Ho*) and expected (*He*) heterozygosities, making it possible to calculate the excess or deficit of heterozygotes within each population, gene flow and inferences of genetic bottlenecks and genetic drift.

Considering that in the traditional method of calculation the genetic variability was based on the assumption that the populations were in Hardy-Weinberg equilibrium, some of the inferences obtained for this population presented a significant bias. As a good example we can consider the inference obtained with the Nei´s statistic for genetic diversity, the observed heterozygosity (*Ho*) was compared to the expected heterozygosity (*He*) which was obtained as the pattern of distribution of all the alleles amplified if the population was in HW equilibrium. To eliminate this bias the Baeysian statistic method was adapted to the analysis of molecular data, promoting a revolution in the parameters that can be calculated using data obtained by the DNA-based markers. The Bayesian method is characterized by the use of the posterior probability to infer the likelihood of occurrence of a particular event. In this method all the assumptions, like HW equilibrium, are discarded and the obtained results with this method are compared with a chain (algorithms - Markov Chain of Monte Carlo MCMC and Metropolis Hastings) of results that come from the analyses of the same data. This method of chained analyses repeats itself in tandem until the obtained "chain" stabilizes and yields a result that come close to the real pattern of genetic variation present in natural populations. With this method some parameters like population bottlenecks, attribution of genotypes to specific populations, pattern of gene flow and migration, and attribution of individual paternity, allowing inferences on the population genetics based on DNA-based markers more complete and trustworthy.

4.2 Consequences of habitat fragmentation to the genetic diversity of tree species

Habitat fragmentation is one of the most important and well diffused consequences to the anthropic soil use dynamics (Brooks et al. 2002). It is characterized by the rupture of landscape unity that initially presented continuity, generating smaller parcels with different dynamics from the original habitat. Such parcels become disconnected from the original biological processes that occurred throughout the area (Dias et al., 2000), behaving like isolated "islands of biodiversity" surrounded by non forest areas (Debinsk and Holt, 2000).

For trees, degradation of primary habitat results from two main processes, fragmentation of forest into patches following clearance, and disturbance of habitat following extraction processes, such as selective logging. Tropical trees are thought to be particularly vulnerable to the effects of habitat degradation due to their demographic and reproductive characteristics (Lowe et al. 2005). Estimations made more than 10 years ago predicted that within fifty years, approximately, twenty-five percent of the vascular plant species would be extinct (Kala, 2000). This loss is still an ongoing process, which is not only linked to the loss of the number of individual plants of a species, but also to the loss of condition of the habitat in which they initially inhabited, as some species cannot persist in small fragments due to alteration in microclimatic conditions and to the intensification of the border effect in small fragments (Lovejoy, 1983).

Tropical trees are predominantly outcrossed, present extensive genetic flow and keep high levels of genetic variability. They frequently experience low density as a consequence of habitat fragmentation, are highly dependent upon animal pollination and present mixed mating systems, so they generally exhibit more genetic diversity among populations than temperate species (Dick et al. 2008).

For neotropical forest species, a reduction in habitat patch size or population density is usually equivalent to a reduction in population size (Lowe et al. 2005), or a genetic bottleneck. The genetic bottleneck leads to a very well known phenomenon in the evolutionary context of habitat disturbance, the genetic drift. Genetic drift can be defined as the sum of random changes in the frequencies of alleles within a population (Futuyma, 2005). This process has as principal outcome the decrease in genetic variability, which can be detected as a drop in the proportion of polymorphic loci and in the number of alleles per locus.

The reduced effective size of populations is also accompanied by the increase in endogamy levels. This can be a result of both the decline in pollen vectors and consequent raise of selfing in species with mixed mating systems, and the increased probability of crossing among relatives, given their close distribution and reduced potential mates. This can lead to a declined heterozigosity, augmented homozigosity and consequent fixation of alleles, independently of their effects over fitness.

The consequences of genetic drift and endogamy are maximized by the isolation of the remaining fragments, resulting in an increased genetic diversity among them. The loss of genetic variability can affect population viability and limit evolutionary opportunities to the populations: they are expected to suffer increased disease and pest susceptibility, loss of incompatibility alleles, fixation of deleterious alleles and decline in fitness (Young et al, 1996).

Lowe and coworkers, based in a computer simulation (Lowe et al. 2005), found that with even relatively low levels of gene flow between remnant populations, loss of diversity can be significantly mitigated by increasing effective population size. Changes in genetic diversity and differentiation following a decrease in population size take a number of generations to become apparent, which is not the case for inbreeding coefficient that increases immediately in the first generation following the occurrence of selfing. Long-lived, historically outcrossing species, such as neotropical trees, are expected to harbour a high genetic load, as deleterious recessive alleles will be masked at multiple heterozygous loci.

Also, although deleterious mutations are expected to be purged by selection over time in these species, mildly deleterious alleles can persist despite strong selection.

Some morpho-physiological and life-history traits could confer differences in plants' vulnerability to the effects of fragmentation. For example, species with long generation times will suffer weaker negative effects of fragmentation than the ones with short generation times. The same for species able to reproduce asexually, that will have an extended time between generations.

Many naturally outcrossed tree species frequently present selfing as a clear result of fragmentation, since low density/fragmented populations tend to be more autogamous than high density populations. In the hypothetical scenario of anthropogenic fragmentation ceasing and landscapes remaining as they are today, the effects on genetic diversity of plants will still be much stronger in the future if mating patterns continue shifting towards selfing (Dick et al. 2008, Aguillar et al. 2008).

Aguillar et al. (2008) showed that outcrossing species, such as neotropical trees, suffer greater losses of alleles and polymorphic loci than non-outcrossing species. For self-incompatible species in particular, this may result in the loss of low frequency self-incompatibility alleles (S) leading to mate limitation and further reduction of effective population size.

Animal pollinated outcrossed species are also strongly negatively affected in terms of effective pollination service and seed production by habitat fragmentation. Considering that more than 98% of tropical trees species are animal pollinated (Bawa 1990), these species are exceptionally vulnerable to fragmentation as a consequence of both, ecological and genetic mechanisms. Not only pollination but also seed dispersion in tropical forests is much more dependent in animals than in wind: more than 70% of all tropical tree species are animal dispersed (Howe and Smallwood 1982). These species share multiple agents and are generally understory (Dick et al. 2008).

For the majority of tropical trees, pollen flow transposes enormously the distances of the seeds flow. Pollen dispersion distances, although dependent on small animals in most cases, can be surprisingly as high as 500m and still higher, a few kilometers, in low density/fragmented populations (reviewed in Dick et al. 2008). Even if sufficient pollen reaches an isolated tree to fertilize all potential ovules, a reduction in diversity of the pollen cloud, due to fewer pollen donors, can reduce population fitness by allowing fertilizations from self, related, or maladapted parents. A reduction in seed set predicted from fragmentation may be due to one or a combination of a lack of pollination (e.g. from loss of pollinators), lack of compatible pollination (from increased self-pollination through restricted pollinator movement or a loss of incompatibility alleles owing to reduced population size), or inbreeding (Lowe et al. 2005).

Removal of primary habitat also usually decreases the probability that migrant seeds will find suitable sites for establishment. Under such circumstances, pioneers or invasive weeds will be favoured and will increase in occurrence. If a species' life history profile is characterized by frequent extinction and colonization events, the metapopulation is under threat of extinction if the two forces are not balanced. Even if site colonization does occur, founder bottlenecks can drastically reduce diversity (Lowe et al. 2005 and citations therein).

Species rarity can also determine its susceptibility to genetic erosion. Because common species have comparatively higher levels of genetic variability than naturally rare species, they are expected to lose more diversity due to recent fragmentation processes (Aguillar et al. 2008). Given the ubiquitous nature of anthropogenic habitat fragmentation in today's landscapes, this is important and of interest to conservation biology as they situate common species in potential risk of genetic erosion, which is counterintuitive to current conservation principles that almost exclusively emphasize efforts on rare or threatened species.

As discussed above, habitat fragmentation has the potential to erode genetic diversity of a species, and the magnitude of its effects is related to the state of several life-history traits. Among the several factors, deserve special attention the compatibility system, mating system, pollination vector, seed dispersal vector, vegetative growth capability, rarity, time elapsed in fragmentation conditions, (reviewed in Aguilar et al. 2008), and sucessional stage, as we a going to discuss below.

5. Local adaptation in tree species

An adaptation is a characteristic that enhances the survival or reproduction of organisms that bear it, relative to alternative character states, especially ancestral condition. The only way to an adaptation to evolve is by means of natural selection, so it can be also stated that adaptation is a characteristic that evolved by natural selection (Futuyma, 2005).

Local adaptation is the set of patterns and processes observed across local populations of the same species connected, at least potentially, by dispersal and gene flow. It is generally the case when resident genotypes in a deme have on average a higher relative fitness in their local habitat than genotypes originated from other habitats (Kawecki and Ebert, 2004 and references therein).

Local adaptation can be observed in a continuous population, in which sampling units are arbitrary, but is more commonly observed in fragmented populations, since they are discrete units of perennial populations in well delimited habitat patches (Kawecki and Ebert, 2004). Considering that the most existing neotropical tree populations have been fragmented by land use and urban occupation, local adaptation is a relevant area of study for neotropical tree species. Even though, most studies on local adaptation in plants are only available for herbaceous plants in temperate regions (Leimu and Fischer, 2008).

We aim to gather information on local adaptation in plants that can be used to evaluate its strength and outcomes for neotropical tree populations, considering their present fragmented character. Our understanding of this topic is mainly based on the papers of Kawecki and Ebert (2004), Savolainen and colleagues (2007), and Leimu and Fischer (2008).

The ability to adapt may be compromised in small populations because of reduced genetic diversity, caused by genetic bottlenecks or founder effects, which have as consequences an increase in genetic drift and inbreeding. In addition to reduced genetic variation and genetic drift, local adaptation can also be constrained by variation in natural selection. Temporal environmental variability may involve opposing selection pressures and thus constrain adaptation. In contrast, spatial heterogeneity of the habitats of plant origin favours selection for reduced dispersal and increases habitat fidelity, which may in turn favour the evolution of local adaptation (reviewed in Leimu and Fischer, 2008).

Reproductive traits have a role in determining the extent of local adaptation in natural populations. Gene flow, for example, can hinder local adaptation. This is true because protected polymorphism in a heterogeneous environment may be maintained even if dispersal results in complete mixing of the gene pool. In such a case demes will not differentiate genetically, i.e. there will be no local adaptation. Nevertheless, the existence of a pattern of local adaptation despite gene flow certifies to the strength of natural selection imposed by particular environmental factors (Kawecki and Ebert, 2004). Also, if local adaptation is constrained by lack of genetic variation, dispersal and gene flow between populations can enhance local adaptation by increasing genetic variation within populations and potential to respond to selection (Leimu and Fischer, 2008).

Spatial environmental heterogeneity favours reduced dispersal and habitat fidelity, which make conditions for local adaptation more favourable. It should be noted that environmental heterogeneity favours the evolution of adaptive phenotypic plasticity. In the absence of costs of and constraints on plasticity, a genotype that in each habitat produces the locally optimal phenotype would become fixed in all demes. Adaptive phenotypic plasticity would thus lead to adaptive phenotypic differentiation, but without underlying genetic differentiation. The failure of the metapopulation to evolve such ideal plasticity is thus a pre-requisite for local adaptation (Kawecki and Ebert, 2004 and reference therein).

Shortlived and self-compatible species tend to be more strongly differentiated at a smaller scale than long-lived and outcrossing species, and so the former are expected to show stronger adaptation to local conditions. Therefore, neotropical tree species, which are long-lived, outcrossing and as a group includes several examples of self incompatibility, are expected to have weak local adaptation (Leimu and Fisher 2008).

Leimu and Fischer (2008) conducted the first quantitative review on local adaptation in plants, assembling papers that reported comparisons of the performance of plants from local and foreign populations. Among these studies, local genotypes performed on average better than foreign genotypes at their site of origin. However, divergent selection favoured locally adapted plants only in less than half of the pair-wise site comparisons. This suggested to them that local adaptation is less widespread than commonly believed. In this study, they also found that local adaptation appeared to be independent of some plant life-history traits, the degree of spatial and temporal habitat heterogeneity, and of the geographic distance between study populations, but was strongly affected by population size. This clear role of population size for the evolution of local adaptation raises considerable doubt on the ability of small plant populations to cope with changing environments. Thus, in the context of fragmentation process in the Neotropic small fragments, featured by low genetic variability as a consequence of genetic drift and endogamy, might not be able to respond to different selective pressures of changing environments and develop local adaptation.

6. Brazilian Atlantic forest fragments: Case studies of species from Tibagi River Basin

6.1 The Tibagi project

In this section of the chapter we intend to describe our experience in population genetic studies of neotropical trees remnants in Brazil. All species that we have studied have a high importance value index or IVI (the sum of relative dominance, relative frequency and

relative density) in the Tibagi River basin, Paraná, Brazil and, due to the rapid degradation of ecosystems associated with this river basin, its present occurrence was limited to highly impacted forest remnants. The Tibagi River Basin has a great importance in the economic and social development of one of the Brazilian states of greatest economic impact in the country, however, the degree of landscape devastation that has taken place since the beginning of the last century has been threatening the biodiversity of the ecosystem and even physical and chemical characteristics of its rivers. For these reasons, a group composed of around 51 researchers from the University of Londrina and other research centers, in partnership with COPAT (Consortium for Environmental Protection of the Tibagi River basin) and funded by Klabin S/A and Araucaria Foundation for the Support of Scientific and Technological Development of Paraná, developed the project "Aspects of Fauna and Flora of the Tibagi River basin" or Tibagi Project. The Tibagi Project produced a wealth of valuable information aimed at the recovery and conservation of the river basin as a whole. We limit ourselves here to present the knowledge gained about the influence of two factors on the genetic diversity of populations of neotropical tree species remnants: habitat fragmentation and local adaptation to seasonally flooded river banks.

The Tibagi River basin is composed of 65 direct tributaries and hundreds of sub-tributaries in an area of approximately 25,000 Km^2, covering 54 counties in the Paraná State, Brazil. Its landscape, belonging to the Atlantic Forest biome, presents important climate and soil variations in the north-south axis that allows it to be divided into Upper, Middle, and Lower Tibagi. On the Upper Tibagi, the predominant vegetation is steppe grassy-woody also known as general fields, with patches of Araucaria forest. In the Middle Tibagi, there is a transition zone between rain forest and mixed semideciduous forest with some patches of fields and scrubs. In the Lower Tibagi, before fragmentation, the dominant vegetation was the semi-deciduous forest, which formed a continuum with the dense rain forest of the Brazilian coast. With massive deforestation due to lumber extraction, urbanization and expansion of the agricultural front, only small forest fragments remain in this area adding up to only 2.4% of the original forest cover in the lower Tibagi and 12.7% in the Tibagi River basin as a whole (Ribeiro, 2009).

The riparian forests are plant formations that surround bodies of water and for that reason, are deeply influenced by them. On the other hand, water quality and other physical characteristics of rivers are also heavily influenced by the presence and degree of conservation of the riparian forests. In river banks, where the vegetation is scarce or not present at all, events such as erosions can be up to 30 times more significant than in areas where the riparian vegetation is present. Moreover, it is estimated that 80-90% of sediments and pollutants generated in agricultural areas are filtered by riparian vegetation before reaching the bodies of water (Naiman & Decamps, 1997). From the ecological point of view, these forests work as ecological corridors, linking different ecosystems and thereby promoting gene flow, both by land, and across the river. These are environments in which the diversity is determined by the occurrence of flood events with consequent flooding of the river margins. These environments can be very heterogeneous when flooding events are short, sporadic and with little intensity; however, they can be slightly heterogeneous when these flooding events are constant or very prevalent and intense.

Flooding is the main limiting factor that acts on the riparian forest, since it changes dramatically the physical and chemical characteristics of the soil, which is the main

substrate for plant growth. In flooded soil, the diffusion of gases is severely reduced, with a sharp drop in oxygen levels, reduction of gas exchanges, accumulation of toxins, changes in pH and nutrient availability (Drew 1992). When flooding extends to total or partial submergence of the plants, the negative effects on vegetation are even greater due to reduced availability of light and carbon dioxide (Crawford, 1993). Once flood period is over, the soil becomes compacted and a series of injuries from re-aeration process arises due to the accumulation of reactive oxygen species in plant tissues during the flooding period (Crawford, 1993).

However, most of the riparian plant species survive the flooding stress through the development of morphologic, anatomic and physiological changes (collectively referred to as the low oxygen escape syndrome - LOES; Bailey-Serres & Voesenek, 2008) that re-established, at least in part, minimum energy levels for plant survival. Other plant species survive with avoidance strategies, completing their life cycles between periods of flooding. There are also those species that have better relative growth rates in flooded soil. The ability to tolerate periods of flooding gives plants a competitive advantage over those that do not tolerate this stress, which is evidenced by the example of the tolerant species *Cecropia pachystachya*, *Sebastiana commersoniana* and species of the genus *Inga*, which are most often found in areas affected by periodic flooding than in other plant formations, where inter-specific competition limits its distribution. In general, local adaptation in challenging environments may either favour the reduction of genetic variation through natural selection or lead to the expression of phenotypic plasticity of identical genotypes (Stöcklin et al., 2009).

In this light, it seems clear that the recovery strategies and/or conservation of genetic diversity of tree species of the riparian forests on the Tibagi River basin should consider: (i) knowledge about the history of fragmentation and its influence on the genetic structure of the affected populations, (ii) knowledge about the diversity of mechanisms of tolerance to flooding and their possible genetic determination, (iii) and the influence of local adaptation to flooding events on the distribution of genetic variation within and among populations. In order to answer these questions, many common species from Tibagi river basin were studied, regarding flood tolerance and occurrence of morphological, anatomical and physiological changes in response to it. For eight of these species, we used DNA-based molecular markers (RAPD, AFLP and SSR) to obtain estimates of population genetic parameters such as polymorphic loci, total genetic diversity and genetic differentiation within and among populations. The eight sampled species have high importance value index (IVI) in Tibagi River basin (five are among the ten species with highest IVI by region) and cross-pollination system, but vary in regard to gender distribution and dispersal mechanisms for pollen and seeds (Table 1).

In Tibagi River basin, the analysis of the pluviometrical events from 1932 to 1990 showed that there were ten significant flood events in the region in this period, while the fragmentation of the vegetation was more pronounced between 1920 and 1960 (Medri et al., 2002). Considering that the first study published by the group comprised a minimum of 76 years between the collection of plant material and the first pronounced flooding event registered in the region and 83 years in relation to the beginning of the fragmentation process; in the meantime, the considered species advanced at least three generations (Table 1). It was expected at first that: (i) the genetic diversity of populations of degraded areas

might have been reduced compared to preserved areas, (ii) the pioneer species have been less affected by the effects of fragmentation than the species of late successional stages, (iii) there has been significant genetic differentiation between fragments and (iv) local adaptation to flooding has caused genetic structure, probably by reducing the diversity in populations that regularly experience stress.

Tree Species	Ss[1]	Ra	MS (SD)[2]	Pd	Sd	Long	Mm	Np	Pp(%)	Ht	Fst	Fis	Ref.
A. sellowiana	Pi/Si	2-4	Ou & Vp (Di)	Bi	Bi	25	SSR	5	77.78-100	0.32-0.48	0.197	0.03(-) 0.33	Oliveira, 2010
							RAPD	2	66.67-67.48	0.17-0.18	0.059	-	Medri et al., 2011
							RAPD	9	35.32-52.38	0.09-0.14	0.495	-	Medri et al., 2011[4]
A. polyneuron	St	50	Ou (He)	Bi	Ab	1200	AFLP	2	88.5-99	0.31-0.37	0.265	-	Damasceno et al., 2011
C. xanthocarpa	St	6-8	Ou (He)	Bi	Bi	60	AFLP	2	92.27-92.82	0.32-0.33	0.180	-	Ruas, 2009
C.gonocarpum	St/Cl	-	Ou (He)	Bi	Bi	-	AFLP	2	79.5-92.1	0.28-0.31	0.300	-	
L. divaricata	Pi	2	Ou (He)	Bi	Ab	60	RAPD	2	80.69-86.9	0.9	0.104	-	De Carvalho et al., 2008
M. aquifolium	St/Cl	2	Ou (Un)	-	Bi	60	RAPD	3	72.80-84.51	0.17-0.21	0.218	-	Sahyn et al., 2010
P. rigida	Pi	10	Ou (Mo-Si)	Bi	Ab	100	RAPD	2	91.95	0.94	0.063	-	Silva et al., 2010
							AFLP	8	48.44-82.81	0.17-0.30	0.228	-	Souza, 2011
S. terebinthifolius	Pi	2-4	Ou (Di)	Bi	Bi	10-15	RAPD	2	64.64-70.72	0.168-0.20	0.137	-	Ruas et al., 2011
							SSR	6	87.5-100	0.48-0.62	0.199	0.04(-)0.03	Ruas et al., 2011[4]

[1] Pi: Pioneer; Si: Secundária inicial; St: Secundária tardia; Cl: Clímax.
[2] Ou: Outcrossing; Vp: Vegetatively propagated; Mo: Monoecious; Mo-Si: Monoecious, Self incompatible; Di: Dioecious; He: Hermaphrodite; Un: unknown.
[3] Bi: Biotic; Ab: Abiotic.
[4] In press.

Table 1. Biological features and genetic parameters of the tree species studied in the Tibagi Project initiative. Information on biological traits was obtained in the literature and by personal communications of experts. When more than two populations of a given species were studied, the Fst values are represented as an average of those populations. Ss: Sucessional stages; Ra: Reprodutive age, in years; MS (SD): Mating system (Sex distribution); Pd: Polen dispersal; Sd: Seed dispersal; Long: Longevity, in years; Mm: Molecular marker; Np: Number of populations; Pp: Percentage of polymorphic loci; Ht: heterozigosity or genic diversity, when either codominant or dominant markers are respectively applied; Ref: references.

6.2 Local adaptation of tree species to flooding

In the Amazonian floodplain forests there are some regions where the water column reaches up to 7m with submersion of trees for long periods, and regions where this stress is less intense. Comparing the different areas of flooding, it is possible to observe a significant variation in the occupation pattern of these areas ranging from monospecific to over 150 species ha^{-1} (Wittmann et al., 2007 and references therein). It is clear that the stress intensity and duration of flooding periods determine the diversity of species occupying these areas. In the Tibagi River basin, the most critical flooding period is the three wettest summer months (December-January-February), with areas that may remain flooded for a few days to several months depending on the intensity of the phenomenon. Torezan & Silveira (2002), conducted several phytosociological studies in forest fragments along the Tibagi River Basin

comparing areas under different intensities of occasional floodings. It was observed that the higher intensity, duration and size of the flooded area, the lower species richness was found. In a fragment with 100% of flooded area, there were 42 species/ha, while in fragments located in non-flooded areas, or with an insignificant area subject to flooding, the number of tree species was often greater than 100/ha, reaching up to 127 in one single area. Thus, we consider as a starting point the hypothesis that the stress of flooding, typically observed in the region, should determine the occupancy of the periodically flooded banks, favoring the establishment of species tolerant to this stress.

Several studies with species that are found in the Tibagi River basin were conducted in a greenhouse to answer this hypothesis. Plant species not found in the wetlands and the species most often found in the wetlands have been challenged in artificial flooding treatments with different intensities and durations. Briefly, the results showed that some tree species naturally distributed in this river basin but that do not occur in flooded areas, did not tolerate the stress of artificial flooding; also, there was a great variation in the response of the species from wetlands. In addition to this information, the provenance trials, conducted to compare the performance of populations of the same species from areas periodically flooded and non flooded, showed that for the studied species, *Luehea divaricata* and *Parapiptadenia rigida*, the plants that originated from populations naturally flooded were more efficient in responding to waterlooging stress (De Carvalho et al., 2008) or tolerated higher levels of stress, which was not tolerated by the plants of the other provenance (Silva et al., 2010). In a similar study with the tree species *Aegiphila sellowiana*, Medri et al. (2011) used plants regenerated from seeds collected in four regions that presented occasional flooding, showing a variation in the response of individual tolerance to waterlogging. In this experiment, 46.7% of the plants died, while the remaining individuals developed morpho-anatomical alterations common to tolerant species, supporting up to 80 days flooding (Medri et al., 2011).

When these results were obtained (2000 - 2002) we believed that there should be a genetic background related to flood tolerance, but the variation in responses observed between the tolerant species pointed to the existence of a large number of genes and/or a strong effect of genotype-environment interaction in determining the tolerance. With these in mind, we used Random Amplified Polymorphic DNA (RAPD) to see if there was a reduction in genetic variation in populations tolerant to flooding, and whether it was possible to detect variation between populations that respond differently to the stress. In our experiments of performance comparison, the percentage of polymorphism observed *in situ* among adults of *Luehea divaricata* and *Parapiptadenia rigida* was greater (approximately 6%) in the areas subjected to periodic flooding than in the highest regions, which is never affected by stress. The studied populations of *L. divaricata* and *P. rigida*, showed genetic differentiation of 10.48 % for the first species and of 6.00% for *P. rigida* (while about 90% of the observed variation was attributed to the variation found within populations). These results suggested that perhaps the expected reduction in genetic diversity of riparian communities has been masked by the balance between different evolutionary forces that could be occurring *in situ*. Considering the proximity of the sampled areas and the fact that both are crossbreeding species, there certainly is gene flow between the flooded and non flooded populations, which in turn must ensure the re-establishment of genetic diversity in each generation. Possibly, the lower inter-specific competition and greater light availability on the river banks may represent, for the pioneer species studied, a more important factor than natural

selection caused by flooding, enabling the maintenance of higher levels of diversity in wetlands.

We needed a model that would allow us to access the isolated effect of flooding on the genetic diversity of populations of plants tolerant and intolerant to stress. In this experimental model we use the tree species *A. sellowiana*, which displays a gradation of responses to flooding, ranging from death to survival periods of up to 80 days of stress. When comparing plants of *A. sellowiana* which survived the stress of artificial flooding (tolerant), with plants which died (intolerant), the genetic differentiation around 6% was detected between the two groups and it was not possible to detect important variation in the percentage of polymorphism between both of them (Medri et al., 2011). Given these results, it became clear that the vast phenotypic variation between tolerant and intolerant plants of *A. sellowiana* can be, at least in part, related to the genetic patterns observed when using the DNA-based marker RAPD. A similar situation was found by Bekessy et al. (2003) who used RAPD markers to study genetic variation in populations of the South American tree *Araucaria araucana*. Even though we have yet to acquire the knowledge of which genes determine the stress tolerance of flooding on tree species it is possible that this character is directly influenced by many genes as suggested by Sairam et al. (2008) in their review entitled "Physiology and biochemistry of waterlogging tolerance in plants".

Interestingly, considering the results obtained by the group, especially with the local adaptation experiment, carried out *ex situ* with *A. sellowiana*, it becomes clear that the genetic diversity within and between populations cannot be considered the only factor in the election of sources for the recovery of degraded riparian forests. In short, the RAPD markers used in these studies allowed us to detect consistent results for the three species studied since more than 94% of genetic variation detected can be attributed to variation found within populations, following the pattern usually observed for pioneer species, with preferably cross-fertilization. In the absence of knowledge on the mechanisms of inter-specific diversity of tropical trees on response to flooding, we may conclude, erroneously, based only on estimates of distribution of the genetic diversity, that both the populations that occupy the margins of the flooded rivers and those never flooded could be used as seed sources for the recovery of degraded riparian areas. However, the difference in performance between plants, as measured by morphological parameters, indicate that populations locally adapted to the stress would provide a better material for the recovery of degraded riparian areas.

6.3 Effects of fragmentation in natural populations from Tibagi River Basin

In several years of studies observing the effects of fragmentation and local adaptation on the genetic diversity of the remaining tree populations of the Tibagi River Basin, eight species were evaluated mostly with dominant molecular markers (Table 1). The information gathered here reflect observations for post-fragmented populations, and the number of generations elapsed since the beginning of the fragmentation process is equal to one or two for centenary species, and no more than ten generations for the species with longevity from 20 to 25 years.

As seen among tropical species (Matallana et al., 2005), all species present higher frequencies of cross-fertilization and animals play an important role in pollination and/or seed dispersal. These characteristics can be readily related to the high levels of genetic diversity (or heterozygosity) observed in the sampled populations (Table 1). In addition, pollen and

seed dispersal by animals, often observed in tropical species, should provide high gene flow between nearby populations in a continuous gradient, while favouring the genetic differentiation over long distances (Givinish, 1999 and references there in). In such cases, the observed genetic distance between populations was positively correlated with their geographic distances. This hypothesis could be verified from the combined analysis of two studies conducted with *P. rigida*, where two nearby populations (Silva et al., 2010) and eight populations distributed over a long geographic scale (Souza et al., 2011) were evaluated. For nearby populations, the Fst value obtained was only 6.2%, whilst the average F_{ST} between the eight populations was 22.8%, and the correlation between genetic and geographic distance was positive and significant. Similarly, the observed values of F_{ST} between geographically close and distant populations of *A. sellowiana* ranged from 9.56% to 50.26%, respectively (Medri et al. 2011; Oliveira, 2010). For this species in particular, the genetic differentiation between distant populations may be even more pronounced than the one seen for *P. rigida* due to the occurrence of vegetative propagation.

When the genetic diversity among populations is compared between the studied outcrossing species, one can see that there are significant variations among the observed values of F_{ST} (6% - 30%, Table 1), which is suggested to be related to successional stage that each species occupies. Fst values for the initial or secondary pioneer species vary between 6% - 19% and the values for the late secondary or climax species vary between 21% - 30%. In both cases, genetic differentiation among populations assume moderate to high values, which is, in principle, expected to occur among tropical species. Tropical species are often pollinated by animals or have their seeds dispersed by them, often have mixed breeding system in which selfing rates can be changed depending on the environment and are represented by low-density populations (Dick et al. , 2008). These characteristics, together, make the tropical species more dependent on the quality of the ecosystem where they are inserted then the temperate species, and therefore more susceptible to the effects of fragmentation. Under the effect of fragmentation, tropical species suffer a reduction in the availability of pollinators and/or seed dispersers with the reduction in the number of individuals, experiencing a higher proportion of inbreed crosses and genetic drift, with a resulting increase in genetic differentiation between populations.

It is suggested that the observed tendency of lower values of Fst between the pioneer species than among the secondary is a reflection of the life history of these species. The pioneer species represent the first successional stage of the forest, being able to invade areas not yet occupied, including harsh environments and forest edges. Thus, given the ecological role they play, the pioneers are heliophyte, experience rapid growth, lower inter-specific competition, increased investment in reproduction (r selection), a higher number of generations per unit time and aggregate distribution. Fragmentation, therefore, does not cause too negative effects on the pioneer species. In fact, while fragmentation reduces the number of individuals of the species of later successional stages and completely alters the environment they occupy, the pioneer species can be instead favored by this process. Also, the pioneer species from the fragments of the Tibagi River Basin have advanced several generations (10 -15 generations) since the beginning of the fragmentation process. So, after several cycles of cross-fertilization, gene flow between populations may have indeed come relatively high to overcome the effects of fragmentation, allowing lower levels of genetic differentiation among populations.

Unlike the pioneers, the late secondary and climax species tend to have less aggregated spatial distribution, to be ombrophilous and to occupy more specialized niches. These species spend their energy more in the inter-specific competition than in reproduction (k selection) and have slow growth and far fewer generations per unit time. The species of late successional stages, considered here, represent the first or second generation post-fragmentation and thus are believed to reflect the immediate consequences of the fragmentation process. Interestingly, one of the secondary species studied, *Aspidosperma polyneuron*, was found in a continuous distribution, in a plateau followed by a high declivity (Damasceno et al., 2011), and the other, *Maythenus aquifolium*, was found in fragments separated by up to 30 km (Sayhun et al., 2010); however, the F_{ST} values observed in the two situations were similarly high, suggesting that part of the genetic differences found between the populations of these species also linked to adaptive characteristics (Sayun et al. 2010; Damasceno et al., 2011).

Another important information that we could extract from the obtained results for the sampled populations in these regions of lower, middle and upper Tibagi (Ruas et al. 2011 In press; Ruas et al. 2011; Medri et al. 2011 In press; Oliveira et al, 2011) is that the intense fragmentation towards middle-lower Tibagi has provided a significant increase in inbreeding coefficient (F_{IS}), loss of alleles and reduced genetic diversity compared to populations from the upper Tibagi (Table 1). These factors also influenced the genetic differentiation between populations of the upper Tibagi (where fragmentation is less evident) and the others. The only exception to this result was observed in *A. sellowiana*. When analyzing the species *A. sellowiana* with microsatellite markers we also observed the formation of two distinct groups of populations (Dendrogram using Nei´s genetic distance, 1978), one comprised of populations from the middle Tibagi region and the other pertaining to lower Tibagi (Oliveira, 2011 submitted). Even though *A. sellowiana* is able to perform vegetative reproduction and also its propagules can reach as far as 10m distance from mother tree, we evidenced highly significant negative values of F_{IS} indicating excess of heterozygosity in tree of the five populations. When considering the effect of recent genetic bottlenecks in these populations, the values obtained by the software bottleneck indicated that four of the five populations showed significant values of heterozygosity excess, when considering the Infinite Allele Model, and only one population presented significant levels of heterozygosity excess for all tree models (Infinite Allele Model, Two-phase Model and Stepwise Mutation Model). Migration rates calculated for these populations demonstrated that exchange of 30% of migrants between two populations from the middle Tibagi region. Such results demonstrate that some of these populations are suffering more than others, the impact of fragmentation and also founder effect and gene flow are playing a key role in the shaping of the genetic variability within these populations of *A. sellowiana* promoting a weak balance between the evolutionary forces of genetic drift and migration in this post fragmentation period landscape.

Lately, this research group has been engaged in the development and inter-specific transference of microsatellite primers (SSR) for the species *L. divaricata* (Ruas et al., 2009), *A. sellowiana* (Ruas et al., 2010), *P. rigida*, *A. polyneuron*, *C. xanthocarpa*, *C. gonocarpum* and *A. polyneuron*. The use of SSR markers will allow more accurate estimates about the distribution of genetic variation and the effects of fragmentation and local adaptation between the species of the Tibagi River basin. Such is possible due to the codominant nature of these markers, allowing to access allelic information and thus to estimate the number of

alleles, heterozygosity, inbreeding and gene flow among natural populations. Although much work is yet to be done with these and other species of the Tibagi River basin, the present studies certainly represent a great step for the comprehension of the present availability of genetic resources and of its relation to the life history of species from forests fragments pertaining to the Tibagi River basin.

7. Conclusion

In summary, the knowledge we have accumulated so far on the genetic structure of the studied populations allows us to infer that the large genetic differentiation that has been maintained between natural populations, especially among those of late successional stages, makes it imperative to conserve all forest remnants. One of the strategies that we believe can be effective for the conservation and restoration of these ecosystems is the establishment of green corridors to restore communication of pollinators and dispersers between forest fragments of the Tibagi River basin. Moreover, the development of SSR markers for tropical tree species would first increase our knowledge on those species genome and help the investigation of the genetic determination and evolution of some important adaptations, such as flooding tolerance. Second, it would advance our understanding of the effects of habitat fragmentation over the riparian forests' diversity.

8. Acknowledgment

The authors are grateful to Conselho Nacional de Pesquisa (CNPq), Fundação Araucária, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Universidade Estadual de Londrina (UEL) and Universidade Estadual do Norte do Paraná (UENP) for financial support.

9. References

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., & Lobo J. (2008). Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Molecular Ecology, Vol. 17, No. 24, (December 2008), pp. 5177-5188, ISSN 1365-294X

Antonelli, A. & Sanmartín, I. (2011). Why there are so many plant species in the Neotropics? Taxon, Vol. 60, No.2, (April 2011), pp. 403 – 414, ISSN 0040-0262

Bailey-Serres, J. & Voesenek, L.A.C.J. (2008). Flooding Stress: Acclimations and Genetic Diversity. Annual Review of Plant Biology, Vol. 59, (December 2008), pp. 313-339, ISSN 1543-5008

Bawa, K.S. (1990). Plant-pollinator interactions in tropical rainforests. Annual Review of Ecology, Evolution, and Systematics, Vol. 21, No. 1, (November 1990), pp. 399-422, ISSN 1543-592X

Bekessy, S.A., Ennosb, R.A., Burgmana, M.A., Newtonc, A.C. & Adesd, P.K. (2003). Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. Biological Conservation, Vol. 10, No. 2, (April 2003), pp. 267-275, ISSN 0006-3207

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands A.b., Konstant, W. R., Oldfield, S., Magin, G. & Hilton-Taylor, G. (2002). Habitat loss and extinction in the hotspots of biodiversity. Conservation Biology,16: p. 910–923.

Crawford, R.M.M. & Braendle, R. (1996). Oxygen deprivation stress in a changing enviroment. Journal of Experimental Botany, Vol. 47, No. 295, (February 1996), pp. 145-159, ISSN 1460-2431

Damasceno, J.O., Ruas, E.A., Rodrigues, L.A., Ruas, C.F., Bianchini, E., Pimenta, J.A. & Ruas, P.M. (2011). Genetic differentiation in Aspidosperma polyneuron (Apocynaceae) over a short geographic distance as assessed by AFLP markers. Genetics and Molecular Research, Vol. 10, No.2, (June 2011), pp. 1180-1187, ISSN 1676-5680

Dean, W. (1996). A ferro e Fogo – A História e a devastação da Mata Atlântica brasileira, Companhia das Letras, ISBN 85-7164-590-6, São Paulo, Brazil

Debinski, D. M. & Holt, R. D. (2000). A survey and overview of habitat fragmentation experiments.Conservation Biology 14: p. 342-355.

De Carvalho, M.C.C.G., Silva D.C.G., Ruas, P.M., Medri M.E., Ruas E.A. & Ruas, C.F. (2008). Flooding tolerance and genetic diversity in populations of Luehea divaricata. Biologia Plantarum, Vol.52, No.4, (December 2008), pp. 771-774, ISSN 1573-8264

Dias, A., Latrubesse, E. M., Galinkin, M. (2002). Projeto corredor ecológico Bananal Araguaia, Brasília, 120 p.

Dick, C.W., Hardy, O.J., Jones, F.A., & Petit, R.J. (2008). Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. Tropical Plant Biology, Vol. 1, No. 1, (March 2008), pp. 20-33, ISSN 1935-9764

Drew, M.C. (1992). Soil aeration and plant root metabolism. Soil Science, Vol. 154, No.4, (October 1992), pp. 259-268, ISSN 1538-9243

Fundação SOS Mata Atlântica & Instituto Nacional de Pesquisas Espaciais (INPE). (2011). Atlas dos remanescentes florestais da mata atlântica. São Paulo, Brazil

Futuyma, D.J. (2005). Evolution. Sinauer, ISBN 0-87893-187-2 (hardcover), Sunderland.

Givnish, T.J. (1999). On the causes of gradients in tropical tree diversity. Journal of Ecology, Vol. 87, No. 2, (March 1999), pp. 193-210, ISSN 1365-2745

Govaerts, R. (2001). How many species of seed plants are there? Taxon, Vol. 50, No.4, (November, 2001), pp.1085–1090, ISSN 0040-0262

Hamrick, J. L., Linhart, Y. B. & Milton, J. B. (1979). Relationships between life history characteristics and electrophoretically-detectable genetic variation in plants. Ann. Rev. Ecology and Systematics. 10:173-200.

Hamrick, J. L. (1983). The distribution of genetic variation within and among natural plant populations. In: Schone-Wald-Cox, C. M., Chambers, S. H.; MacByde, B., Thomas, L. (Ed.). Genetics and Conservation. Menlo Park: Benjamin Cummings, 1983. p. 335-348.

Howe, H.F., & Smallwood, J. (1982). Ecology of seed dispersal. Annual Review of Ecology, Evolution, and Systematics, Vol. 13, No. 1, (November 1982), pp. 201–228, ISSN 1543-592X

Instituto Brasileiro de Geografia e estatística (IBGE). (2004). Mapa de Biomas do Brasil, Rio de Janeiro, Brazil

IBGE – Instituto Brasileiro de Geografia e Estatística. (2006). Manual técnico de uso da terra. Rio de Janeiro: IBGE, 2 ed. (Manuais técnicos em Geociências, n.7). 91 p.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2007). Censo Populacional 2005, Rio de Janeiro, Brazil

Kageyama, P. Y. (1987). Conservação in situ de recursos genéticos de plantas. Instituto de Pesquisas e Estudos Florestais - IPEF, 35: p. 7-37.

Kala, C. P. (2000). Status and conservation of rare and endangered medicinal plants in the Indian trans-Himalaya. Biological Conservation, 93: p. 371-379.

Kawecki, T.J., & Ebert, D. (2004). Conceptual issues in local adaptation. Ecology Letters, Vol. 7, No. 12, (December 2004), pp. 1225-1241, ISSN 1461-023X

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. PLoS ONE, Vol. 3, No. 12, (December 2008), pp. e4010, ISSN 1932-6203

Lovejoy, T. E. (1983). Ecological dynamics of tropical forest fragments. In: Sutton, S. L.; Whitmore, T. C., Chadwick, A. C. Tropical Rain Forest: Ecology and Management. Oxford: Blackwell Scientific Publication, p. 377-384.

Loveless, M. D. & Hamrick, J. L. (1987). Ecological determinants of genetic structure in plant populations. Annual Review of Ecology and Systematics, 15: p. 65-95.

Lowe, A.J., Boshier, D., Ward, M., Bacles, C.F.E., & Navarro, C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. Heredity, Vol. 95, No. 4, (August 2005), pp. 255–273, ISSN 13652-540

Matallana,G., Wendt, T., Araujo, D.S.D. & Scarano, F.R. (2005). High abundance of dioecious plants in a tropical coastal vegetation. American Journal of Botany, Vol. 92, No. 9, (September 2005), pp. 1513–1519, ISSN 1537-2197

Medeiros, R.; Irving, M. & Garay, I. (2004). A proteção da natureza no Brasil: Evolução e conflitos de um modelo em construção. Revista de desenvolvimento econômico, Vol. No.9, (January, 2004) ISSN 1516-1684

Medri, C., Ruas, E.A., Medri, M.E., Ruas, C.F., Sayhun, S.A., Medri, P.S., Silva, D.C.G., Bianchini, E. & Ruas, P.M. (2011). Genetic diversity and flooding survival in Aegiphila sellowiana (Lamiaceae), a typical tree species from upland riparian forests. Genetics and Molecular Research, Vol. 10, No. 2, (June 2011), pp. 1084-1091, ISSN 1676-5680

Medri C., Ruas E.A., Ruas C.F., Medri P.S., Medri M.E. & Ruas, P. M. (2011). Population genetic structure of the tropical tree species Aegiphila sellowiana (Lamiaceae). Genetic and Molecular Research (in press).

Medri, M.E., Bianchini, E., Shibatta, O.A. & Pimenta, J.A. (2002). A Bacia do Rio Tibagi, Câmara Brasileira do Livro, ISBN 85902390-1-2/85-902392-1-7/85-902394-1-1/85-902395-1-9, Londrina, Brazil

Naiman, R.J. & Decamps, H. (1997). The ecology of interfaces: riparian zones. Annual Review of Ecology and Systematics, Vol. 28, (November 1997), pp. 621-658, ISSN 0066-4162

Nei, M. (1973) Analysis of gene diversity subdivided populations. Proceedings of the National Academy of Science, USA 70: p. 3321-3323.

O'Gorman, E. (1972). The invention of America: An inquiry into the historical nature of the New World and the meaning of its history. Greenwood Press, ISBN 0837164702, Westport, USA

Oliveira, R.B.R. (2010). Estrutura genética de cinco populações de Aegipila sellwiana Cham. (Lamiaceae) da bacia do Rio Tibagi por marcadores microssatélites. Master degree dissertation, Universidade Estadual de Londrina, Londrina, Brazil, 63 p.

Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J. & Hirota, M.M. (2009). The Brazilian Atlantic Forest: How much is left, and how is the remaining forest

distributed? Implications for conservation. Biological Conservation, vol. 142, No.6, (March 2009), p. 1141-1153, ISSN 0006-3207

Ruas, E.A., Conson, A.R.O., Costa, B.F., Damasceno, J.O., Rodrigues, L.A., Reck, M., Vieira, A.O.S., Ruas, P.M. & Ruas, C.F. (2009). Isolation and characterization of ten microsatellite loci for the tree species Luehea divaricata Mart. (Malvaceae) and intergeneric transferability. Conservation Genetics Resources, Vol.1, No.1, (August 2009), pp.245-248, ISSN 1877-7260

Ruas, E.A., Ruas, C.F., Medri, P.S., Medri, C., Medri, M.E., Bianchini, E., Pimenta, J.A., Rodrigues, L.A. & Ruas, P.M. (2011). Anatomy and genetic diversity of two populations of Schinus terebinthifolius (Anacardiaceae) from the Tibagi River basin in Paraná, Brazil. Genetics and Molecular Research, Vol. 10, No. 1, (March 2011), pp. 526-536, ISSN 1676-5680

Ruas, E.A., Damasceno, J.O., Conson, A.R.O., Costa, B.F., Rodrigues, L.A., Reck, M., Vieira, A.O.S., Ruas, C.F., Medri, C. & Ruas, P.M. (2011). Isolation and characterization of eleven polymorphic microsatellite loci in Aegiphila sellowiana and their transferability. Biologia Plantarum, Vol. 55, No. 2, (June 2011), pp. 396-399, ISSN 1573-8264

Sahyun, S.A., Ruas, E.A., Ruas, C.F., Medri, C., Souza, J.R.P., Johan, L.A.P.S., Miranda, L.V. & Ruas, P.M. (2010). Genetic Variability of three Natural Populations of Maytenus aquifolium (Celesteraceae) from Telêmaco Borba, Paraná, Brazil. Brazilian Archives of Biology and Technology, Vol. 53, No. 5, (September/October 2010), pp.1037-1042, ISSN 1516-8913

Savolainen, O., Pyhäajärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. Annual Review of Ecology, Evolution, and Systematics, Vol. 38, No. 1, (December 2007), pp. 595-619, ISSN 1543-592X

Schultz, J. (2005). The ecozones of the world: The ecological divisions of the geosphere. Springer, ISBN 978-3-540-20014-7, New York, USA

Shannon, C.E. & Weaver, W. (1949) The mathematical theory of communication. Urbana, The Univeristy of Illinois Press 117pp.

Silva, D.C.G., Carvalho, M.C.C.G., Ruas, P.M., Ruas, C.F. & Medri, M.E. (2010). Evidence of ecotypic differentiation between populations of the tree species Parapiptadenia rigida due to flooding. Genetics and Molecular Research, Vol. 9, No. 2, (May 2010), pp. 797-810, ISSN 1676-5680

Souza L. B. 2011. Estrutura genética de populações de Parapiptadenia rigida (Benth) Brenan (Leguminosae-Momosideae) na região sul do Brasil por marcadores AFLP. Master degree Dissertation, Universidade Estadual de Londrina, Londrina, Brazil, 57p.

Stöcklin, J., Kuss, P. & Pluess, A.R. (2009). Genetic diversity, phenotypic variation and local adaptation in the alpine landscape: case studies with alpine plant species. Botanica Helvetica, Vol. 119, (November 2009), pp.125–133, ISSN 0253-1453

Takhtajan, A. 1986. Floristic regions of the world. University of California Press [An English translation of the original Russian book published in 1978], ISBN 0520040279 London and Berkeley, USA

Thomas, W. W. (1999). Conservation and monographic research on the flora of Tropical America. Biodiversity and Conservation, Vol. 8, No. 8, (August 1999), pp. 1007–1015, ISSN 0960-3115

Torezan, J.M.D. & Silveira, M. (2002). Fatores ambientais, diversidade e similaridade em florestas da bacia do rio Tibagi, In: A Bacia do rio Tibagi, Medri, M.E., Bianchini, E., Shibatta, O.A. and Pimenta, J.A, pp. 125-131, Câmara Brasileira do Livro ,ISBN 85902390-1-2/85-902392-1-7/85-902394-1-1/85-902395-1-9, Londrina, Brazil

Wittmann, A.O., Piedade, M.T.F., Wittmann, F., Schöngart, J. & Parolin, P. (2007). Patterns of structure and seedling diversity along a flooding and successional gradient in amazonian floodplain forests. Pesquisas, série Botânica, No. 58, pp. 119-138, ISSN 0373-840X

Wright, S. (1951). The genetical structure of populations. Annals of Eugenics 15: p. 323-354.

Young, A., Boyle, T., & Brown, A.H.D. (1996). The population genetic consequences of habitat fragmentation for plants. Trends in Ecology & Evolution, Vo. 11, No. 10, (October 1996), p. 413-418, ISSN 01695-347

21

A Brief Review of a Nearly Half a Century Wheat Quality Breeding in Bulgaria

Dobrinka Atanasova, Nikolay Tsenov and Ivan Todorov
Dobrudzha Agricultural Institute
General Toshevo
Bulgaria

1. Introduction

Wheat quality is a complex feature determined by the level of various indices and dependent on a number of factors. It is defined by different interactions involving glutenin and gliadin composition and abiotic stresses. The efforts of the wheat breeders are directed towards combining a high level of the indices in the new varieties with a stable expression of these indices in different years and environments.

The aim of this chapter is 1) to describe the glutenin and gliadin composition of Bulgarian winter bread wheat varieties, bred since the middle of 20th century; 2) to asses the quality indices of wheat varieties; to compare the differences between wheat quality groups as well as the different period of creating the varieties; 3) to determine the level of influence of various environmental factors on the quality indices of wheat; to follow the response of common winter wheat varieties to various combinations of growing conditions; to study the ability of the varieties to realize their genetic potential for quality under certain environments.

2. Genetic diversity of Bulgarian winter wheat varieties in relation to glutenin and gliadin compositions

Wheat storage proteins (gliadins and glutenins) are the main components of gluten which is the major determinant of end use quality. Glutenin proteins are divided into two groups: High and Low Molecular Weight Glutenin Subunits (HMW-GS and LMW-GS, respectively). HMW-GS are encoded by two type of genes (x and y) that are located on three loci (Glu-A1, Glu-B1 and Glu-D1) placed on the long arm of the group 1 chromosomes (Lawrence & Shepherd, 1981). LMW-GS are classically divided into B, C and D groups on the basis of molecular weight and isoelectric point (Jackson et al., 1983). They are coded by gene families located on the short arm of the group 1 chromosomes (Glu-A3, Glu-B3 and Glu-D3 loci). Gliadins are monomeric proteins and they are classified as α-, β-, γ- and ω-gliadins. Most of the α/ β-gliadins are encoded at the Gli-2 loci, located on the short arms of chromosomes 6A, 6B and 6D. Most of γ- and ω-gliadins are encoded at the Gli-1 loci, located on the short arms of chromosomes 1A, 1B and 1D (Masci et al., 2002). It has been reported a close linkage

between the Glu-3 loci encoding LMW-GS and the Gli-1 loci for gliadins (Sing and Shepherd, 1988).

The bread making characteristics of wheat flour are closely related to the elasticity and extensibility of the gluten proteins. The seed storage proteins are associated with agriculturally significant traits and they are used in a legal protection of cultivars (Knoblochova & Galova, 2000). Many studies have been made in order to investigate the genetic diversity of the initial material (Atanasova et al., 2009b; Bradova, 2008; Bushehri et al., 2006; Li et al., 2005; Tabasum et al., 2011; Tohver, 2007; Tsenov et al., 2009).

Bulgarian wheat varieties, especially those developed at Dobrudzha Agricultural Institute - General Toshevo (DAI), are known to have good and very good end-use quality (Atanasova et al., 2008; Panayotov et al., 2004, Todorov, 2006; Tsenov et al., 2010a). Their glutenin composition has been investigated to various degrees (Atanasova et al., 2009b; Todorov et al., 2006; Tsenov et al., 2009).

In this review eighty-nine cultivars developed at Dobrudzha Agricultural Institute (DAI) – General Toshevo, Bulgaria, during 1962–2010, and nine cultivars of the Institute of Plant and Genetic Resources (IPGR) – Sadovo, Bulgaria were investigated. Some of these cultivars were heterogeneous and their biotypes were considered as separate varieties when determining the ratio of high- and low- molecular weight glutenin and gliadin subunits and spectra. The allelic frequency in loci *Glu-1, Glu-3,* and *Gli-1* was studied and the genetic variability in each allele was calculated. The percent of observed spectra was determined by decades.

In high-molecular weight glutenins, highest variability was registered in locus *Glu-B1* (fig. 1). Five alleles were identified – *a, b, c, d, f*, which expressed subunits 7, 7+8, 7+9, 6+8 and 13+16, as well as one untypical fraction pair 6+9 which was observed in a biotype of cultivar Skitiya. Allele *c* had highest frequency (65.4%), followed by allele *b* (17.7%) and allele *d* (10.8%). Three alleles were identified in locus *Glu-A1*: *a, b, c*, expressing the respective subunits 1 (24.6%), 2* (41.5) and N (33.8%). In locus *Glu-D1* allele *a* (subunit 2+12) had frequency 27.7%, and allele *d* (subunit 5+10) - 69.2%. Biotypes with the untypical fraction pair 5+12 were observed in cultivar Levent.

In the low-molecular weight glutenins, locus *Glu-B3* had highest variability, where 7 alleles were identified – *b, d, f, g, h, i, j*, followed by locus *Glu-A3* with 5 alleles – *b, c, d, e, f* ; in locus *Glu-D3* two alleles were identified – *a* and *c*. Allele *Glu-A3c* (69.2%) had highest frequency, followed by allele *Glu-A3e* (18.5%). The other alleles had frequency less than 10 %. Allele *Glu-B3b* had frequency 47.7%, followed by *Glu-B3j* (13.8%), *Glu-B3f* – 13.1%, *Glu-B3h* – 10.0%.

The gliadin fraction composition of the investigated wheat accessions was more variable than the fraction composition of the high- and low-molecular weight glutenins. In each of all three gliadins loci 9 alleles were observed. In locus *Gli-A1* the frequency of alleles *Gli-A1b* (40.8%), *Gli-A1m* (18.5%) and *Gli-A1a* (17.7%) was highest. In locus *Gli-B1* the alleles with highest frequency were: *Gli-B1b* (47.7%), *Gli-B1l* (13.8%), *Gli-B1g* (13.1%) and *Gli-B1d* (10.0%). In locus *Gli-D1* the alleles with highest frequency were *Gli-D1b* (56.2%), *Gli-D1j* (13.1%) and *Gli-D1a* (11.5%).

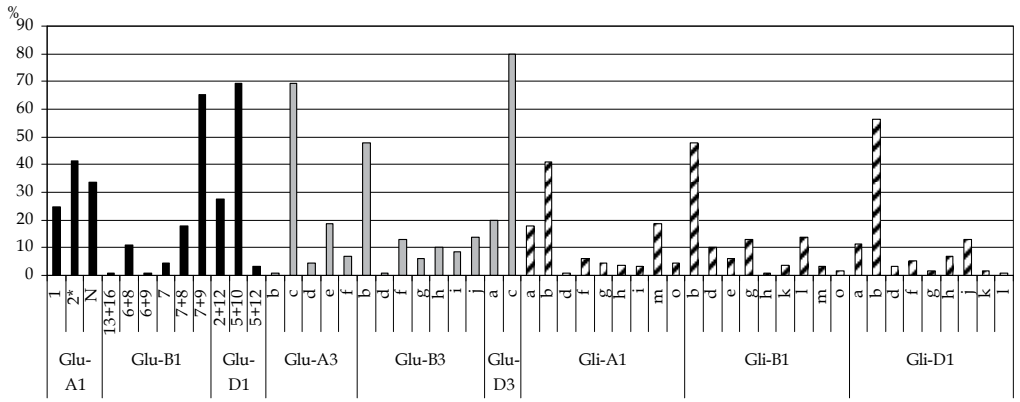


Fig. 1. The allelic frequency in loci *Glu-1, Glu-3,* and *Gli-1*

Twenty-four high-molecular weight, twenty-six low-molecular weight glutenin and fifty-seven gliadins spectra were identified in the investigated wheat accessions. The most frequent high-molecular weight spectra were 2* 7+9 5+10 (22.14%), N 7+9 5+10 (14.5%), 2* 7+9 2+12 (8.4%), 1 7+9 5+10 (8.4%) (table 1). Among the low-molecular weight glutenin spectra, the following had highest frequency: *c b c* (22.9%), *e b c* (12.2%), *c f c* (8.4%), *c h c* (6.9%), *c j c* (6.9%). In gliadins, most frequent were spectra *b b b* (12.2%), *m b b* (6.9%), *m b j* (6.1%).

Loci	Spectrum	%
Glu-1	2* 7+9 5+10	22.14
	N 7+9 5+10	14.5
	2* 7+9 2+12	8.4
	1 7+9 5+10	8.4
Glu-3	c b c	22.90
	e b c	12.21
	c f c	8.40
	c h c	6.9
	c j c	6.9
Gli-1	b b b	12.2
	m b b	6.9
	m b j	6.1

Table 1. The most frequent spectra in the *Glu-1*, *Glu-3* and *Gli-1* loci

The highest frequency of *Glu-1* spectra in the 70's was for N 7+9 5+10 (19.4%), 2* 7+9 2+12 (13.9%), 2* 7+9 5+10 (13.9%), N 7+9 2+12 (11.1%) (table 2). In the 80's these were 2* 7+9 5+10 (20.7%), 1 7+9 2+12 (10.3%), N 6+8 5+10 (10.3%), N 7+9 5+10 (10.3%). The most frequent in

the 90's were the spectra 2* 7+9 5+10 (30.4%), N 7+9 5+10 (21.7%), 1 7+9 5+10 (17.4%), N 7+8 5+10 (13.0%). In the new century widespread were: 2* 7+9 5+10 (25.6%), 2* 7+8 5+10 (14.0%), 2* 7+9 2+12 (11.6%), 1 7+8 5+10 (9.3%), 1 7+9 5+10 (9.3%), N 7+9 5+10 (9.3%). The frequency of *Glu-3* and *Gli-1* spectra in the different decades is shown in the table.

Decade	Glu-1	Glu-3	Gli-1
70	N 7+9 5+10 (19.4)*	c b c (33.3)	b b b (14.3)
	2* 7+9 2+12 (13.9)	c i c (16.7)	a b b (5.7)
	2* 7+9 5+10 (13.9)	c h c (13.9)	g b b (5.7)
	N 7+9 2+12 (11.1)	c i a (11.1)	o b d (5.7)
	1 7+9 5+12 (5.6)	d b c (8.3)	b d b (5.7)
	N 6+8 5+10 (5.6)	c b a (2.8)	b d j (5.7)
80	2* 7+9 5+10 (20.7)	c f c (20.7)	m b b (10.3)
	1 7+9 2+12 (10.3)	e b c (17.2)	m b j (6.9)
	N 6+8 5+10 (10.3)	c h c (10.3)	b e a (6.9)
	N 7+9 5+10 (10.3)	c b c (6.9)	b g f (6.9)
	1 7 2+12 (6.9)	c f a (6.9)	I g f (6.9)
	1 7+8 5+10 (6.9)	c g a (6.9)	f l b (6.9)
	1 7+9 5+10 (6.9)	f j c (6.9)	b b a (3.45)
90	2* 7+9 5+10 (30.4)	c b c (21.7)	b b b (13.0)
	N 7+9 5+10 (21.7)	c j c (13.0)	a e b (8.7)
	1 7+9 5+10 (17.4)	c g a (8.7)	m l b (8.7)
	N 7+8 5+10 (13.0)	c g c (8.7)	a b b (4.3)
	1 7+9 2+12 (8.7)	c j a (8.7)	f b b (4.3)
	N 7+9 2+12 (8.7)	e j c (8.7)	b b g (4.3)
After 2000	2* 7+9 5+10 (25.6)	c b c (25.6)	b b b (16.3)
	2* 7+8 5+10 (14.0)	e b c (23.3)	m b b (13.95)
	2* 7+9 2+12 (11.6)	c j c (11.6)	m b j (9.3)
	1 7+8 5+10 (9.3)	c b a (7.0)	b l b (9.3)
	1 7+9 5+10 (9.3)	c f c (7.0)	g b b (4.7)
	N 7+9 5+10 (9.3)	d b c (4.7)	o b d (4.7)
	N 6+8 5+10 (4.7)	f b c (4.7)	b b h (4.7)

* - number in the parentheses indicates the percentage of frequencies

Table 2. Distribution of the most common spectra in decades

In order to improve the quality of new varieties, a narrowing of the genetic diversity was observed especially after the mass penetration of the variety Bezostaya 1 in the breeding

programs after the 80's. The direct usage as a parent in the hybridization or indirectly as a participant in the pedigree of other varieties, led to unification of the spectra of new varieties with that of Bezostaya 1. It is necessary to broaden the genetic basis by including alleles with good influence over the wheat quality. Such alleles are *Glu-B1f* (subunit 13+16), *Glu-B1i* (subunit 17+18), *Glu-A3f*, *Glu-A3b*.

3. Progress in development of quality parameters of Bulgarian wheat varieties

The breeding for wheat quality started in 1962 when the first contemporary wheat breeding program in the Dobrudzha Agricultural Institute was adopted. In the two breeding centers (DAI and IPGR) 26 high quality varieties were created till now which was 1/3 of the entire variety list. During 2006 and 2008 sixty nine varieties were tested for their quality parameters. The varieties belonged to different quality groups according to Bulgarian State Standard (BSS). Based on the level of indices sedimentation value, wet glutenin content and valorimeter, varieties from this investigation were divided in three groups (table 3).

Level of indices	Varieties
	Strong wheat
sedim. > 47 WGC > 23 val. > 47	Aglika, Albena, Antitsa, Bezostaya 1, Goritsa, Ideal, Iveta, Laska, Lazarka, Merilin, Milena, Miziya, Pobeda, Slavyanka 196, Zlatina
	Medium wheat
sedim. 38 - 46 WGC 18 - 23 val. 40 - 47	Bolyarka, Charodeika, Enola, Karina, Liliya, Neda, Sadovo 552, Vratsa, Yanitsa, Yunak
	Weak wheat
sedim. < 38 WGC < 18 val. < 40	Antonovka, Kaliakra 2, Karat, Kristi, Petya, Pliska, Prelom, Svilena, Todora, Yantar, Zlatitsa

Table 3. Grouping of varieties according to the level of sedimentation (sedim.), wet glutenin content (WGC) and valorimeter (val.)

Some of the strong wheat varieties, according to BSS, failed to enter the first group in this investigation. Such varieties were: Demetra - with lower WGC; Preslav – with lower WGC and valorimeter; Dona – with lower sedimentation and WGC and Ludogorka and Momchil with lower levels for all three indices. A tendency in breeding of high quality varieties with low wet gluten content, even in comparison to medium and low-quality varieties, has been reported by Atanasova et al. (2010), Panayotov & Rachinsky (2002). One of the reasons for this is the use of Russian and Ukrainian sources in the breeding programs for wheat quality improvement. These sources have high quality of gluten but in lower amounts.

Wheat varieties, belonging to second group by BSS, in this investigation showed level of indices for the first group. With higher level of all three indices were varieties Bozhana and Stoyana. With higher sedimentation and valorimeter were Slaveya and Venka 1. Varieties Galateya, Prostor, Zagore had higher WGC and valorimeter value. Some varieties from third

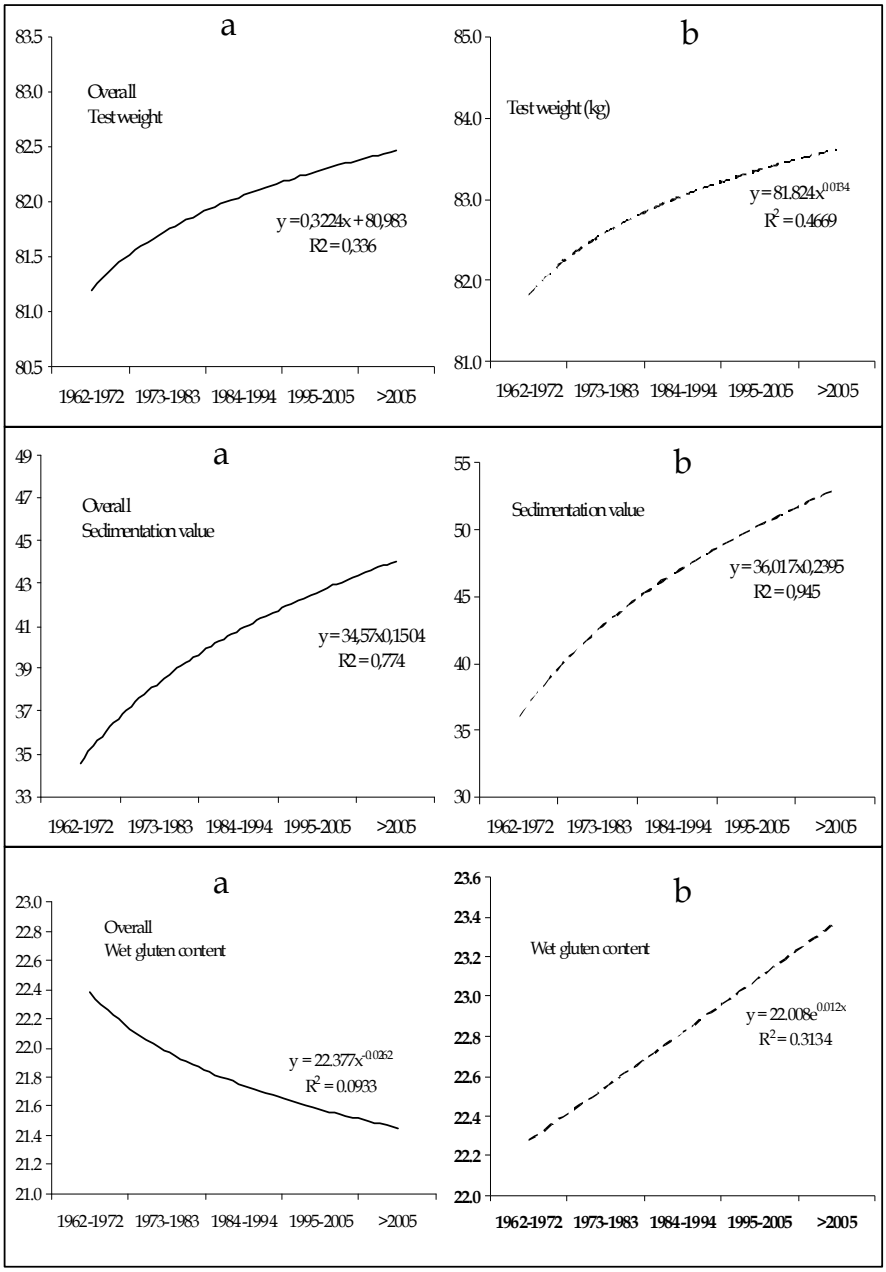


Fig. 2. Changes of quality indices by decades for the entire group (a) and for the quality group (b) of wheat varieties

group by BSS, also, showed higher level of some of the indices. For example varieties Trakiya and Ogosta had higher level of valorimeter value. Variety Pryaspa had higher WGC and varieties Charodeika and N 100-10 had higher level of the two indices. Similar discrepancies in the level of indices were observed in other investigations (Atanasova et al.,

2009a; Atanasova et al., 2010). This demonstrates the complex nature of the wheat quality, which is affected not only by genotype but by environments as well.

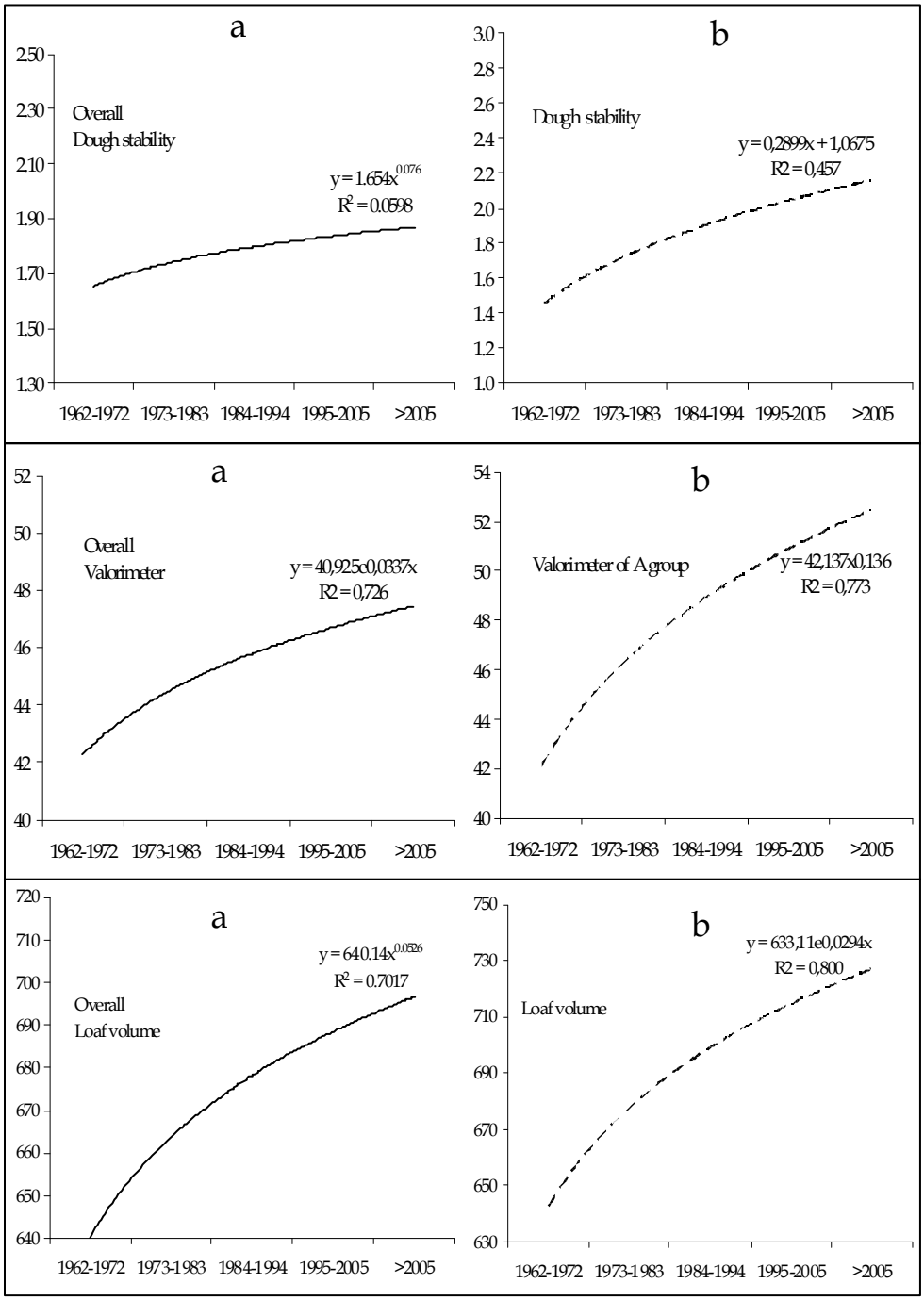


Fig. 2. Continue

As a result of breeding a progress was determined in almost all the indices defining end-use quality in wheat. The breeding periods were divided into decades as follows: 1962-1972; 1973-1983; 1984-1994; 1995-2005 and after 2005. The average values of the whole group of sixty-nine wheat varieties and that of the high-quality group were compared. The most strongly modification was observed for the indices: dough stability, where the enlargement was 24% for the entire group and 60% for the quality group; sedimentation value – 26%for whole group and 44% for quality group (figure 2). For the indices valorimeter and loaf volume a high level had been reached which changed very slowly through the years. The group of quality wheat varieties had greater progress compared to the all varieties regardless of the quality index and its behavior. It was valid even to the indices for which the breeding progress was not statistically proven.

4. Evaluation of genotype, environment and their interaction for quality parameters of Bulgarian varieties

The complex nature of wheat quality is determined by the level of various indices and dependent on a number of factors. One of the problems of breeding for quality is not only achieving a high level of the indices in the new varieties, but also achieving a stable expression of these indices in different years and environments (Atanasova et al., 2008; Johansson et al., 2001; Williams et al., 2008). It is valuable for breeding to determine the nature and direction of the effects of different genetic and environmental factors on the specific quality indices (Hristov et al., 2010; Panozzo & Eagles, 2000; Tsenov et al., 2004), and to study the performance of the varieties under changeable growing conditions. This would allow predicting to some degree their response to certain combinations of environmental factors (Atanasova et al., 2010; Drezner et al., 2006; Yong et al., 2004). The variation of the quality indices when growing each individual variety under different environments is always helpful both for the distribution of the new varieties and for their improvement in breeding programs (Gomez-Becera et al., 2010). The data on the variation of winter wheat quality is contradictable with regard to the level of genotype effect (Gomez-Beccera et al., 2010; Williams et al., 2008). In breeding there is always the question whether the high quality of a given variety under favorable environments is a prerequisite for the realization of its high-quality potential under unfavorable environments, too, or is it and impediment for its stability (Hristov et al., 2010; Tsenov et al., 2004).

Sixteen Bulgarian winter wheat varieties from first and second quality group according to BSS were tested at two locations: Dobrudzha Agricultural Institute – General Toshevo (DAI) and Institute of Agriculture and Seed Science, Obraztsov Chiflik, Rouse (OCH) during 2004–2007. The expression of 5 grain quality indices which give information about various quality aspects was analyzed: test weight (kg) (BSS 7971-2:2000), sedimentation value of flour (ml) (Pumpyanskii, 1971), wet gluten content in grain (%) (BSS 13375-88), valorimeric value (valorimeter, conditional units) (BSS 16759-88), and loaf volume, determined according to the methods adopted at the DAI laboratory.

The effect of the location was highest on wet gluten content and valorimeric value (table 4). The traits test weight and loaf volume were affected most by the year conditions. The distribution of the varieties into groups according to their quality had highest influence on the expression of sedimentation. The independent influence of the genotype on the variation of most of the investigated indices was lowest. Atanasova et al. (2008) and Zhang et al.

(2005) have found out that the genotype had higher effect on the expression of sedimentation, while Drezner et al. (2006, 2007), Panozzo and Eagles (2000), Tsenov et al. (2004) have proved that the environment has greater effect on the quality indices of wheat. Among the various combinations between the factors, the interaction *quality group x genotype* had the highest effect on almost all investigated traits. The exception was wet gluten content, which was affected most by the interaction *location x year*. Hristov et al. (2010), Mladenov et al. (2001), Williams et al. (2008) have found out that in spite of the significant effect of the interaction *genotype x environment* on the expression of the quality parameters of wheat, this effect was less significant than the independent influences of the genotype and the location.

Source of variation	TW	SDS	WGC	Val	Lvol
Main effects					
A: Location	161.1***	1250.0***	170.4***	1058.0***	43512.5***
B: Year	205.5***	876.6***	91.6***	173.0***	72179.4***
C: Group	76.3***	2032.0***	7.9 ns	666.1***	39903.1***
D: Genotype	11.2***	28.9**	22.2***	48.7**	3868.2***
Interactions					
A x B	4.7***	94.6***	76.1***	8.4ns	703.6ns
A x C	7.3***	101.5***	17.3**	63.3**	2032.0ns
A x D	1.1*	4.0ns	1.5 ns	13.5ns	942.4ns
B x C	0.3ns	15.6ns	13.0**	26.4*	3650.5***
B x D	0.4ns	21.5**	5.3ns	9.0ns	2273.9***
C x D	17.4***	172.7***	13.2**	157.4***	11036.6***

*P<0.1; **P<0.05; ***P<0.01; *ns*-not significant

Table 4. Analysis of variance for mean squares of 16 wheat varieties

The growing conditions at DAI were more favorable for expressing genetic potential of wheat varieties as it is revealed by mean values of the indices (table 5). It is logical to expect higher values for the indices of the varieties from the first quality group. The results from the variance analysis were once again confirmed; the values for wet gluten content were not significant and there were no considerable differences between the two quality groups.

It is difficult in breeding for wheat quality to develop varieties with high quality indices, which remain stable under various growing conditions (Atanasova et al., 2008; Johansson et al., 2001; Williams et al., 2008). According to Finlay & Wilkinson (1963), Sudaric et al. (2006) genotypes with coefficient of regression (b_i) under 0.7 were considered unresponsive to changeable environments; with b_i between 0.7 and 1.3 had average stability; and >1.3 were considered responsive to good environments. Most of the varieties involved in this investigation have moderate stability for all the indices with some exceptions (table 6). There wasn't any regularity in manifestation of the stability. For some traits varieties showed stable reaction as b_i was near 1.0. For other traits varieties were specifically adapted to

growing in unfavorable environments as b_i was below 0.7. For example variety Enola was with stable reaction for test weight, sedimentation value, WGC (b_i was 1.01, 1.15, 1.08, respectively) and with b_i=0.598 for loaf volume.

Factors	TW	SDS	WGC	Val	Lvol
Location					
DAI	81,2^{b}	39,4^{b}	21,6^{b}	43,7^{b}	673,8^{b}
OCH	78,9^{a}	33,1^{a}	19,3^{a}	38^{a}	636,9^{a}
Year					
2004	81,5^{c}	36,5^{b}	20,0^{a}	42,4^{b}	641,1^{a}
2005	76,5^{a}	43,4^{c}	22,4^{b}	41,8^{b}	725,6^{b}
2006	80,2^{b}	31,3^{a}	18,4^{a}	41,7^{b}	633,1^{a}
2007	82,2^{c}	33,8^{ab}	21,0^{ab}	37,4^{a}	621,7^{a}
Group of quality					
I-group	80,8^{b}	40,2^{b}	20,7^{a}	43,1^{b}	673,0^{b}
II-group	79,3^{a}	32,3^{a}	20,2^{a}	38,5^{a}	637,7^{a}

Numbers with different letters differ significantly

Table 5. Mean values of the investigated traits according to different factors of variation

The valorimeric value is an index which determines gluten quality and which correlates strongly with end-use quality. Most of the investigated varieties possess moderate stability (b_i from 0.7 to 1.3) with regard to this trait. Varieties Milena and Aglika (b_i = 1.00 and b_i = 1.25, respectively) from first quality group were an evidence that high values of the traits can be observed in stable genotypes, too (Hristov et al., 2010; Mladenov et al., 2001; Sudaric et al., 2006) and that stability is not necessarily related to low mean values, as previously stated (Becker and Leon, 1988). The ecovalence values of the individual varieties for this trait were very high, which brings forth the strong effect of the environment. The variation expressed through the ecovalence was much lower in the wheat varieties from the second group in comparison to the varieties from the first group (56.8 and 99.6, respectively) (table 6).

A number of researchers have found out that the varieties with lower quality potential are considerably more stable, especially under stress (Atanasova et al., 2010; Tsenov et al., 2004). The values of the regression coefficient of the separate groups are quite indicative in this respect. The regression coefficient in the first group was above 1.0. This means that that the varieties from this group realized their quality potential under conditions more favorable for its formation. Therefore the W_i values of variety Zlatina, which is a quality variety, were several times higher than the mean values for this group. On the other hand, varieties like Aglika, Milena and Preslav demonstrated considerably lower variation of their quality indices in comparison to the other varieties from the group. It can be suggested that such stability is due to the fact that the absolute values of the indices in this case were low. These data once again prove that when comparing the response of the varieties to the environment

with regard to their quality, conclusions should be made very carefully. The results for the individual varieties and their indices are not unidirectional and therefore when selecting a variety for mass production it is necessary to consider the purpose for its cultivation, the potential of the given variety and its ability to realize this potential with stability and adequacy.

Variety	Test weight		Sedimentation		Wet gluten content		Valorimetric value		Loaf volume	
	bi	Wi	bi	Wi	bi	Wi	bi	Wi	bi	Wi
Igroup										
Pobeda	1.09	2.75	0.84	8.48	1.174	21.2	1.552	121.2	1.300	7.04
Albena	0.92	1.85	1.29	15.4	0.854	12.2	1.020	59.67	0.650	9.67
Preslav	1.05	2.73	1.43	8.01	1.226	9.37	0.648	26.26	1.015	10.59
Milena	1.07	1.54	0.91	10.7	1.219	36.5	1.000	21.33	0.834	4.56
Aglika	1.26	6.24	1.01	18.2	1.022	10.5	1.251	89.42	1.166	12.45
Progres	1.40	9.48	1.28	6.39	0.656	36.7	0.700	42.86	1.283	9.32
Zlatina	1.10	2.40	1.11	10.2	0.800	23.1	1.855	274.2	1.350	12.36
Demetra	0.95	3.16	1.62	22.3	1.709	40.4	0.754	162	0.987	2.46
Average	1.104	3.77	1.186	12.46	1.080	23.72	1.100	99.62	1.073	8.56
II group										
Sadovo 1	1.3	5.24	1.21	5.52	1.154	14.5	1.331	133.1	1.303	5.45
Enola	1.01	3.53	1.15	16.3	1.078	82.6	0.963	86.11	0.598	11.00
Pliska	1.13	2.24	1.20	11.5	0.645	47.0	0.814	14.58	1.019	4.22
Pryaspa	1.07	3.00	0.88	5.33	1.229	14.3	0.912	23.51	0.905	10.67
Yantar	0.4	25.8	0.65	6.46	0.860	15.4	0.658	41.83	0.909	2.72
Kristi	0.61	8.35	0.75	5.86	0.900	25.6	0.736	70.67	0.789	8.09
Prelom	0.85	3.05	0.85	9.18	0.671	19.3	0.648	28.39	0.960	11.77
Boryana	1.19	3.52	0.69	2.88	0.803	86.6	1.158	56.17	1.198	8.69
Average	0.944	6.84	0.923	7.88	0.920	38.16	0.900	56.79	0.960	7.83

Table 6. Stability parameters of varieties from different quality groups

When compare the mean values of varieties with their stability it can be seen that for the conditions of DAI varieties Demetra, Zlatina, Aglika differ from the total group with higher levels of the indices (figure 3). For the other location (OCH) these varieties were Demetra, Enola, Aglika. The most poorly performing variety for the two locations was Prelom.

Projection to the ordinate of the two-dimensional matrix genotype x character, regardless of the direction, determines the stability of the varieties (Yan et al., 2007). For the conditions of DAI as most stable and exceeding the mean level were varieties Aglika, Sadovo 1, Preslav, Demetra. Varieties Milena and Preslav had most stable reaction at OCH location as they had almost null projection to the ordinate. Under the same conditions with high variability stand out varieties Enola and Pliska.

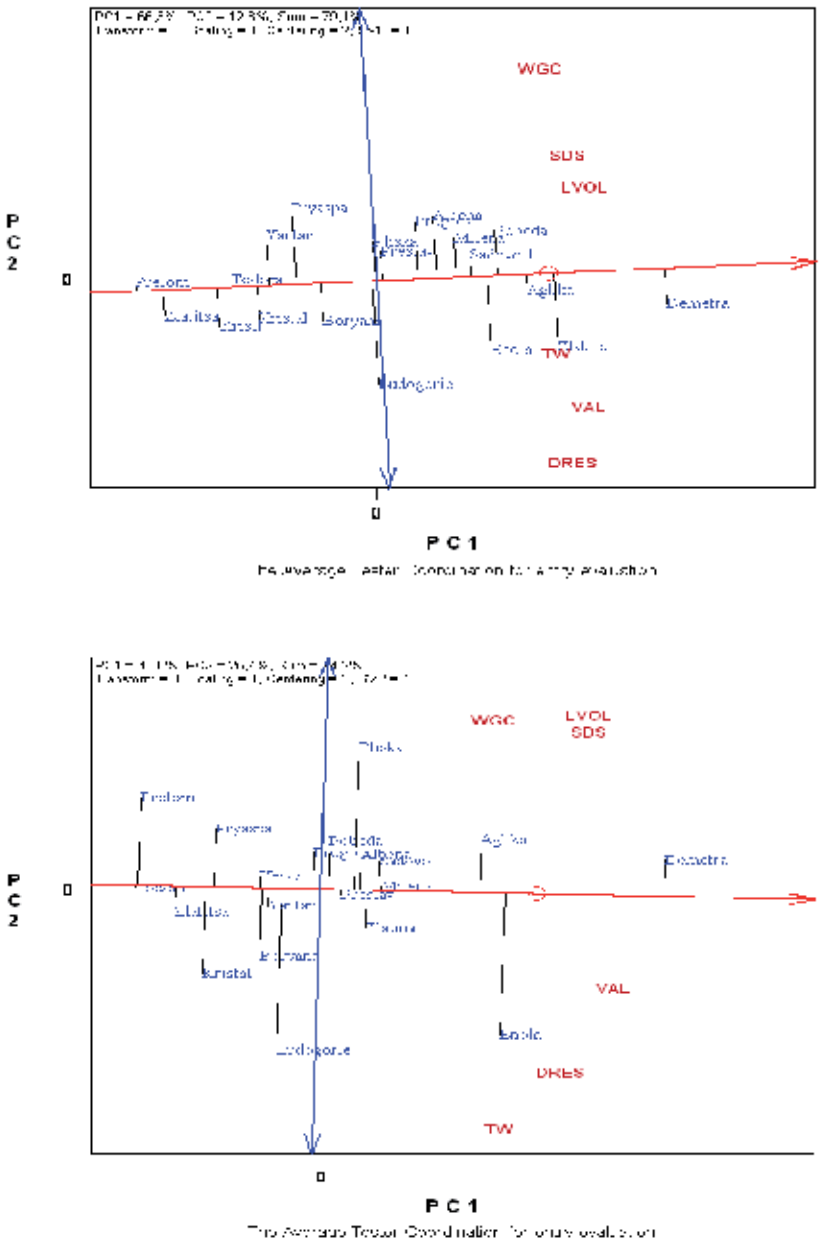


Fig. 3. Comparison between mean values of the varieties and their stability for investigated indices for conditions of DAI (up) and OCH (down)

Another investigation was carried out during 2001-2003 with 10 varieties with different quality in two typical for grain production in Bulgaria regions – Dobrudzha Agricultural Institute (DAI) in North-East Bulgaria and Chirpan in South Bulgaria. Data presentation by Principal Component Analysis (PCA) allows more comprehensive explanation of the

variation of quality indices of studied varieties in the two locations (Stoeva et al., 2006). There was a similar component structure of the factors which mainly influence the variation of the quality over locations. About 65% of the total variation of the two-dimensional matrix variety x character for each location is for first and second PC. The biggest part is for PC1 (table 7).

Indices	DAI		Chirpan	
	PC1	PC2	PC1	PC2
Test weight, kg	0.647	-0.385	0.587	0.395
Vitreousness, %	0.603	-0.459	0.358	0.845
Endosperm hardiness, %	0.835*	-0.258	0.674	0.678
Wet gluten content, %	0.493	0.421	0.325	-0.361
Sedimentation value, ml	0.910**	0.206	0.916**	-2,362
Dough stability, min	0.805*	1,841	0.971**	-2,149
Valorimeric value,	0.922**	2,983	0.990**	-2,017
Bread loaf, ml	0.778*	0.181	0.742*	-0.505

Table 7. Correlation between quality indices, PC1 and PC2 over locations

Indices endosperm hardiness, sedimentation value, dough stability, valorimeter and bread loaf were determinative for the quality of the studied varieties in DAI as the levels of the PC1 were positive and higher than 0.700. For the other location (Chirpan) these were sedimentation value, dough stability, valorimeter and bread loaf.

Varieties from this investigation had different reaction to the environment in the two locations. Varieties with lower quality such as Pryaspa, Karat, Todora showed low variability connected with not linear reaction to the environments as their PC1 and PC2 were negative (table 8). High quality varieties Aglika, MIlena, Galateya showed high stability in the two locations irrespective of the direction in the changing environments. The different sing in the reaction of varieties show that some of them react specifically to the environments. For example in poor environments some varieties managed to save higher levels of indices and vice versa.

Variety	DAI		Chirpan	
	PC1	PC2	PC1	PC2
Yantar	0.334	0.684	0.452	0.731
Pryaspa	-0.166	-0.490	-0.121	-0.395
Milena	0.722	0.764	0.801	0.884
Aglika	2.001	-0.103	1.120	0.030
Galateya	0.922	-0.386	0.793	-0.268
Albena	-0.507	0.957	-0.317	0.784
Enola	-0.080	1.130	0.009	0.922
Karat	-0.406	-0.814	-0.566	-0.753
Kristal	-1.668	0.388	-1.284	0.299
Todora	-0.493	-2.131	-0.569	-1.977

Table 8. Values of principal components of varieties

Wheat growing depends on environments although the right technologies and new varieties can manage to lower this dependence. That's why it is valuable to select varieties which give high quality in different environments.

5. Breeding of wheat quality in Bulgaria – Steps, drawbacks, prospects

Breeding of new varieties with good quality has always been on the attention of the breeding programs in Bulgaria (Boyadjieva et al., 1999; Panayotov et al., 1994; Rachinski, 1966; Stoeva & Ivanova, 2009). Combining high yield and quality in wheat is a challenge for the contemporary breeding as it is connected with many obstacles with different nature (Baenziger et al., 2001; Dencic et al., 2007; Eagles et al., 2002; Trethowan et al., 2001). Analysis of wheat quality begins with studying the initial material (Tsenov et al., 2010c). When foreign samples are received they are study for three years for a set of quality indices. Those with high levels are included in hybridization programs especially if they show high yield potential, stress tolerance, disease resistance. The basic foreign parent components for hybridization are from breeding centers with traditions and excellent achievements in breeding for wheat quality (table 9).

The quality of Bulgarian wheat varieties is based on the widespread use of variety Bezostaya 1 and other Russian and Ukraine sources (Panayotov & Kostov, 2007; Todorov et al., 1998; Tsenov et al., 2010b). During the last years samples with different origin are widely used as they possess high level of productivity combined with high end-used quality. The quality analysis during the breeding process is made according the scheme described in the table 10.

The beginning of quality analysis of the breeding lines is in the screening nursery. The main tests in this unit are sedimentation value of the flour and grain protein content. Quality index of each line and quality standard varieties is calculated by these two parameters. These indices are defined as they possess high and positive correlation with the main parameters, connected with the strength of flour and dough (Bona et al., 2003, Dacheva & Boydjieva, 2002).

Country	Grain properties	Dough properties	Gluten properties	Bread making	End-use quality
Bulgaria	*	*	*	*	*
Romania	*		*		*
Serbia	*				*
Turkey	*		*		*
Odessa, Ukraine	*	*	*	*	*
Harkov, Ukraine	*	*		*	
Mironovka, Ukraine		*		*	*
Ktrasnodar, Russia	*	*	*		*
Nebraska, USA	*	*		*	
Oklahoma, USA	*	*		*	
Texas, USA	*	*		*	
Australia		*		*	
Canada		*		*	
Argentina	*		*		*

Table 9. Origin of the initial breeding material for wheat quality

Quality traits	Screening Nursery (SN)	Preliminary Yield Trail (PYT)	Competitive Yield Trails (CYT)
Sedimentation value	+	+	+
Grain Protein Content	+	+	+
Quality index*	+	+	+
Test weight		+	+
Farinograph characteristics			+
End-use quality			+
Bread making characteristics			+
SDS-PAGE			+

* Quality index=Sedimentation/Protein content

Table 10. Quality indices, used for breeding in the separate trial units

In the next trial units (PYT, CYT) the analysis extend and cover almost all aspects of quality of grain, flour, dough and bread. Along with this an analysis is made to determine the allelic diversity of the lines. At least three year data from CYT, which is the last level of screening of breeding materials, are needed in order to be able to assign each line in the following quality group: A – strong wheat; B – medium and C – wheat with soft endosperm (figure 4).

In the yield trails of Executive Agency of Variety Testing, Field Inspection and Seed Control (EAVTFIS) the candidate-varieties are also testing in these three groups at least 2 or 3 years. Each group has specific standards for comparison.

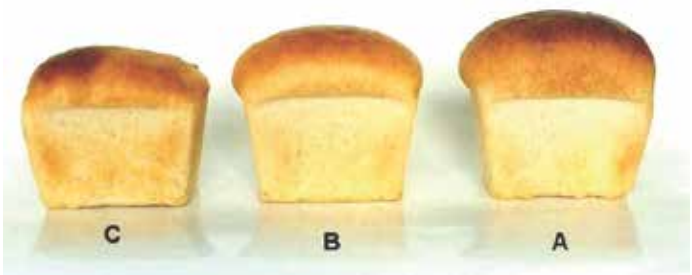


Fig. 4. Bread-making quality groups

The evaluation and selection for quality is associated with many difficulties from different nature. They can be determined as follows:

1. There is a wide genetic variation for each index caused by three genomic structure of the crop.
2. Environments have an enormous share in wheat quality formation and embarrassed the genetic expression of the varieties (Dencic et al., 2011, Tayyar, 2010).
3. Wheat quality is reduced in the presence of biotic and abiotic stress (Atanasova et al., 2010).
4. The inheritance of quality indices is complicated and polygenic (Tsenov, 1994, Tsenov et al., 1995, Tsenov & Stoeva, 1997).

5. Samples with high combining ability for quality are limited.
6. Complex genetics suggests a special selection procedure (Tsenov & Stoeva, 1998).
7. The selection in the early hybrid populations is not effective because of fading forming process and restricted amount of grain for analysis.
8. There is a lack of breeding indices for parallel selection for quality and yield potential and other parameters.

6. Conclusions

As a result of a nearly half a century breeding in Dobrudzha Agricultural Institute 26 varieties from the first quality group are created. This is nearly one third of the entire variety list. The massive usage of variety Bezostaya 1 in the breeding programs direct or indirect narrowed the genetic diversity of the materials and to some extent unified their glutenin and gliadin spectra. The efforts to include foreign initial materials in hybridization, with alleles with good influence over the wheat quality, will probably be rewarded in the near future.

A breeding progress is determined for almost all the indices, defining end-use quality in wheat. The group of high quality wheat varieties has greater progress compared to all varieties irrespective of the indices.

Different factors influence on the wheat quality and on the expression of different indices. The knowledge of the nature and direction of the effect of these factors is valuable for predicting to some extend the performance of wheat varieties in certain growing conditions. Most of the Bulgarian varieties show moderate stability at different environments. But the results for the individual varieties and their indices are not unidirectional and therefore when selecting a variety for mass production it is necessary to consider the purpose for its cultivation, the potential of the given variety and its ability to realize this potential with stability and adequacy.

7. Acknowledgment

The authors give their acknowledgments to all colleagues from the different laboratories for their help in determine different aspects of wheat quality.

8. References

Atanasova, D., Tsenov, N., Stoeva, I. & Dochev, V. 2008. Genotype x environment interaction for some quality traits of Bulgarian winter wheat varieties, In: *Modern Variety Breeding for Present and Future Needs,* J. Prohens and M. L. Badenes (Eds), Proceedings of *The 18th EUCARPIA General Congress,* 9-12 September 2008, Valencia, Spain, pp. 532-537.

Atanasova, D., Dochev, V., Tsenov, N. & Todorov, I. 2009a. Influence of genotype and environments on quality of winter wheat varieties in Northern Bulgaria. *Agricultural science and technology*, 1 (4): 121-125.

Atanasova, D., Tsenov, N., Todorov, I. & Ivanova, I. 2009b. Glutenin composition of winter wheat varieties bred in Dobrudzha Agricultural Institute. *Bulgarian Journal of Agricultural Science*, 15 (1): 9-19.

Atanasova, D., Tsenov, N., Stoeva, I. & Todorov, I. 2010. Performance of Bulgarian winter wheat varieties for main end-use quality parameters under different environments. *Bulgarian Journal of Agricultural Science*, 16 (1): 22-29.

Baenziger, P.S., Shelton, D.R., Shipman, M.J. & Graybosch, R.A. 2001. Breeding for end-use quality: Reflections of the Nebraska experience. *Euphytica*, 119 (1-2): 95-100.

Becker, H.C. & Leon, J. 1988. Stability analysis in plant breeding. *Plant Breeding*, 101: 1-23.

Bona, L., Matuz, J. & Acs, E. 2003. Correlations between screening methods and technological quality characteristics in bread wheat. *Cereal Research Communications*, 31 (1-2): 201-204.

Boyadjieva, D., Dacheva, V. & Mungova, M. 1999. A study on the rheological properties of T. aestivum germplasm from different geographic areas. Bulgarian Journal of Agricultural Science, 5: 703-712.

Bradova, J. 2008. Allelic diversity of HMW- and LMW-glutenin subunits wheat varieites (Triticum aestivum L.) registered in the Czech Republic. In: *Modern Variety Breeding for Present and Future Needs*, J. Prohens and M. L. Badenes (Eds), *Proceedings of The 18th EUCARPIA General Congress*, 9-12 September 2008, Valencia, Spain, pp. 553-557.

BSS 7971-2:2000 – Bulgarian State Standard for test weight.

BSS 13375-88 - Bulgarian State Standard for wet gluten content in grain.

BSS 16759-88 - Bulgarian State Standard for valorimeric value.

Bushehri, A.A.S., Gomarian, M.& Samadi, B. Y. 2006. The high molecular weight glutenin subunit composition in old and modern bread wheats cultivated in Iran. *Australian Journal of Agricultural Research*, 57: 1109-1114.

Dacheva, V. & Boydjieva, D. 2002. The quality index as a selection tool in winter wheat (T. aestivum L.) breeding. *Proceedings of "120 years Agricultural Science in Sadovo", Scientific Reports*, pp. 88-91. (In Bul).

Dencic, S., Kobiljsli, B., Mladenov, N., Hristov, N. & Pavlovic, M. 2007. Long-term breeding for bread making quality in wheat. In: *Wheat production in stressed environments*, Buck, H.T., Nisi, J.E., Salomyn, N. (Eds.), pp: 495-501.

Dencic, S. Mladenov, N. & Kobiljski, B. 2011. Effects of genotype and environment on bread making quality in wheat. *International Journal of Plant Production*, 5 (1): 71-81.

Drezner, G., Dvojkovic, K., Horvat, D., Novoselovic, D., Lalic, A., Babic, D. & Kovacevic, J. 2006. Grain yield and quality of winter wheat genotypes in different environments. *Cereal Research Communications*, 34 (1): 457-460.

Drezner, G., Dvojkovic, K., Horvat, D., Novoselovic, D. & Lalic, A. 2007. Environmental impacts on wheat agronomic and quality traits. *Cereal Research Communications*, 35 (2): 357-360.

Eagles, H.A., Hollamby, G.J., Eastwood, R.F. 2002. Genetic and encironmental variation for grain quality traits routinely evaluated in southern Australian wheat breeding programs. *Australian Journal of Agricultural Research*, 53: 1047-1057.

Finlay, K. W. & Wilkinson, G. 1963. The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research*, 14, 742-754.

Gomez-Beccera, H.F., Abudalieva, A., Morgunov, A., Abdullaev, K., Bekenova, L., Yessimbekova, M., Sereda, G., Shpigun, S., Tsygankov, V., Zelenskiy, Y., Pena, R.J. & Chakmak, I. 2010. Phenotypic correlations, G x E interactions and broad sense

heritability analysis of grain and flour quality characteristics in high latitude spring bread wheats from Kazahstan and Siberia. *Euphytica*, 171 (1): 23-38.

Hristov, N., Mladenov, N., Djuric, V., Kondic-Spika ,A., Marjanovic-Jeromela, A. & Simic, D. 2010. Genotype by environment interactions in wheat quality breeding programs in southeast Europe. *Euphytica*, 174 (3): 315-324.

Jackson, E.A., Holt, L.M. & Payne, P.I. 1983. Characterization of high-molecular-weight gliadin and low-molecular-weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localization of their controlling genes. *Theoretical and Applied Genetics*, 66: 29-37.

Johansson, E., Prieto-Linde, M.L. & Jonsson, J.O. 2001. Breeding for stability in bread milling quality. In: *Wheat in a Global Environment*, Z. Bedo and L. Lang (eds.) Kluwer Academic Publishers, Netherlands, 229-235.

Knoblochova, H. & Galova, Z. 2000. High molecular weight glutenin variation and relative quantization in winter spelt wheat cultivars. *Rostlinna Vyroba*, 46: 255-260.

Lawrence, G.L. & Shepherd, K.W. 1981. Inheritance of glutenin protein subunits of wheat. *Theoretical and Applied Genetics*, 60: 333-337.

Li, L., He, Z., Yan, J., Zhang, Y, Xia, X & Pena, R.J. 2005. Allelic variation at the Glu-1 and Glu-3 loci, presence of the 1B.1R translocation, and their effects on mixographic properties in Chinese bread wheats. *Euphytica*, 142(3): 197–204.

Masci, S., Rovelli, L., Kasarda, D.D., Vensel, W.H. & Lafiandra, D. 2002. Characterisation and chromosomal localisation of C-type low-molecular-weight glutenin subunits in the bread wheat cultivar Chinese Spring. *Theoretical and Applied Genetics*, 104: 422-428.

Mladenov, N., Misic, T., Przulj, N. & Hristov, N. 2001. Bread-making quality and stability of winter wheat grown in semiarid conditions. *Rostlinna vyroba*, 47: 160-166.

Panayotov, I., Boyadjieva, D., Todorov, I., Stankov, I., Dechev, D. & Tsvetkov, S. 1994. Situation and problems in wheat breeding in Bulgaria, *Plant Science*, 31 (3-4): 48-57.

Panayotov, I. & Kostov, K. 2007. Combining Bulgarian and Ukrainian breeding for improving wheat quality and productivity. *Proceedings of International Scientific Conference "Plant Genetic Stocks – The Basis of Agriculture of Today"*, 13-14 June 2007, Sadovo, Bulgaria, vol. 2-3: 371-374.

Panayotov, I. & Rachinsky, T. 2002. Wheat breeding as a basis of grain production in Bulgaria. *Proceedings of 50th Anniversary of Dobrudzha Agricultural Institute, "Breeding and Agrotechnics of field crops"*, Dobrich, June 2001, vol. 1: 21-37. (Bg)

Panayotov, I., Todorov, I., Stoeva, I. & I. Ivanova. 2004. High quality wheat cultivars created in Bulgaria during the period 1994-2004 - achievements and perspectives. *Field Crops Studies*, 1 (1): 13-19 (In Bul).

Panozzo, J.F. & Eagles, H.A. 2000. Cultivar and environmental effects on quality characters in wheat. 2. Protein. *Australian Journal of Agricultural Research*, 51, 629–636.

Pumpyanskii, A.Y. 1971. Technological characteristics of common wheat. L., pp. 22 (In Rus).

Rachinski, T. 1966. Results from testing of some Russian wheat varieties in North-West Bulgaria. *Plant Science*, 3 (10): 11-21. (In Bul)

Singh, N.K. & Shepherd, K.W. 1988. Linkage mapping of the genes controlling endosperm proteins in wheat. 1. Genes on the short arms of group 1 chromosomes. *Theoretical and Applied Genetics*, 66: 628-641.

Stoeva, I. & Ivanova, A. 2009. Interaction of the technological properties of common winter wheat varieties with some agronomy factors. *Bulgarian Journal of Agricultural Science*, 15 (5): 417-422.

Stoeva, I., Tsenov, N., & Penchev, E. 2006. Environmental impact on the quality of bread wheat varieties. *Field Crop Studies*, 3 (1): 7-17.

Sudaric, A., Simic, D. & Vrataric, M. 2006. Characterization of genotype by environment interactions in soybean breeding programmes of southeast Europe. *Plant Breeding*, 125, 191-194.

Tabasum, A., Iqbal, N., Hameed, A. & Arshad, R. 2011. Evaluation of Pakistani wheat germplasm for bread quality based on allelic variation in HMW glutenin subunits. *Pakistan Journal of Botany,* 43 (3): 1735-1740.

Tayyar, S. 2010. Variation in grain yield and quality of Romanian bread wheat varieties compared to local varieties in north-western Turkey. *Romanian Biotechnological Letters*, 15(2): 5189-5196.

Todorov I. 2006. Investigation of grain storage proteins and their use as markers in wheat breeding. *PhD Thesis*, DAI, General Toshevo, pp. 398 (In Bul).

Todorov, I., Ivanov, P. & Ivanova, I. 1998. The HMW-GS composition of wheat genotypes from Western Europe. *Proceedings of 2nd Balkan Symposium on Field crops*, Novi Sad, 16-20 June, vol. 1: 33-37.

Todorov, I., Ivanov, P.& Ivanova, I. 2006. Genetic diversity of high molecular weight glutenin alleles in varieties with different origin, *Field Crop Studies*, III (4): 487-498 (In Bul).

Tohver M. 2007. High molecular weight (HMW) glutenin subunit composition of Nordic and Middle European wheats. *Genetic Resources and Crop Evolution*, 54: 67-81.

Trethowan, R., Pena, R.J. & van Ginkel, M. 2001. The effect of indirect tests for grain quality on the grain yield and industrial quality of bread wheat. *Plant Breeding*, 120 (6): 509-512.

Tsenov, N. 1994. Combining ability of the main economic traits associated with productivity and quality in combinations of winter wheat varieties. *PhD Thesis*, pp. 185 (In Bul).

Tsenov, N., Atanasova, D., Todorov, I., Ivanova, I. & Stoeva, I. 2009. Allelic diversity in Bulgarian winter wheat varieties based on polymorphism of glutenin subunit composition, *Cereal Research Communications*, 37 (4): 551-558.

Tsenov, N., Atanasova, D., Todorov, I., Ivanova, I. & Stoeva, I. 2010a. Quality of winter common wheat advanced lines depending on allelic variation of Glu-A3. *Cereal Research Communications*, 38 (2): 250-258.

Tsenov, N., Atanasova, D., Stoeva, I. & Petrova, T. 2010b. Grain yield, end-use quality and stress resistance of winter wheat cultivars Aglika and Slaveya. *Scientific Works Agricultural University*, Plovdiv, 55 (1): 27-34.

Tsenov, N., Stoeva, I. & Atanasova, D. 2010c. Breeding of end-use quality of winter wheat in Dobrudzha Agricultural Research Institute – present and prospects, *Field Crop Studies,* 6 (2): 217-233 (In Bul).

Tsenov, N., Kostov, K., Gubatov, T. & Peeva, V. 2004. Study on the genotype x environment interaction in winter wheat varieties. I. Grain quality. *Field Crop Studies*, 1 (1), 20-29. (In Bul).

Tsenov, N. & Stoeva, I. 1997. Potentialities to combine high productivity with grain quality of winter common wheat hybrids. *Plant Science*, 34 (2): 9-13 (In Bul).

Tsenov, N. & Stoeva, I. 1998. Response to selection for improvement of grain quality in winter bread wheat crosses. *Proceedings of the 9th International Wheat Genetics Symposium*, Saskatoon, Canada, 2-7 August, Vol. 4, pp. 285-287.

Tsenov, N., Stoeva, I. & Penchev, E. 1995. Combining ability for sedimentation value in diallel crosses of winter bread wheat (Triticum aestivum). *Proceedings of "The first Balkan symposium on breeding and cultivation of wheat, sunflower and legume crops"*, 26 - 29. 06, Albena, Bulgaria: 244-247.

Williams, R. M., O'Brien, L., Eagles, H.A., Solah, V.A. & Jayasena, V. 2008. The influence of genotype, environment, and genotype x environment interaction on wheat quality. *Australian Journal of Agricultural Research*, 59 (2), 95-111.

Yan, W., M. S. Kang, B. Ma, S. Woods and P.L. Cornelius. 2007. GGEbiplot vs. AMMI analysis of genotype-by-environment data. Crop Science, 47: 643-655.

Yong, Z., Zhonghu, H., Ye, G., Aimin, Z. & van Ginkel, M. 2004. Effect of environment and genotype on bread-making quality of spring-sown spring wheat cultivars in China. *Euphytica*, 139 (1): 75-83.

Zhang Y, Zhang Y, He Z and Ye G, 2005. Milling quality and properties of autumn-sown Chinese wheats evaluated through multi-location trials. *Euphytica*, 143 (1-2), 209-222.

22

Hevea Germplasm in Vietnam: Conservation, Characterization, Evaluation and Utilization

Lai Van Lam, Tran Thanh, Le Thi Thuy Trang,
Vu Van Truong, Huynh Bao Lam and Le Mau Tuy
Rubber Research Institute of Vietnam
Ho Chi Minh City
Vietnam

1. Introduction

Germplasm collections have provided original materials for plant breeding program and crop improvement. Because of their genetic diversity and possible occurrence of particular desirable genes, germplasm collections are useful targets for plant breeders as well as other biologists. Recently, many of germplasm are being lost worldwide due to habitat destruction, invasion of foreign species, and reliance on fewer high yielding strains. Therefore, maintaining germplasm of agricultural crops is very important. Normally, a germplasm collection will be utilized based on its characters of immediate perceived value or its potential variation. Furthermore, it can also be used to better understand the properties and performances of the plants, particularly at the genomic level. Nowadays, countries all over the world have set up facilities for conservation, characterization and utilization of germplasm collections of various crops either directly or indirectly.

Rubber tree (*Hevea brasiliensis* Willd. ex A. de Juss. Müell. Arg) which produces natural rubber is a tall deciduous perennial tree belonging to the Euphorbiaceae family. *Hevea brasiliensis* is a native of the Amazon basin and was introduced to countries in the tropical belts of Asia and Africa during late 19th century. It can be termed as one of the most far reaching and successful introductions in plant history, resulting in plantations with about 10.6 million hectares in the world for providing the industry with natural rubber of 9.62 million tonnes in the year 2009 (IRSG, 2009). Approximately 78% of the cultivation has taken place in Southeast Asia, 15% in other Asia countries, 5% in Africa, and 2% in Latin America (IRSG, 2009). The main producing countries were Thailand, Indonesia, Malaysia, India and Vietnam (IRSG, 2010). As the fifth natural rubber producer in the world, Vietnam produced 723,700 tonnes that shared about 7.5% of the world's natural rubber production and total area under rubber trees were 674,200 hectares that shared 6.1% of the world's rubber area (Hoa, 2010). In Vietnam, areas under rubber trees are mainly in the South-eastern region (65.2%), followed by the Highlands (23.4%), central coastal area (9.7%) and the new areas developed in the North-western region (1.6%) (Hoa, 2010).

Hevea brasiliensis was introduced into Vietnam in 1897 from Bogor (Indonesia) by Alexandre Yersin. Since then, rubber tree has been considered as one of the most important crops and

widely cultivated throughout the country, particularly in the Southeast and Highlands regions of Vietnam. Currently, about 3,500 accessions of *Hevea brasiliensis* have been collected and conserved in Vietnam. The majority of this germplasm were derived from the IRRDB'81 germplasm collected in the Amazonian habitats of the genus. This collection has been considered as the key factor contributing to the improvement of rubber tree through breeding programs. The characterization and evaluation of the germplasm are considered to be important aspects of *Hevea* germplasm conservation. Without proper characterization and evaluation, valuable genetic variation in the collections cannot be used for rubber improvement effectively. Understanding the genetic diversity of different genetic resources of *H. brasiliensis* would be important in `order to optimize their management and to utilize *Hevea* germplasm in breeding programs.

2. *Hevea* genetic resources

The genus *Hevea* is basically composed of 10 species: *H. brasiliensis*, *H. guianensis, H. benthamiana*, *H. pauciflora*, *H. spruceana*, *H. microphylla*, *H. rigidifolia*, *H. nitida*, *H. camporum*, and *H. camargoana* (Schultes, 1990). According to Clement-Demange et al. (2000), it is generally considered that there is no biological barrier between them, and some species proved to be inter-crossable by hand-pollination; therefore, the *Hevea* species can be considered as a species complex. *H. paludosa* has been identified in Brazil by Ule in 1905 as the eleventh species (Gonçalves et al., 1990; Priyadarshan and Gonçaalves, 2002). A review on the elaborate description of taxonomical and botanical aspects of *Hevea* has been reported by Schultes (1977, 1987, 1990) and Wycherley (1992). As its natural habitat, *Hevea* species have presented in Brazil, Bolivia, Peru, Ecuador, Colombia, Venezuela, Surinam and French Guiana as shown in Figure 1. These *Hevea* species have 2n = 36 chromosomes, with the possible exception of one triploid clone of *H. guianensis* (2n = 54) and the possible existence of one genotype of *H. pauciflora* with 2n = 18 (Baldwin, 1947; Majumder, 1964), and *H. brasiliensis* behaves as an amphidiploid (Ramaer, 1935; Ong, 1975; Wycherley, 1976).

It is known that all high-yielding cultivars of rubber tree (*Hevea brasiliensis* Muell. Arg.) in the world originated from breeding programs initially developed in Southeast Asia with a very narrow genetic base. Historically, the introduction of the rubber trees into Asia began with the transfer of 70,000 seeds to England by Henry Wickham in 1876 (Wycherly, 1968). During the years of 1876-1877, a total of 2,397 *Hevea* seedlings were sent to several Asian countries such as Sri Lanka (1,919 seedlings), Bogor (Java) (18 seedlings) and Singapore (22 seedlings) (Wycherly, 1968; Dean, 1987; Baulkwill, 1989). Because of the centrally geographical position of Singapore in Asia and the influences of the British to the development of the rubber industry during these years, the collection of 22 seedlings planted in the Singaporean botanical garden became the main source of the rubber trees which was distributed to and planted in Asian countries later. Since this introduction, the rubber tree has become an important perennial crop as the major source of commercial rubber in the world (Fig. 2) and this collection was mentioned as Wickham collection (W). Since the current *Hevea* varieties all came from such a single population, it is necessary to enlarge the genetic basis for *Hevea* breeding program.

In order to enlarge the genetic basis for *Hevea*, a large collection of *H. brasiliensis* accessions from various areas in Colombia was gathered by Schultes after 1945 and then duplicated in

Ivory Coast (Nicolas, 1985). In addition, a collecting survey in the Madre de Dios basin in Peru was organized in 1948 by the Peruvian Ministry of Agriculture (Rands and Polhamus, 1955); the resulting seedlings were multiplied by grafting, and then introduced and studied in Liberia, Guatemala and Brazil under the name of MDF accessions (Bos and McIndoe, 1965). Moreover, another collecting expedition was carried out in the Brazilian states of Acre and Rondonia by both the Brazilian Agricultural Research Corporation (EMBRAPA) and French Institute for Rubber Research (IRCA) in 1974 (Hallé and Combe, 1974; Seguin et al., 2003); 42 resulting wild elite-tapped trees were collected and propagated by grafting under the name as EMBRAPA/IRCA accessions. Remarkably, in 1981, the International Rubber Research and Development Board (IRRDB) conducted an expedition covering three western states of Brazil, namely Acre (AC), Rondonia (RO), and Mato Grosso (MT), in 16 different districts and in 60 different locations overall to collect wild *Hevea* germplasm (refered as IRRDB'81 collection). As a result, a total of 63,768 seeds, 1,413 meters of budwood from 194 high yielding trees and 1,160 seedlings were collected (Tan, 1987; Simmonds, 1989; Onokpise, 2004). Of these, 12.5% and 37.5% of the seeds were sent to Malaysia and Ivory Coast, respectively, and the remaining 50% of the collections were maintained in Brazil (Clément-Demange et al., 2007). The genotypes issued from budwood collection were also then brought to Malaysia and Ivory Coast. The collection planted in Malaysia and Ivory Coast has then been distributed to all IRRDB members as clones since 1984.

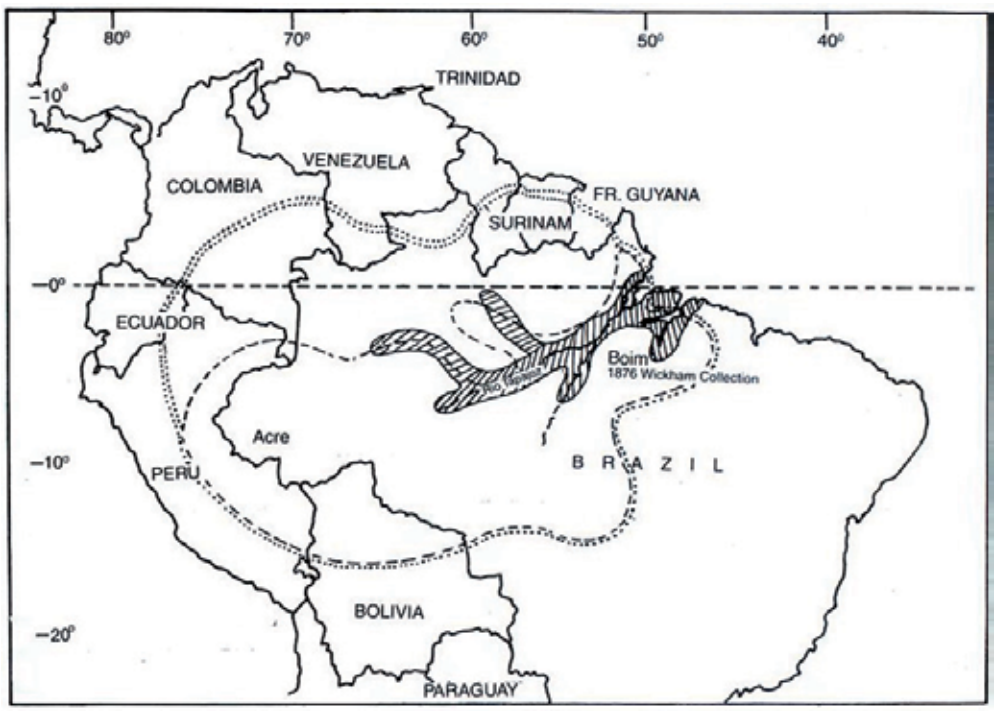


Fig. 1. Geographical origin of *Hevea* (adapted from George (2000))

Fig. 2. Distribution of rubber tree (*H. brasiliensis*) in the world where this species has been planted (adapted from Orwa et al. (2009)). Native range: Bolivia, Brazil, Colombia, Peru and Venezuela; Exotic range: Brunei, Cambodia, China, Ethiopia, India, Indonesia, Laos, Liberia, Malaysia, Myanmar, Philippines, Singapore, Sri Lanka, Thailand, Uganda, Vietnam, Guatemala, Cameroon, Ivory Coast, Ghana, Gabon, Guinea, Liberia, Nigeria, Congo, Bangladesh, Papua New Guinea, and Mexico.

3. Long-term conservation of *Hevea* germplasm

In situ and *ex situ* conservations are the two major strategies used in the conservation of plant genetic resources. *In situ* conservation, the conservation of diversity in its natural habitat, involves the designation, management and monitoring of the population at the location where it is currently found and within the community to which it belongs whereas *ex situ* conservation, collection of which the biodiversity is preserved outside its natural habitat, involves the sampling, transfer and storage of a population of a certain species away from the original location (Maxted et al., 1997). Because of a number of advantages, *the ex situ* method has been used to primarily conserve numerous plant genetic resources. In this method, plant diversity is safely preserved and concentrated in a small number of controlled places under consistent environmental conditions and is readily accessible to breeders.

An alternative solution to rubber tree *in situ* genetic conservation is the management of existing *ex situ* collections (Le Guen et al., 2009). Because of the ease of vegetative propagation by grafting, many *ex situ* collections of *Hevea* were established in various rubber producing countries. In Vietnam, the *ex situ* conservation of *Hevea* germplasm was established in 1985. This germplasm included the collection of local *Hevea* accessions, the introduction of *Hevea* clones from other countries. The majority of this germplasm collection was derived from the IRRDB'81 collection expedition in the Amazon forests of Brazil which is the primary center of diversity of the crop and the source of wild rubber trees. This germplasm was introduced into Vietnam in the form of budwood in 1984. All of the materials were sent to Lai Khe experimental station of the Rubber Research Institute of Vietnam (RRIV) in Binh Duong province, which is located in the traditional rubber growing belt. On receipt of the budwood, each accession was first multiplied by bud grafting and

then planted in the field genebank for conservation in the form of source-bush garden (Fig. 3). This source-bush garden was laid out in randomized complete block design, in which each accession was represented in two replications of five trees with a planting distance of 1.5 m x 1.2 m. The genetic resources of *Hevea* germplasm conserved in Vietnam were showed in Table 1. The source-bush garden is cut back every year to maintain the conservation and also to generate budwood for various evaluation trials. Several preliminary field evaluation trials for most of *Hevea* germplasm accessions have been being established in several representative locations to evaluate their agronomical and morphological characteristics.

Fig. 3. The source-bush garden for *ex situ* conservation of *Hevea* germplasm in Vietnam

Genetic sources	Number of accessions
South America	**3082**
Acre (AC)	959
Mato Grosso (MT)	901
Rodonia (RO)	1116
Others	106
Africa	**38**
Ivory Coast	38
Asia	**422**
Vietnam	338
Malaysia	55
Indonesia	7
Sri Lanka	16
China	5
Cambodia	1
Total	**3542**

Table 1. The genetic resources of *Hevea* germplasm conserved in Vietnam

4. Characterization and evaluation of *Hevea* germplasm

In order to use the germplasm in breeding programs, it must be characterised and evaluated. There is often a delay between collection of germplasm and its evaluation, particularly for rubber trees because of the time required for them to reach maturity. Evaluation is useful if it considers the traits wanted by plant breeders. We are fortunate that our program of germplasm acquisition and evaluation is very closely linked to our program of *Hevea* breeding with the same people usually involving in both. Preliminary evaluation can help indicate those accessions that need more detailed evaluation, but those that appear not to be of immediate values should not be simply discarded.

4.1 Agronomical characteristics

Standard characterization and evaluation of germplasm collection may be routinely performed using different methods including traditional practices such as the use of descriptive lists of morphological characters. They may also involve evaluation of the agronomical performances under various environmental conditions. Understanding the nature and the magnitude of variability of important traits existing among plant genetic materials is vital for the effective utilization of such materials for breeding purposes. In rubber tree, high latex yield is always the exclusive objective of breeding programs. To this ultimate objective, many different factors are associated. For instance, the main components of productivity are the growth of the trunk determined during the immature period before the beginning of tapping, the resistance to various diseases and the tolerance to stress factors such as high altitude, low temperature, wind damage and moisture deficit.

In Vietnam, a part of *Hevea* germplasm, especially IRRDB'81 collection, has been agronomically evaluated in various clonal trials including arboreta and small scale clonal trials at different locations. In the view of latex production, IRRDB'81 collection exhibited very poor performance with an average latex yield of around 16% of the level of the currently developed Wickham clones after 5 years of tapping. This result was similar to that of other studies conducted on IRRDB'81 collections in Malaysia, Indonesia, Ivory Coast and China (Ramli et al., 2004; Aidi et al., 2002; Clément-Demange et al., 2002; Hu et al., 2002). Outstandingly, in the first three years of tapping, some IRRDB'81 accessions such as AC56/276, AC62/54, MT8/27, MT/I/2 and RO62/26 produced 30.0 – 45.0 gram/tree/tapping, or 102.7 – 153.8% of the production of the control clone (GT1). These accessions have been being used as parents in *Hevea* breeding programs. Considering geographical origins, accessions derived from Mato Grosso seemed to be better yielder than those from Acre and Rondonia. This might indicate their better adaptability to the experimental areas where the climate featured by a distinct dry spell of six months is similar to that of the original region which is known to have a dry spell of three to four months annually (Chevallier, 1988). A large number of IRRDB'81 accessions showed very good growth performance with girth at opening ranging from 59.9 – 74.0 cm, or 124.8 – 133.5% of that of the control clone. Remarkably, some IRRDB'81 accessions such as MT29/68, RO24/58 and RO32/104 had significantly higher girth than the control when planted in the highland area which was considered as a non-traditional rubber region in Vietnam. The superiority of these accessions will be of considerable value in advanced breeding programs.

Recently, IRRDB'81 collection has been considered as an important source for timber selection and rubber wood production. The average wood volume of the IRRDB'81

collection conserved in Vietnam was initially recorded, yielding 1.48 m^3/tree at the age of 21 years, 43.4% higher than the Wickham population. Among these, several accessions appeared to be the best with clear bole volume ranging from 1.55 to 2.07 m^3/tree, which could be considered as suitable clones for timber production purposes. Similarly, several high timber yield accessions of the IRRDB'81 collection in Indonesia and Malaysia were also reported. At the age of 13 years, a total of 28 accessions in Indonesia and 20 accessions in Malaysia were selected for timber yield with clear bole volume at a range of 0.90 to 2.56 m^3/tree (Aidi et al., 2002) and 1.0 to 1.6 m^3/tree (Ramli et al., 2004), respectively.

It is known that unlike other clonally multiplied species, *Hevea* is not affected by viral diseases (Simmonds, 1989). Other diseases which are considered as economic importance are *Gloeosporium* leaf disease (*Colletotrichum gloeosporioides* Pen. Sacc.), pink disease (*Corticium salmonicolor* Berk. & Br.), powdery mildew (*Oidium heveae* Stein.), *Corynespora* leaf fall (*Corynespora cassiicola* Berk. & Curt. Wei.), *Phytophthora* leaf fall (*Phytophthora* sp.) and SALB (South American Leaf Blight - *Microcyclus ulei* P. Henn von Arx.). Among these diseases, *Corynespora* leaf fall and SALB are the most important in rubber plantations. Wycherly (1969) noted that the clonal and location specificity was evident towards resistance to these diseases. Differences in the level of resistance to important diseases have been observed in IRRDB'81 collections from different countries. In India, a large number of IRRDB'81 accessions were showed to be resistant to powdery mildew (140 accessions) and *Corynespora* leaf fall (70 accessions) (Varghese et al., 2002). Similarly, a total of 21 accessions in IRRDB'81 collection are resistant to powdery mildew in bush-wood garden in China (Huang et al., 2002). The resistance to SALB was observed on 298 accessions from *ex situ* germplasm collection planted in both French Guyana and Brazil, of which the accessions from Acre and Rondonia were the most resistant (Le Guen et al., 2002). In Vietnam, it seemed that Wickham and IRRDB'81 accessions showed no significant difference in susceptibility to pink disease, powdery mildew and *Gloeosporium* leaf diseases (Lam et al., 2002). In general, evaluations of IRRDB'81 collection for biotic and abiotic stresses are in progress in major rubber growing countries.

4.2 Genetic diversity based on biochemical and molecular markers

In conventional plant breeding, many morphological traits have been used as markers for genetic analyses and cultivar identification, but specific genetic information on Mendelian traits are rare in *Hevea*. In contrast to traditional practices, genetic characterization refers to the description of the attributes that follow Mendelian inheritance or involve specific DNA sequences. In this way, biochemical assays, which detect differences between isozymes, or molecular markers were applied to determine the genetic diversity of the germplasm. In addition, the development of molecular and biochemical markers help researchers not only to identify genotypes, but also to assess and exploit the genetic variability (Whitkus et al., 1994). Insights into the relative genetic diversity among *Hevea* collections would be useful in *Hevea* breeding as well as *ex situ* conservation of *Hevea* genetic resources. The commercial value associated with identifying useful traits, especially yield and growth, would create a direct value in genebanks, ensuring long-term preservation of a collection. Moreover, *Hevea* germplasm characterization using molecular and biochemical markers will contribute to the knowledge of genetic relationships not only among wild accessions but also between accessions of wild and cultivated gene pool, and hence help to facilitate the breeding programs. In Vietnam, significant progress has been made in evaluation and

characterization of *Hevea* germplasm by applying genetic markers. Among several efficient methods to reveal the genetic variability within and among plant populations, the most widely applied methods are isozyme electrophoresis and random amplified DNA polymorphism (RAPD). Both markers are useful to analyze genetic diversity of *Hevea* germpasm, and to select good *Hevea* clones for future breeding or cultivation purposes.

4.2.1 Isozymes marker

As first described by Hunter and Markert (1957), isozymes were defined as the different molecular forms in which proteins may exist with the same enzymatic specificity (Buth, 1984). This means that different variants on the same enzymes have identical or similar functions and are present in the same individual. Isozyme had played a minor role in research on plant biochemistry until genetic polymorphism for isozymes within the same population was discovered in 1966 (Stebbins, 1989; Wendel, 1989). In the 1980s, analysis of isozymes was developed at CIRAD (French Agricultural Research Centre for International Development) with 13 polymorphic isozymic systems to formulate a diagnostic kit associated with a clonal identification database. This kit has proved to be able to differentiate a large set of cultivated clones (Leconte et al., 1994). Since then, isozymes have been used as genetic markers for identification of rubber tree cultivars, genetic diversity analysis, controlling progenies issued from hand pollination and reproductive biology (Chevallier, 1988; Leconte et al., 1994; Paiva et al., 1994; Sunderasan et al., 1994). The first study on using isozyme markers for a precise understanding of the genetic diversity of the wild *Hevea* germplasm was carried out using ten isozyme markers on a set of 263 accessions from the IRRDB'81 collection (Chevallier, 1988).

Isozyme markers were firstly used in Vietnam to identify and confirm rubber clones in budwood gardens. Moreover, it was also used in *Hevea* breeding for hybrid genealogical legitimacy and genetic diversity research of *Hevea* germplasm. A total of 12 isozyme systems were used in studying genetic diversity of the IRRDB'81 *Hevea* germplasm in Vietnam. Banding patterns of representative IRRDB'81 accessions generated by isozyme electrophoresis are shown in Figure 4.

The study was performed on both IRRDB'81 collection and Wickham population with 117 accessions from 15 districts of the states of Acre, Rondonia and Mato Grosso of Brazil and 24 Wickham's clones (Fig. 5 and Table 2). The quantity of accessions sampled for each district was more or less proportional to the quantity of accessions currently conserved for the district. The result of isozymatic analysis showed that a total of 60 alleles were detected (Table 2). Out of them, 60 alleles were detected in IRRDB'81 accessions and 26 alleles in Wickham population. The result showed that the alleles detected in Wickham population were also detected in the IRRDB'81 collection, many new alleles were found in the IRRDB'81 collection only, thus underlining the genetic enrichment provided by the wild *Hevea* collections. Among IRRDB'81 collection, Acre accessions had the largest number of polymorphic alleles (51/60 alleles), followed by the Rodonia (45/60 alleles) and Mato Grosso accessions (45/60 alleles). This result revealed that Mato Grosso was obviously less polymorphic than Rondonia because Mato Grosso and Rondonia groups had the same number of detected polymorphic alleles but the volume of samples of the Mato Grosso groups was bigger than that of the Rondonia groups. The result also showed the significant polymorphic differences among the districts. The number of alleles detected in the districts Brasileia, Taurauca of Acre and district Cartriquacu of Mato Grosso was high in spite of the

small volume of the samples whereas the number of alleles detected in districts Itanba and Aracatuba of Mato Grosso was quite low.

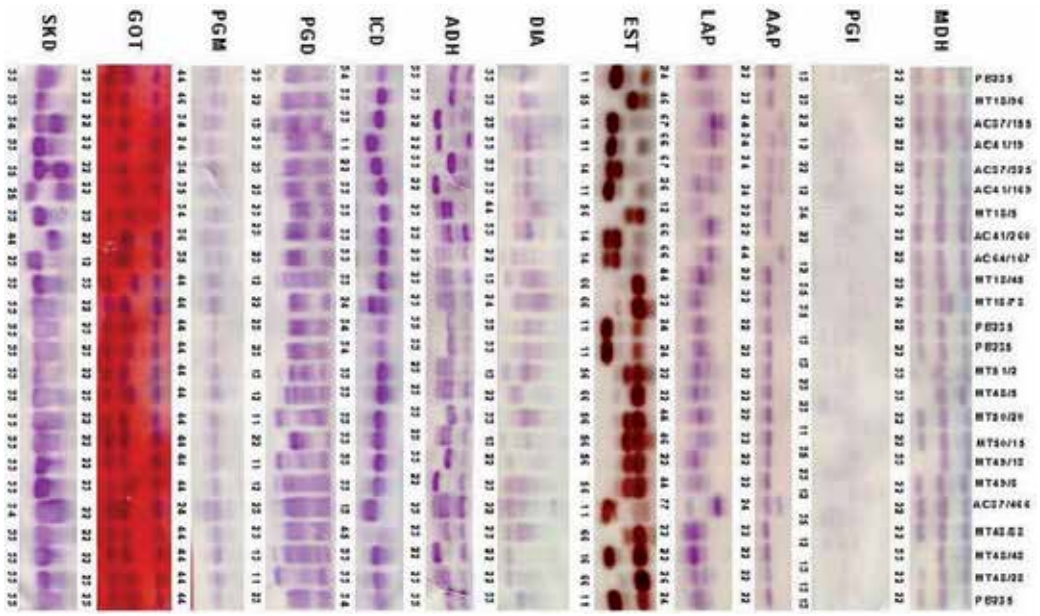


Fig. 4. Banding patterns of representative *Hevea* accessions generated by isozyme electrophoresis. Twelve isozyme systems used in studying genetic diversity of the IRRDB'81 *Hevea* germplasm had very different banding patterns. MDH, malate dehydrogenase; PGI, phospho glucose isomerase; AAP, alanyl amino peptidase; LAP, leucine amino peptidase; EST, esterase; DIA, diaphorase; ADH, alcohol dehydrogenase; ICD, isocitrate dehydrogenase; PGD, phosphogluconase dehydrogenase; PGM, phosphoglucomutase; GOT, glutamate oxaloacetate transaminase; SKD, shikimate acid

The study affirmed that the genetic base of *Hevea* germplasm in Vietnam, especially IRRDB'81 collection, was prosperous and diversified. Because of the high genetic variability level, this collection would contribute effectively to the long term progress of *Hevea* breeding and selection program in the country. In contrast, the Wickham population showed a low level of genetic variability, which is the consequence of the oriented selection through many years in a narrow geographical origin.

The result of cluster analysis based on isozymes database revealed the noticeable relationship between genetic clusters (Fig. 6). According to the genetic distance between the accessions, Acre accessions and Rondonia accessions were close to each other, meanwhile the genetic distance between Acre or Rondonia accessions were far from Mato Grosso accessions, except that the accessions from Vila Bella district of Mato Grosso (MT/VB) were close to those of Rondonia. Among IRRDB'81 accessions, Mato Grosso population was relatively close to Wickham collection based on isozyme analysis. In general, the genetic distance between the accessions conformed to the geographical origins of *Hevea*. However, several accessions of Arce and Rondonia were not separated distinctly using isozyme electrophoresis although they were distributed widely according to the geographical origins.

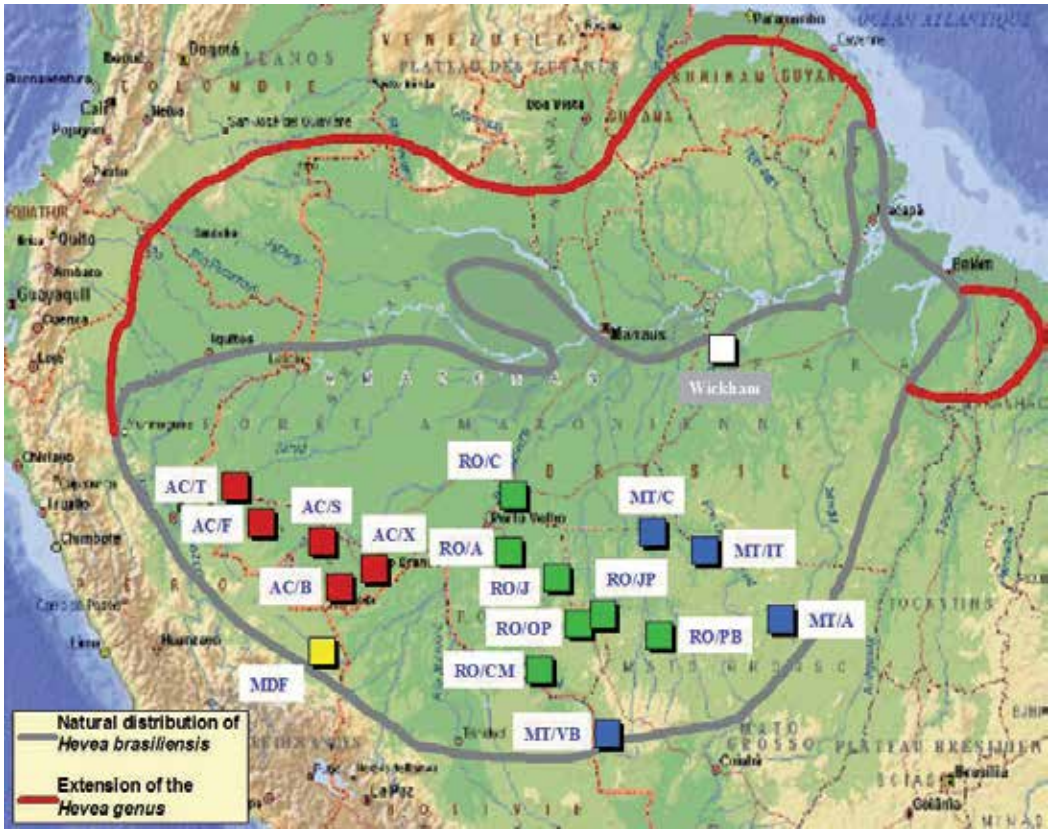


Fig. 5. Geographical origins of *Hevea* IRRDB'81 collection

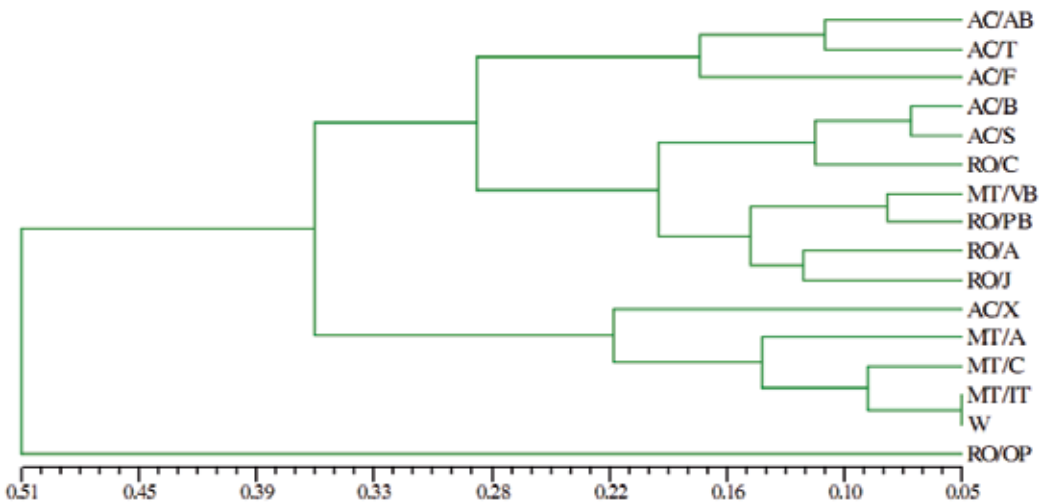


Fig. 6. Dendrogram based on Nei's genetic distance between IRRDB'81 collection and Wickham population.

Genetic resources	Number of accessions	Number of alleles/locus*												Total of alleles
		MDH	PGI	AAP	LAP	EST	DIA	ADH	ICD	PGD	PGM	GOT	SKD	
IRRDB'81	**117**	**4**	**5**	**5**	**7**	**7**	**5**	**2**	**5**	**3**	**6**	**4**	**7**	**60**
Acre	**58**	**4**	**5**	**3**	**6**	**6**	**5**	**2**	**4**	**3**	**6**	**3**	**4**	**51**
Assis-brasil	2	1	2	2	3	2	1	2	2	2	2	1	3	23
Brasileia	9	3	3	2	3	4	3	2	4	3	5	2	3	37
Feijo	15	3	3	2	4	3	4	2	4	3	4	2	4	38
Sena Madureira	21	2	5	2	4	4	3	2	4	3	6	3	4	42
Tarauaca	8	3	3	2	3	2	2	2	3	2	4	1	3	30
Xapuri	3	1	2	1	3	2	2	2	1	3	4	3	2	26
Mato Grosso	**32**	**3**	**5**	**1**	**6**	**5**	**4**	**2**	**4**	**3**	**4**	**3**	**5**	**45**
Aracatuba	9	2	4	1	3	3	3	2	3	3	1	2	1	28
Cartriquacu	7	2	5	1	5	5	4	2	3	3	3	3	1	37
Itanba	13	2	3	1	4	3	2	2	3	2	2	3	2	29
Vila Bela	3	1	3	1	1	3	1	1	2	2	3	1	4	23
Rondonia	**27**	**3**	**5**	**4**	**4**	**4**	**2**	**2**	**4**	**2**	**5**	**3**	**7**	**45**
Ariquemes	5	1	2	3	2	3	2	2	2	2	5	2	4	30
Calama	18	2	5	1	3	4	2	2	4	2	5	3	6	39
Jaru	2	1	1	1	1	2	1	2	2	2	2	1	3	19
Ouro Preto	1	1	1	2	1	2	1	1	2	1	2	1	2	17
Pimenta Bueno	1	2	2	1	1	2	1	1	1	2	2	1	2	18
WICKHAM	**24**	**2**	**3**	**1**	**3**	**3**	**1**	**2**	**4**	**3**	**1**	**2**	**1**	**26**
TOTAL	**141**	**4**	**5**	**5**	**7**	**7**	**5**	**2**	**5**	**3**	**6**	**4**	**7**	**60**

* MDH, malate dehydrogenase; PGI, phospho glucose isomerase; AAP, alanyl amino peptidase; LAP, leucine amino peptidase; EST, esterase; DIA, diaphorase; ADH, alcohol dehydrogenase; ICD, isocitrate dehydrogenase; PGD, phosphogluconase dehydrogenase; PGM, phosphoglucomutase; GOT, glutamate oxaloacetate transaminase; SKD, shikimate acid

Table 2. Genetic variability of *Hevea* germpalsm based on isozymes

The results of isozymes analysis indicated that the *Hevea* germplasm conserved in Vietnam is very diversed. This characterisation would help to utilize the new genetic resources more effectively in *Hevea* breeding programs. The combination of morphological characterization and isozyme markers could help breeders to constitute a core collection of *Hevea* IRRDB'81 germplasm to ensure the conservation of the genetic variability. In addition, maintaining the genetic variability in *Hevea* germplasm would help to reduce gene erosion. Moreover, isozyme markers could be used as an assistant tool to orient a long term plan to advance heterosis to improve Wickham materials based on the recombination between Wickham clones and IRRDB'81 accessions. However, isozyme-based analysis is limited by the rather small number of marker loci available and a general lack of polymorphism for these loci. In addition, the analysis has to be carried out near the field sites owing to the fragility of the isozymes to varied temperatures or otherwise the samples need to be freeze-dried and transported to the laboratory. In spite of such limitations, isozyme was still a helpful marker to evaluate the genetic variability of the *Hevea* germplasm in Vietnam.

4.2.2 RAPD marker

The random amplified polymorphic DNA (RAPD) technique, first described by Williams et al. (1990), despite some limitations, has provided a useful approach for evaluating population's genetic differentiation, particularly in species that are poorly genetically known (Silva and Russo, 2000; Nybom, 2004). Recently, a large number of studies have pointed out that DNA-based markers, such as RAPD, were superior to isozymes in detecting genetic diversity (Garkava et al., 2000; Matos et al., 2001; Ochiai et al., 2001; Sharma et al., 2008). It is known that isozymes represent allelic expression of the same locus, while DNA fragments produced by RAPD are independent genetic markers (Ochiai et al., 2001) with a lower proportion of non-neutral markers than formerly (Bartish et al., 2000). Hence, isozyme and RAPD analyses often give discordant patterns, suggesting the importance of using multiple molecular marker systems in studies of population structure (Wendel and Doyle, 1998; Bartish et al., 2000; Lebot et al., 2003). RAPD marker was also used to evaluate the levels of gene flow between species (Arnold et al., 1991) and detection of gene introgression in various plant species (Waugh et al., 1992; Orozco-Castillo et al., 1994; Gomez et al., 1996). In rubber tree, RAPD has become a useful maker for investigating genetic diversity within and between *Hevea* populations, especially the IRRDB'81 collection (Varghese et al., 1997; Venkatachalam et al., 2002; Lam et al., 2009). Moreover, this marker was also used to identify a dwarf genome-specific marker (Venkatachalam et al., 2004) or certain homology to proline-specific permease gene (Venkatachalam et al., 2006) in rubber tree. The accumulated data on *Hevea* RAPD analysis from different accessions give information on genetic relations and *Hevea* origin, and provide the initial basis for clonal distinction and germplasm evaluation of agronomical interest. Therefore, the data can also be used in *Hevea* improvement programs.

In Vietnam, RAPD was firstly used to study genetic diversity of *Hevea* germplasm (Lam et al., 2009). The study was performed on IRRDB'81 collection with 59 accessions from 13 districts of the states of Acre, Rondonia and Mato Grosso of Brazil (Fig. 5). Using 6 oligonucleotide primers, the percentage of polymorphic loci calculated for individual districts ranged from 15.38% in Assis-Brasil district to 70.77% in Sena Madureira district of Acre, which totally had 10 and 46 polymorphic banding patterns, respectively (Table 3, Fig. 7). Although the sample sizes might have certain effects on the extent of the polymorphism of various districts, in the cases of Assis-Brasil of Acre, Ariquemes of Rondonia, and Vila Bela of Mato Grosso, they were quite different in the extent of polymorphism with the same sample sizes. In addition, the Jiparana district of Rondonia was very polymorphic regardless of its small sample size.

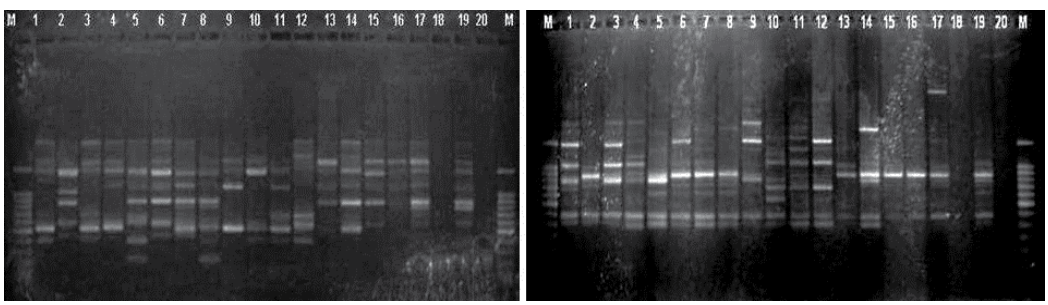


Fig. 7. DNA fingerprints of representative IRRDB'81 accessions generated by primer A18 (left) and OPB-12 (right)

State/District	No. of accessions	No. of total/polymorphic band patterns	Observed number of alleles	Mean heterozygosity	Shannon index	Genetic distance*
Acre	**25**					
Assis-brasil	2	41/10	1.154	0.064	0.093	0.139
Brasileia	4	52/37	1.569	0.223	0.328	0.299
Feijo	6	56/42	1.646	0.240	0.355	0.282
Sena Madureira	9	58/46	1.708	0.246	0.368	0.256
Tarauaca	4	56/32	1.492	0.196	0.287	0.225
Mato Grosso	**21**					
Aracatuba	6	51/38	1.585	0.227	0.333	0.264
Cartriquacu	5	51/38	1.585	0.222	0.328	0.306
Itanba	8	56/42	1.646	0.231	0.344	0.264
Vila Bela	2	46/21	1.323	0.134	0.195	0.296
Rondonia	**13**					
Ariquemes	2	45/16	1.246	0.102	0.149	0.216
Calama	4	53/36	1.554	0.217	0.319	0.290
Jaru	4	51/39	1.600	0.239	0.348	0.323
Jiparana	3	54/44	1.677	0.264	0.389	0.469
Total/Mean	**59**	**65/62**	**1.522**	**0.200**	**0.295**	**0.279**

* Genetic distance is mean genetic distance among accessions in individual district

Table 3. Patterns of genetic diversity of *Hevea* IRRDB'81 collection based on RAPD analysis

In general, the IRRBD'81 collection conserved in Vietnam showed the high level of genetic diversity detected by RAPD. In fact, the mean values of heterozygosity or Nei's genetic diversity (Nei, 1978) within individual districts varied from 0.064 to 0.264 over 65 loci with the average of 0.2 across the districts (Table 3). This parameter differed substantially among 13 districts studied. The Jiparana district of Rondonia showed the highest estimated heterozygosity over 62 polymorphic loci, whereas the Assis-Brasil district of Acre showed the lowest one. In spite of high standard errors of mean heterozygosities, probably due to the small sample sizes, recorded in Assis-Brasil (Acre), Ariquemes (Rondonia), and Vila Bela (Mato Grosso) populations, the remarkable variations of mean heterozygosity clearly showed differences in genetic variability among 13 districts. The average degree of diversity within individual districts using Shannon's diversity index (Shannon and Weaver, 1949) was 0.296 and ranged from 0.093 for the Assis-Brasil district of Acre to 0.389 for the Jiparana district of Rondonia (Table 3). Shannon index was correlated strongly with the percentage of polymorphic loci in a district. In fact, the districts with high Shannon's diversity index also exhibited the high percentages of polymorphic loci (Table 3). Similar to mean heterozygosity, differences in values of Shannon's diversity index also showed genetic differentiation among the districts. The largest average genetic distance among accessions within the districts was detected in the Jiparana district of Rondonia and the smallest was found in the Assis-Brasil district of Acre, of which the average genetic distance values were 0.469 and 0.139, respectively (Table 3). According to previous studies, high genetic diversity is usual in IRRDB'81 accessions (Chevallier, 1988; Besse et al., 1994; Lekawipat et al., 2003). All the accessions had unique RAPD genotypes. Nei's genetic distance values between pairs of districts ranged from 0.046 for Catriquacu and Itanba of Mato Grosso to 0.304 for Tarauaca of Acre and Aracatuba of Mato Grosso (Table 4). The dendrogram constructed by UPGMA cluster analysis showed that *Hevea* IRRDB'81 collection of 13 different districts were in five clusters with Ariquemes of Rondonia quite different from the others (Fig. 8).

Among the districts of Mato Grosso, Vila Bela was in the same cluster with the districts from Rondonia (Jaru and Jiparana), meanwhile other districts (Itanba, Catriquacu and Aracatuba) were grouped into one cluster. This showed that Vila Bela was quite different from other districts of Mato Grosso, which was also noted by other researchers using RFLP markers (Besse et al., 1994). The distribution of those districts in genetic cluster analysis seemed to conform to geographical origins of *Hevea* IRRDB'81 collection, except Calama district of Rondonia.

District	AC/AB	AC/B	AC/F	AC/S	AC/T	MT/A	MT/C	MT/IT	MT/VB	RO/A	RO/C	RO/J	RO/JP
AC/AB		0.147	0.201	0.146	0.160	0.297	0.244	0.236	0.202	0.286	0.188	0.236	0.266
AC/B	0.147		0.106	0.107	0.155	0.190	0.135	0.161	0.216	0.275	0.135	0.160	0.164
AC/F	0.201	0.106		0.105	0.145	0.125	0.115	0.129	0.213	0.230	0.107	0.144	0.150
AC/S	0.146	0.107	0.105		0.190	0.153	0.151	0.146	0.158	0.239	0.103	0.164	0.169
AC/T	0.160	0.155	0.145	0.190		0.304	0.227	0.234	0.193	0.236	0.230	0.205	0.212
MT/A	0.297	0.190	0.125	0.153	0.304		0.094	0.086	0.296	0.295	0.171	0.191	0.183
MT/C	0.244	0.135	0.115	0.151	0.227	0.094		0.046	0.202	0.286	0.125	0.145	0.141
MT/IT	0.236	0.161	0.129	0.146	0.234	0.086	0.046		0.189	0.289	0.147	0.145	0.147
MT/VB	0.202	0.216	0.213	0.158	0.193	0.296	0.202	0.189		0.233	0.123	0.146	0.149
RO/A	0.286	0.275	0.230	0.239	0.236	0.295	0.286	0.289	0.233		0.222	0.189	0.205
RO/C	0.188	0.135	0.107	0.103	0.230	0.171	0.125	0.147	0.123	0.222		0.137	0.172
RO/J	0.236	0.160	0.144	0.164	0.205	0.191	0.145	0.145	0.146	0.189	0.137		0.097
RO/JP	0.266	0.164	0.150	0.169	0.212	0.183	0.141	0.147	0.149	0.205	0.172	0.097	

Table 4. Nei's genetic distance among 13 districts of *Hevea* IRRDB'81 collection

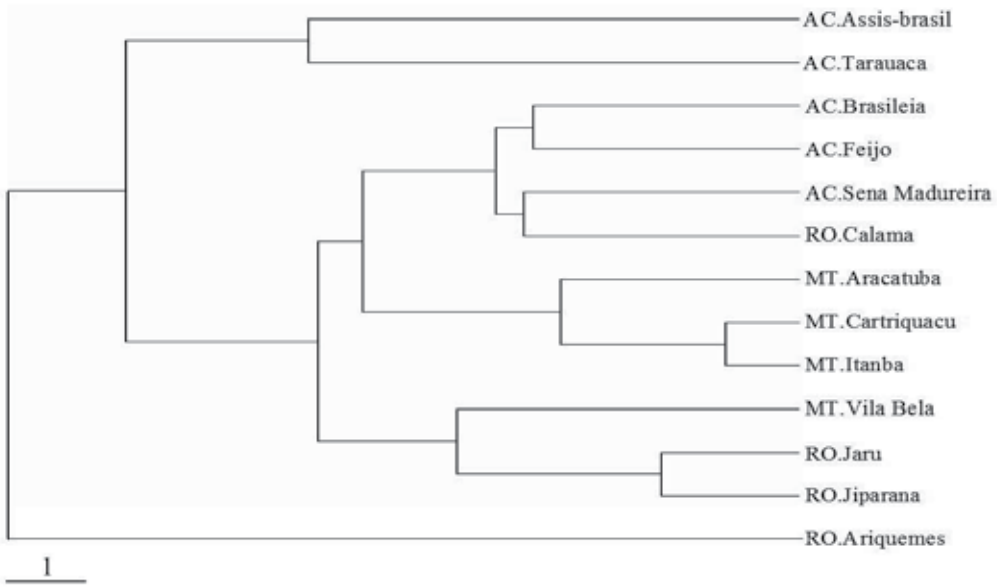


Fig. 8. Dendrogram of UPGMA cluster analysis based on Nei's genetic distance

Analysis of molecular variance (AMOVA) for 59 accessions of 13 districts revealed that the interpopulation (interdistrict) variation accounted for only 14.1% of the total genetic variance while intrapopulation (within district) variation accounted for 85.9% (Table 6). The results indicated that the majority of genetic diversity was within district variation (85.9%). The results suggests that there might be a certain gene flow among the districts, possibly owing to the species' outcrossing as a mode of reproduction and dispersion of seeds by a network of rivers in Amazon basin. However, it should be treated with caution due to small samples brought into the study.

Source of variation	df	SS	CV	%Total
Among districts	12	198.6	1.57	14.1
Within districts	46	439.6	9.56	85.9

Significant at P=0.001

Table 5. Summary of the AMOVA analysis

The IRRDB'81 *Hevea* germplasm in Vietnam exhibited large genetic variability by RAPD markers. The intradistrict source accounted for most of the genetic variation of the germplasm. Based on the genetic distance analysis, the collection could be classified into five groups which could help in planning crosses based on genetic distance in the hope of looking for heterosis and maintaining the abundant genetic diversity. The RAPD markers could also help in checking the genetic variability of the *Hevea* breeding program. Moreover, in association with the agronomical characteristics, morphological traits and isozymes analysis, RAPD markers are now suitable tools for genetic diversity studies of *Hevea* germplasm and can be useful for accumulation and management of genetic-breeding resources of *Hevea brasiliensis*.

5. Utilization of *Hevea* germplasm in Vietnam

One of the major objectives of conservation of *Hevea* genetic resources is to make genetic diversity available for immediate or future use. The widest possible range of the genetic diversity has to be conserved in order to meet future, as yet unknown, needs. *Hevea* germplasm conservation program in Vietnam is expected to promote and facilitate the use of conserved materials through the maintenance of healthy and readily accessible and adequately characterised/evaluated materials, and proper documentation of the relevant information. Evaluation data of the agronomic performances and the morphological characteristics gathered during cultivation of the accessionshave been being recorded continuously using a specifically constructed program (Fig. 9). Currently, a total of about 3,500 *Hevea* accessions have been collected and *ex situ* conserved. This germplasm comprises of three main genetic resources: the Amazonian (A) (most of which belong to IRRDB'81 collection collected in the Amazonian habitats of the genus), the Wickham (W), and the Wickham x Amazonian (WA) resources. The majority of this germplasm were derived from the IRRDB'81 collection with a total of 2,999 accessions, each of which is a clone derived from originally collected seedlings. Most of them have been evaluated for the agronomical and morphological characteristics. In the view of the limitations of the agronomical and morphological traits, isozyme and RAPD markers were used to analyze the genetic diversity and structure of the IRRDB'81 collection for more effective utilization of the germplasm in *Hevea* breeding programs in Vietnam.

Since the IRRDB'81 collection exhibited very poor profiles in agronomical characteristics, especially latex productivity, the chance for direct use of this collection for latex purposes seemed to be impractical although certain accessions could be planted for timber purposes. Regarding widening the genetic base for genetic improvement, several promising IRRDB'81 accessions have been included in hand pollination program in RRIV since 1997. Based on the agronomical and morphological traits as well as the genetic diversity analysis, recently, many attempts have been made to enlarge the genetic base of *Hevea* breeding materials by polycrossing among different genetic resources. In this way, many crosses between maternal

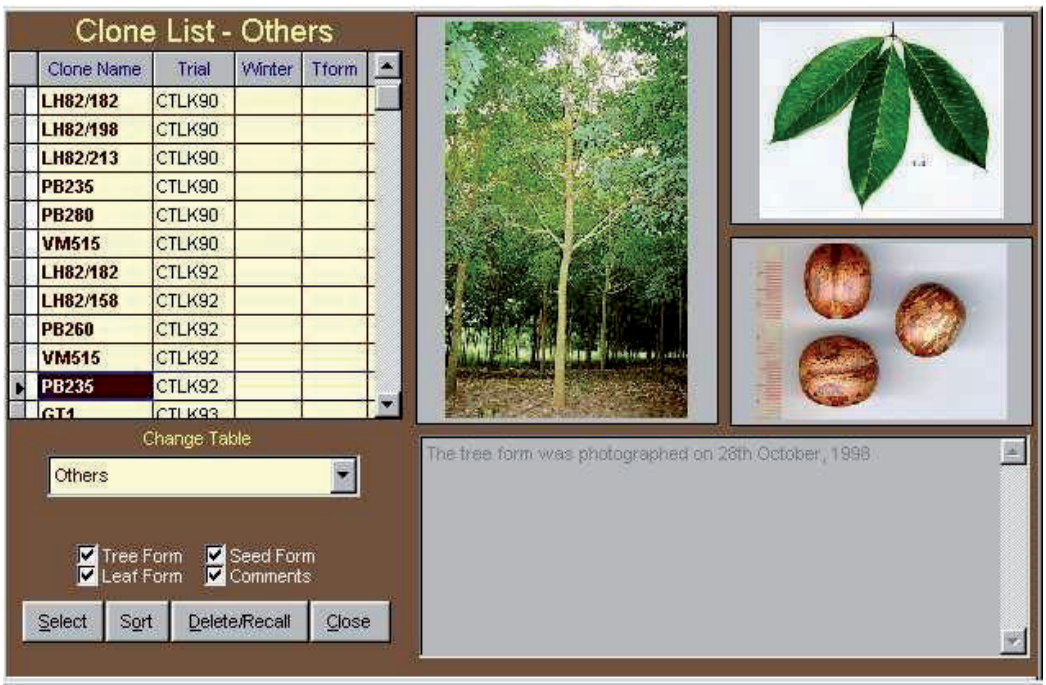


Fig. 9. Genebank documentation program in conservation of *Hevea* germplasm

W and paternal A accessions have been preferably made and the progenies of which are in various phases of evaluation with the expectation that the W x A progenies could combine the good agronomical characteristics from parents while optimizing the genetic variability in this population. In fact, Amazonian accessions were crossed with Wickham high yielding and good set fruit clones such as PB260 and RRIC110, and progenies derived from these hand pollination crosses were disbudded into fields of early selection trials for evaluation of agronomical performances such as latex yield, growth and diseases incidence. In general, these progenies exhibited rather good in girth but very poor in latex productivity. Most of the progenies produced a very little or negligible amount of latex; this agreed closely with the previous finding in Ivory Coast for the progenies from W x A crosses (Clement-Demange et al, 1990). However, some progenies had the production of 1.9 – 3.2 gram/tree/tapping or 106 – 132% of the production of the control clone (i.e. PB 260), and a large number of the progenies showed very good in growth performance with girth at 34 months old after planting ranging from 15.6 – 21.2 cm, or 101 – 161% of that of the control clone. These progenies are being further tested in the small scale clonal trials

and would be included as parents in future breeding programs. This gives a way for the opportunity of genetic improvement in breeding programs, particularly in latex productivity of the *Hevea* IRRDB'81 collection. In diseases incidence, all of the progenies exhibited varying degrees of infection to powdery mildew. The progenies derived from different paternal accessions showed significant differences in susceptibility to powdery mildew. For instance, the progenies derived from AC6/23 and AC35/114 were more susceptible to powdery mildew than other progenies. In contrast, the progenies derived from RO44/268 and RO44/71 seemed to be lightly susceptible to powdery mildew in comparison to other progenies. This result has contributed to the development of clonal disease resistance by genetic recombination using IRRDB'81 collection as paternal clones in *Hevea* breeding programs.

The results obtained so far can be considered as a basis to continue combining Wickham and IRRDB'81 genetic resources in breeding programs at RRIV. The W x A progeny population provides a valuable source for selecting multi-clones recommended for developing rubber cultivation in non-traditional regions in Vietnam.

6. Problems and challenges

6.1 Problems

Because the genetic structure of natural *H. brasiliensis* populations is based upon both hydrographical network patterns and long-range isolation by distance, it is likely that the known genetic diversity represents only part of its true natural diversity (Le Guen et al., 2009). In order to enlarge the genetic resources of *H. brasiliensis*, it is necessary to conduct additional expeditions in other areas that were not yet covered previously, such as Amazonian basin in Colombia, Peru and Bolivia, and the Brazilian states of Pará and Amazonas. Besides, since a very small part of the diversity of *Hevea* germplasm has been collected and conserved in Vietnam, a much greater diversity of the germplasm should be imported from other countries such as Brazil, Malaysia and Ivory Coast. Moreover, it is urgent need to duplicate the *Hevea* germplasm accessions, particularly the IRRDB'81 collection, in all IRRDB country members to prevent the loss of accessions and to increase the genetic resources as raw materials in *Hevea* breeding programs. Additionally, molecular tools may contribute in *ex situ* conservation of *Hevea* germplasm to the sampling, management and development of "core" collections as well as the utilization of genetic diversity. However, the use of such molecular tools is limited in the rubber growing countries including Vietnam due to their cost.

Rubber tree is traditionally propagated through bud grafting on unselected seedlings, which maintains intraclonal heterogeneity for vigour and productivity. Therefore, variation among a bud-grafted population is significant and can influence the growth and productivity levels. Therefore, a great improvement may be expected by using *in vitro* micropropagation. *In vitro* techniques have currently made a commercial impact in rubber, and their propagation systems can circumvent the influence of the stock-scion interactions in *Hevea* clones (Priyadarshan, 2007). Recently, there has been an increasing interest in the induction of somatic embryogenesis in rubber trees. However, successful somatic embryo formation and plant regeneration have been reported by a few researchers in different countries using limited genotypes of *Hevea*. In addition, the frequency of somatic embryo induction was

found to be very low and non-synchronous, its germination remains very difficult and thus *Hevea* embryogenic system needs further investigation. There has been no large scale commercial application of tissue culture techniques for mass propagation of clonal *Hevea* as yet, either by microcutting or by somatic embryogenesis. However, there is sufficient progress at the research level to suggest that tissue culture of *Hevea* can and should be further developed.

6.2 Challenges

At the scientific and technical levels, challenges are posed by genetic erosion, genetic vulnerability and utilization. Genetic erosion is defined as "the loss of genetic diversity, including the loss of individual genes, and the loss of particular combinations of genes (i.e. of gene-complexes) such as those manifested in locally adapted landraces" (FAO, 1997). There is no consensus on the optimal balance of *in situ* and *ex situ* conservation methods to combat genetic erosion (Fraleigh, 2006). A gradual erosion of the genetic variability of *Hevea* in all of natural rubber plantations was realized in the 1970s. This erosion occurred because most of *Hevea* clones in cultivation were derived from the few surviving seeds collected by Henry Wickham. Therefore, widening the genetic base of *Hevea* in production was seen as a prerequisite to generate new clones with new and valuable characteristics with regard to productivity, disease resistance, and tolerance to many environmental conditions.

Genetic vulnerability was described as "the condition that results when a widely planted crop is uniformly susceptible to a pest, pathogen or environmental hazard as a result of its genetic constitution, thereby creating a potential for widespread crop losses" (FAO, 1996). It is known that genetic vulnerability pertains to the level of the crop genetic diversity actually being used. Because of the very narrow genetic base in the commercially cultivated *Hevea* clones, the commercial rubber cultivation, due to their genetic vulnerability, is under a constant threat of attack by native as well as exotic diseases and insect pests. The changes in the weather parameters due to the increasing trend in climate change have further complicated the above issues. Climate change, which is clearly felt in the traditional rubber growing regions of Vietnam, may possibly alter the host-pathogen interactions. This will lead to the emergence of otherwise minor disease, and *Corynespora* leaf fall disease may represents this scenario. This pathogen is rapidly progressing into new areas, thus highlighting the need for stronger and advanced resistance breeding approaches.

It is known that the potential uses and values of *Hevea* genetic resources need to be understood by characterizing, evaluating and documenting them. Methods still need to be developed to improve and facilitate productive utilization. Although biotechnological methods are now increasingly available to facilitate productive utilization of *Hevea* germplasm, not all countries have the capacity to use such new technologies.

Another set of challenges is posed for taking action. For instance, how the necessary cooperation can be organized among countries and among disciplines, particularly in order to link the conservation and the use of *Hevea* genetic resources, and how the resources which need to address these issues can be mobilized. Besides, cryogenic preservation of endangered seedling trees is yet another important aspect to be looked into urgently.

7. Conclusions and prospects

Vietnam has received a large share of the *Hevea* germplasm collection of the 1981 IRRDB expedition. A total of 2,999 wild accessions belonging to three states of Brazil *viz*., Acre, Rondonia and Mato Grosso are being conserved. Systematic efforts are underway for conservation, characterization, utilization and documentation of these valuable genetic resources. In general, the genetic base of *Hevea* in Vietnam which is prosperous and diversified has been contributing effectively to the long term progress of *Hevea* breeding program in the country. As management of the germplasm is a herculean task, IRRDB'81 accessions have been evaluated in a phased manner. Studies in different sets of this germplasm have been carried out since 1985 onward to access the extent of variability present in the collection for various agronomical characteristics such as latex yield, girth, wood volume, diseases incidence, biotic and abiotic stresses, in order to exploit them in the improvement programs. A large number of IRRDB'81 accessions are now in various evaluation stages. The evaluated IRRDB'81 collection displayed unimproved characteristics of a wild population and was far inferior to the Wickham clones in agronomical performances, especially latex productivity. However, this germplasm had a much broader genetic variability which can help in broadening the narrow genetic base of cultivated rubber and also in developing location specific rubber clones for cultivation in the marginal and non-traditional rubber regions of the country.

The genetic parameters obtained from isozyme and RAPD analyses indicated that the *Hevea* germplasm conserved in Vietnam exhibited large genetic diversity. The biochemical and molecular markers have also shown to be the effective techniques for breeders to manipulate the *Hevea* germplasm. These markers could be used to select the parents with far genetic distance aiming at enlarging the genetic variability in their progenies and could help in checking the genetic variability of the *Hevea* breeding program. Furthermore, information on the structure of genetic diversity could help establish global *Hevea* collections for long-term conservation with minimum maintenance activity, and help define working collections for medium-term targeted utilization and breeding purposes.

The conventional solution to the conservation of *Hevea* genetic resources has been the establishment of *ex situ* genebanks. *Ex situ* conservation is the only effective method for the long-term conservation of *Hevea* germplasm. *Ex situ* conservation may also represent a last resort for many species and varieties including *Hevea* that would otherwise die out as their habitats are destroyed. Moreover, the management of existing *ex situ Hevea* collections is an alternative solution to rubber trees *in situ* genetic preservation (Le Guen et al., 2009). The main benefit of *ex situ* conservation is to provide breeders with ready access to a wide range of genetic materials with useful traits. *Hevea* germplasms have their own share of problems. Although only a small proportion of *Hevea* genetic resources are actually used by breeders, partly because of the inadequate characterization of the accessions, the costs of characterizing, evaluating and cataloguing genetic resource materials need to be carefully considered. Because of the severe limitations faced by large germplasm collection, establishing a core collection of this germplasm is necessary to facilitate a speedy and more efficient evaluation, and to get a collection which is conserved better and exploited more effectively.

8. References

Aidi, D.; Sekar, W. & Irwan, S. (2002). Report on The Evaluation and Utilization of The 1981 IRRDB *Hevea* Germplasm in Indonesia, *IRRDB Joint Workshop on Breeding, Agronomy and Socioeconomy,* Malaysia & Indonesia, 28 August - 7 September, 2002

Arnold, M.L.; Buckner, C.M. & Robinson, J.J. (1991). Pollen-Mediated Introgression and Hybrid Speciation in Louisiana Irises. *Proc. Natl. Acad. Sci. USA*, 88, 1398-1402

Baldwin, J.J.T. (1947). *Hevea*: A First Interpretation. A Cytogenetic Survey of a Controversial Genus, With a Discussion of Its Implications to Taxonomy and to Rubber Production. *Journal of Heredity*, 38, 54-64, ISSN 0022-1503

Bartish, I.V.; Garkava, L.P., Rumpunen, K. & Nybom, H. (2000). Phylognetic Relationship and Differentiation Among and Within Populations of *Chaenomeles* Lindl. (Rosaceae) Estimated with RAPDs and Isozymes. *Theor. Appl. Genet.*, 101, 554–563

Baulkwill, W.J. (1989). The History of Natural Rubber Production. In: *Rubber*, C.C. Webster & W.J. Baulkwill (Ed.), 1–56, Longman Scientific and Technical, ISBN 0-470-40405-3, Essex, UK

Besse, P.; Seguin, M., Lebrun, P., Chevallier, M. H., Nicholas, D., & Lanaud, C. (1994). Genetic Diversity among Wild and Cultivated Populations of *Hevea brasiliensis* Assessed by Nuclear RFLP Analysis. *Theor. Appl. Genet.*, 88, 199–207, ISSN 0040-5752

Bos, H. & McIndoe, K.G. (1965). Breeding of *Hevea* for Resistance Against *Dothidella ulei* P. Henn. *J. Rubber Res. Inst. Malaya*, 19, 98–107

Buth, D.G. (1984). The Application of Electrophoretic Data in Systematic Studies. *Ann. Rev. Ecol. Syst.*, 15, 501–522, ISSN 0066-4162

Chevallier, M.H. (1988). Genetic Variability of *Hevea brasiliensis* Germplasm Using Isozyme Markers. *J. Nat. Rubb. Res.*, 3, 42-53, ISSN 0127-7065

Clément-Demange, A.; Chapuset, T. & Seguin, M. (2002). Report on The IRRDB 1981 *Hevea* Germplasm by CIRAD (France) Year 2002. *IRRDB Joint Workshop on Breeding, Agronomy and Socioeconomy,* Malaysia & Indonesia, 28 August - 7 September, 2002

Clément-Demange, A.; Legnate, H., Seguin, M., Carron, M.P., Le Guen, V., Chapuset, T. & Nicolas, D. (2000). Rubber Tree. In: *Tropical Plant breeding*, A. Charrier, M. Jacquot, S. Hamon & D. Nicolas (Ed.), 455-480, Collection Repères, ISBN 978-1-57808-144-8, Montpellier, France.

Clément-Demange, A.; Priyadarshan, P.M., Hoa, T.T.T. & Venkatachalam, P. (2007). *Hevea* Rubber Breeding and Genetics. In: *Plant Breeding Review*, 29, J. Janick (Ed.), 177-281, John Wiley & Sons, Inc., ISBN 9780470052419, New York, USA

Dean, W. (1987). *Brazil and the Struggle for Rubber*, Cambridge Univ Press, ISBN 0-521-33477-2, New York, USA

FAO. (1996). First Report on The State of The World's Plant Genetic Resources for Food and Agriculture. FAO, Rome, Italia, Available from http://www.fao.org/WAICENT/FAOINFO/AGRICULT/AGP/AGPS/Pgrfa/pdf/swrshr_e.pdf

FAO. (1997). First Report on The State of The World's Plant Genetic Resources for Food and Agriculture. FAO, Rome, Italia, Available from http://www.fao.org/WAICENT/FAOINFO/AGRICULT/AGP/AGPS/pgrfa/pdf/swrfull.pdf

Fraleigh, B. (2006) Global Overview of Crop Genetic Resources. In: *The Role of Biotechnology in Exploring and Protecting Agricultural Genetic Resources*, J. Ruane & A. Sonnino (Ed.), 21-32, FAO, ISBN 92-5-105480-0, Roma, Italia

Garkava, L.P.; Rumpunen, K. & Bartish, I.V. (2000). Genetic Relationships in *Chaenomeles* (Rosaceae) Revealed by Isozyme Analysis. *Sci. Hortic*, 85, 21–35

George, P.J. (2000). Germplasm Resources. In: *Natural rubber: Agromanagement and Crop Processing*, P.J. George & C.K. Jacob (Ed.), 47-58, Rubber Research Institute of India, Kottayam, India

Gomez, R.; Angel, F., Bonierbale, M.W., Rodriguez, F., Tohme, J. & Roca, W.M. (1996). Inheritance of Random Amplified Polymorphic DNA Markers in Cassava (*Manihot esculenta* Crantz). *Genome*, 39, 1039-1043

Gonçalves, P. de S.; Cardoso M. & Ortolani A.A. (1990). Origin, variability and domestication of *Heve*a. *Pesquisa Agropecuária Brasileira*, 25, 135-156, ISSN 0100-204X

Hoa, T.T.T. (2010). Overview of Vietnam Rubber Industry and Prospects of Development by 2020, *Proceedings of the 9thAsian Workshop on Polymer Processing in Vietnam*, pp. xxiii-xxvii, Hanoi, Vietnam, December 7-10, 2010

Hu, Y.; Zeng, X., Chen, H., Fang, J. & Huang, H. (2005). A Study Report on The Main Characters Evaluation of New Amazon *Hevea* Germplasm, *International Natural Rubber Conference,* pp. 50-54, Cochin, India, November 6-8, 2005

Huang, H.; Yu, D. & Zhou, J. (2002). Studies of The 1981 IRRDB *Hevea* Germplasm in China, *IRRDB Joint Workshop on Breeding, Agronomy and Socioeconomy,* Malaysia & Indonesia, 28 August - 7 September, 2002

Hunter, R.L. & Markert, C.L. (1957). Histochemical Demonstration of Enzymes Separated by Zone Electrophoresis in Starch Gels. *Science*, 125, 1294-1295

IRSG. (2009). *Latest Rubber Statistical Bulletin*. International Rubber Study Group, London, October – December, 2009, Available from http://www.rubberstudy.com/default.aspx?aspxerrorpath=/newsarticle.aspx.

IRSG. (2010). *Latest Rubber Statistical Bulletin*. International Rubber Study Group, London, April – June, 2010, Available from http://www.rubberstudy.com/default.aspx?aspxerrorpath=/newsarticle.aspx

Lam, L.V.; Lam, H.B., Hoa, T.T.T. & Trang, L. T.T. (2002). Status Report of The IRRDB'81 *Hevea* Germplasm in Vietnam. *IRRDB Joint Workshop on Breeding, Agronomy and Socioeconomy,* Malaysia & Indonesia, 28 August - 7 September, 2002

Lam, L.V.; Thanh, T., Chi, V.T.Q., Tuy, L.M. (2009). Genetic Diversity of *Hevea* IRRDB'81 Collection Assessed by RAPD Markers. *Mol. Biotechnol.*, 42, 292–298

Le Guen, V. ; Garcia, D., Mattos, C.R.R. & Clément-Demange, A. (2002). Evaluation of Field Resistance to *Microcyclus ulei* of A Collection of Amazonian Rubber Tree *(Hevea brasiliensis)* Germplasm. *Crop Breeding and Applied Biotechnology*, 2, 141-148, ISSN 1518-7853

Le Guen, V.; Doaré, F., Weber, C. & Seguin, M. (2009). Genetic Structure of Amazonian Populations of *Hevea brasiliensis* is Shaped by Hydrographical Network and Isolation by Distance. *Tree Genetics & Genomes*, 5, 673-683, ISSN 1614-2942

Lebot, V.; Herail, C., Gunua, T., Pardales, J., Prana, M., Thongjiem, M. & Viet, N. (2003). Isozyme and RAPD Variation Among *Phytophthora colocasiae* Isolates from South East Asia and The Pacific. *Plant Pathol.*, 52, 303–313

Leconte, A.; Lebrun, P., Nicolas, D. & Seguin, M. (1994). Electrophoresis Application to *Hevea* Clone Identification. *Plant Research and Dévelopment*, 1, 28-36

Lekawipat, N., Teerawatanasuk, K., Rodier-Goud, M., Seguin, M., Vanavichit, A., Toojinda, T. & Tragoonrung, S. (2003). Genetic Diversity Analysis of Wild Germplasm and

Cultivated Clones of *Hevea brasiliensis* Muell. Arg. by Using Microsatellite Marker. *J. Rubber Res.*, 6, 36–47, ISSN: 1511-1768

Majumder, S.K. (1964). Chromosome Studies of Some Species of *Hevea*. *J. Rubb. Res. Inst. Malaysia*, 18, 269-273.

Matos, M.; Pinto-Carnide, O. & Benito, C. (2001). Phylognetic Relationships Among Portuguese Rye Based on Isozyme, RAPD and ISSR Markers. *Hereditas*, 134, 229–236

Maxted, N.; Ford-Lloyd, B.V. & Hawkes, J.G. (1997). Complementary Conservation Strategies. In: *Plant Genetic Conservation: The In Situ Approach*, N. Maxted, B.V. Ford-Lloyd & J.G. Hawkes (Ed.), 20-55, Chapman & Hall, ISBN 0-412-63730-8, London, UK

Nei, M. (1978). Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics*, 89, 583–590.

Nicolas, D. (1985). Acquisition of *Hevea* Material Derived from Colombian Schultes Collections, *International Rubber Conference*, Kuala Lumpur, Malaysia, October 18-19, 1985

Nybom, H. (2004). Comparison of Different Nuclear DNA Markers for Estimating Intraspecific Genetic Diversity in Plants. *Molecular Ecology*, 13, 1143–1155

Ochiai, T.; Nguyen, V.X., Tahara, M. & Yoshino, H. (2001). Geographical Differentiation of Asian Taro, *Colocasia esculanta* (L.) Schott, Detected by RAPD and Isozyme Analyses. *Euphytica*, 122, 219–234

Ong, S.H. (1975). Chromosome Morphology at the Pachytene Stage in *Hevea brasiliensis:* A Preliminary Report, *Proceedings of the International Rubber Conference*, pp. 3-12, Kuala Lumpur, Malaysia

Onokpise, O.U. (2004). Natural Rubber, *Hevea brasiliensis* (Willd. ex A. Juss.) Müll. Arg., Germplasm Collection in The Amazon Basin, Brazil: A Retrospective. *Econ. Bot.*, 58, 544–555, ISSN 0013-0001

Orozco-Castillo, C.; Chalmers, K.J., Waugh, R. & Powell, W. (1994). Detection of Genetic Diversity and Selective Gene Introgression in Coffee Using RAPD Markers. *Theor. Appl. Gent.*, 87, 934-940

Orwa, C.; Mutua, A., Kindt, R., Jamnadass, R. & Simons, A. (2009). *Agroforestree Database: A Tree Reference and Selection Guide version 4.0*, Available from http://www.worldagroforestry.org/af/treedb/

Paiva, J.R.; Kageyama, P.Y., Vencorsky, R. & Contel, P.B. (1994). Genetics of Rubber Tree *Hevea brasiliensis* (Willd. ex Adr. de Juss.) Muell. Arg. 1. Genetic Variation in Natural Populations. *Silvae Genetica*, 43, 307-312

Priyadarshan, P.M. & Gonçalves P. de S. (2002). Use of *Hevea* gene pool in rubber tree (*Hevea brasiliensis* Muell. Arg) breeding. *The Planter*, 78, 123-138

Priyadarshan, P.M. (2007). Rubber. In: *Genome Mapping and Molecular Breeding in Plant: Tropical Crops*, C. Kole (Ed.)., 143-174, Springer-Verlag, ISBN 978-3-540-34537-4, New York, USA

Raemer, H. (1935). Cytology of *Hevea*. *Genetics*, 17, 193, ISSN 0016-6731

Ramli, O.; Masahuling, B., Nasaruddin M.M.A. & Zarawi, A.G. (2004). IRRDB 1981 Expedition: Harnessing Genetic Potential of *Hevea* Germplasm, *Proceedings of IRRDB Conference*, pp. 11-25, Kunming, China, September 7-8, 2004

Rands, R.D. & Polhamus, L.G. (1955). *Progress Report on The Cooperative Hevea Rubber Development in Latin America*, U.S. Dept. of Agriculture, Washington, USA

Schultes, R.E. (1977). A New Infrageneric Classification of *Hevea*. *Botanical Museum Leaflets of Harvard University*, 25, 243–257

Schultes, R.E. (1987). Studies in the Genus *Hevea* VIII. Notes on intrageneric variants of *Hevea brasiliensis* (Euphorbiaceae). *Econ. Bot.*, 41, 125-147

Schultes, R.E. (1990). A Brief Taxonomic View of the Genus *Hevea*. *MRRDB Monograph* 14, ISSN 0127-7782, Kuala Lumpur, Malaysia.

Seguin, M.; Flori, A., Legnate, H., Clément-Demange, A. (2003). Rubber Tree. In: *Genetic diversity of tropical crops*, P. Hamon, M. Seguin, X. Perrier & J.C Glaszmann JC (Ed.), 277–305, Science Publisher, Enfield, USA

Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*, University of Illinois Press, ISBN 0-252-72546-8, Illinois, USA

Sharma, K.; Mishra, A.K. & Misra, R.S. (2008). A Simple and Efficient Method for Extraction of Genomic DNA from Tropical Tuber Crops. *Afr. J. Biotechnol.* 7, 1018–1022

Silva, E.P. & Russo, C.A.M. (2000). Techniques and Statistical Data Analysis in Molecular Population Genetics. *Hydrobiologia*, 420, 119–135

Simmonds, N.W. (1989). Rubber Breeding. In: *Rubber*, C.C. Webster & W.J. Baulkwill (Ed.), 85-124, Longman Scientific and Technical, ISBN 0-470-40405-3, Essex, UK

Stebbins, G.L. (1989). Introduction. In: *Isozymes in Plant Biology*, D.E. Soltis & P.S. Soltis (Ed.), 1-4, Dioscorides Press, Oregon, USA

Sunderasan, E.; Wickneswari, R., Aziz, M.Z.A. & Yeang, H.Y. (1994). Incidence of Self- and Cross-Pollination in Two *Hevea brasiliensis* Clones. J. Nat. Rubber Res., 9, 253-257, ISSN 0127-7065

Tan, H. (1987). Strategies in Rubber Tree Breeding. In: *Improving Vegetatively Propagated Crops*, A.J. Abbott & R.K. Atkin. (Ed.), 28-54, Academic Press, ISBN 0-120-41410-4, London, UK.

Varghese, Y.A.; Abraham, S.T., Mercy, M.A., Madhavan, J., Reghu, C.P., Rao, G.P., Ammal, S.L., Idicula, S.P. & Joseph, A. (2002). Management of The 1981 Germplasm Collection in India, *IRRDB Joint Workshop on Breeding, Agronomy and Socioeconomy,* Malaysia & Indonesia, 28 August - 7 September, 2002

Varghese, Y.A.; Knaak, C., Sethuraj, M.R. & Ecke, W. (1997). Evaluation of Random Amplified Polymorphic DNA (RAPD) Markers in *Hevea brasilinesis*. *Plant Breeding*, 116, 47-52

Venkatachalam, P.; Priya, P., Gireesh, T., Saraswathyamma, C.K. & Thulaseedharan, A. (2006b). Molecular Cloning and Sequencing of a Polymorphic Band from Rubber Tree [*Hevea brasiliensis* (Muell.) Arg.]: The Nucleotide Sequence Revealed Partial Homology with Proline-Specific Permease Gene Sequence. *Current Science*, 90, 1510-1515

Venkatachalam, P.; Priya, P., Saraswathyamma, C.K. & Thulaseedharan, A. (2004). Identification, Cloning and Sequence Analysis of a Dwarf Genome-Specific RAPD Marker in Rubber Tree *Hevea brasiliensis* Muell. Arg. *Plant Cell Reports*, 23, 327-332

Venkatachalam, P.; Thomas, S., Priya, P., Thanseem, I., Gireesh, T., Saraswathyamma, C.K. & Thulaseedharan, A. (2002). Identification of DNA Polymorphism with The Cultivated Clones of Rubber Tree (*Hevea brasiliensis* Muell. Arg.). *Indian J. Nat. Rubber Res.*, 15, 72-181, ISSN: 0970-2431

Waugh, R.; Baird, E. & Powell, W. (1992). The Use of RAPD Markers for The Detection of Gene Introgression in Potato. *Plant Cell Rep*., 11, 466-469

Wendel, J.F. & Doyle, J.J. (1998). Phylogenetic Incongruence: Window into The Genome History and Molecular Evolution. In: *Molecular Systematics of Plants II. DNA Sequenceing*, D.E. Soltis, P.S. Soltis & J.J. Doyle (Ed.), 265–296, Kluwer Academic Publishing, ISBN 0-412-11121-7, Dordrecht, The Netherlands

Wendel, L.F. (1989). Visualization and Interpretation of Plant Isozymes. In: *Isozymes in Plant Biology*, E.D. Soltis & P.S. Soltis (Ed.), 5–45, Dioscorides Press, Oregon, USA

Whitkus, R.; Doebley, J., Wendel, J.F. (1994). Nuclear DNA Markers in Systematics and Evolution. In: *DNA Based Markers in Plants*, L. Phillips & I.K. Vasil (Ed.), 116–141, Kluwer Academic Publishers, ISBN 0-792-36865-7, Dordrecht, The Netherlands

Williams, J.G.K.; Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. (1990). DNA Polymorphisms Amplified by Arbitrary Primers Are Useful As Genetic Markers. *Nucleic Acids Research*, 18, 6531–6535

Wycherley, P.R. (1992). The Genus *Hevea*: Botanical Aspects. In: *Natural Rubber: Biology, Cultivation and Technology*, M.R. Sethuraj & N.M. Mathew (Ed.), 50–66, Elsevier, ISBN 0-444-88329-0, Amsterdam, The Netherlands

Wycherly, P.R. (1969). Breeding of *Hevea. J. Rubber Res. Inst. Malaya*, 21, 38–55

23

Characterisation of the Amaranth Genetic Resources in the Czech Gene Bank

Dagmar Janovská[1], Petra Hlásná Čepková[2] and Mária Džunková[3]
[1]*Department of Gene Bank, Crop Research Institute*
[2]*Department of Crop Sciences and Agroforestry in Tropics and Subtropics*
Institute of Tropics and Subtropics, Czech University of Life Sciences Prague
[3]*Department of Genomics and Health,*
Centre for Public Health Research (CSISP), Valencia
[1,2]*Czech Republic*
[3]*Spain*

1. Introduction

The human species depends on plants. These constitute the basis for food, supply most of our needs (including clothes and shelter) and are used in industry for manufacturing fuels, medicines, fibres, rubber and other products. However, the number of plants that humans use for food is minimal, compared to the number of species existing in nature. Only 30 crops, the most outstanding of which are rice, wheat and maize, provide 95% of the calories needed in the human diet (Jaramillo & Baena, 2002). However, agricultural biodiversity is in sharp decline due to the effects of modernisation, such as concentration on a few competitive species and changes in diets. Since the beginning of agriculture, the world's farmers have developed roughly 10 000 plant species for use in food and fodder production. Today, only 150 crops feed most of the world's population, and just 12 crops provide 80% of dietary energy from plants, with rice, wheat, maize and potato providing 60%. It is estimated that about three quarters of the genetic diversity found in agricultural crops have been lost over the past century, and this genetic erosion continues (EC, 2007).

Humans need to add to their diet those crops of high yield and quality that can adapt to environmental conditions and resist pests and diseases. Advantage must be taken of native and exotic species, with nutritional or industrial potential, or new varieties must be developed. Improving crops, however, requires reserves of genetic materials whose conservation, management and use have barely begun to receive the attention that they deserve. Humans take advantage of plant genetic resources in as much as they are useful to us, which means that we must understand them, and know how to manage, maintain and use them rationally (Jaramillo & Baena, 2002). Information on genetic diversity and relationships within and among crop species and their wild relatives is essential for the efficient utilization of plant genetic resource collections for the efficient explanation of taxonomic relationships (Chan & Sun, 1997; Drzewiecki et al., 2003).

Amaranthus L. is a genus from *Amaranthaceae* family probably originated in America. This genus contains approximately 70 species of worldwide distribution including pigweeds,

waterhemps, and grain amaranths (Sauer, 1967). The origin of various species of cultivated amaranths is not easy to trace because wild ancestors are pantropical cosmopolitan weeds (Espitia-Rangel, 1994). For human consumption there are cultivated grain amaranths – *A.caudatus, A. cruentus* and *A. hypochondriacus* and vegetable amaranths – mainly *A.dubius, A.tricolor* and *A.cruentus.* Grain amaranths are crop species of New World origin; *A. caudatus* from Andean Peru and Ecuador, *A. cruentus* and *A. hypochondriacus* from Mexico and Central America (Sauer, 1950; Drzewiecki, 2001). Nowadays, the grain amaranths are cultivated from the temperate to tropical zone and the vegetable amaranths mainly in the South Africa and South Asia (Jarošova et al., 1997).

Amaranths are very promising crops. The main reasons could be content of protein, fat and active substances. The content of seed protein is in the range 13 – 18% with very good balanced amino acids. The lysine content is relatively high in the comparison with common cereals. The content of crude proteins in leaves is from 27 to 49% in d.m. what is more than in the leaves in the spinach (Segura-Nieto, 1994). Amaranths have comparable or higher amounts of essential amino acids as whole egg protein (Drzewiecki et al., 2003). The fat content is in the range 0.8-8.0%. The linoleic acid is the predominant fatty acid, with lesser amount of oleic and palmitic acids. The oil also contains squalene, precursor of cholesterol, which is used in the cosmetics and as a penetrant and lubricant (Becker, 1994). Many compounds and extracts from amaranths possessed anti-diabetic, anti-hyperlipidemic, spermatogenic and anti-cholesterolemic effects (Sangameswaran & Jayakar, 2008; Girija et al., 2011), antioxidant and antimicrobial activity (Alvarez-Jubete et al., 2010; Tironi&Anon, 2010). Many consumers purchase amaranth because they want a wheat- and gluten-free product, like the nutritional profile of amaranth, or enjoy "exotic" foods in their diet (Brenner et al., 2000). Amaranth can be used also as a feed for pigs, hens, etc. (Pisarikova et al., 2005). From the cultivation point of view, amaranth is interesting for its heat and drought resistance and very low susceptibility to diseases and pests (Barba de la Rosa, 2009). Considering its agronomic importance, attention should be given to the cultivation, conservation, and sustainable utilization of this promising crop (Ray & Roy, 2009).

Unfortunately, amaranths are also very harmful weeds spread in all over the world. Weedy *Amaranthus* species (pigweeds) have been and continue to be a major problem in agronomic production. The weed amaranth *A. retroflexus* is considered one of the world's worst weeds. A major contributor to the noxious nature of these weedy species is their ability to efficiently adapt to the changes in agricultural management practices that are specifically designed to control and prevent colonization. For example, numerous populations of pigweeds have evolved herbicide resistance (Drzewiecki, 2001; Rayburn et al., 2005).

In the Czech Republic the cultivation of amaranth was introduced in the early 1990s (Michalova 1999; Moudry et al. 1999) and the collection of amaranth genetic resources was established in 1993 in the Czech Gene Bank. Due to the very positive effects on the human health, we try to find out genotypes suitable for the Czech conditions with utilization in the Czech cuisine. On the Czech market, there is very popular food made from amaranth flour such as chips, cookies, and breakfast cereals, etc. However, all amaranth seeds are imported into the Czech Republic from other countries. The demand for vegetable amaranth is also increasing. Presently, in the Gene Bank, there are stored 103 evaluated accessions. In the working collection (in the different stages of evaluations), there are more than 30 accessions. Seed samples of amaranth are obtained from other gene banks, universities, private subjects

or from collecting missions from all over the world. It corresponds with international agreements and with The Czech National Programme on Conservation and Utilization of Plant Genetic Resources and Agro-biodiversity. For maintenance and utilization of plant genetic resources of amaranths, it is very important to know them from all sides. Genetic resources studies are oriented on evaluation of the most important biological characters, with respect to the effective utilization of genetic resources in breeding and agricultural practice. Good characterization and evaluation of genetic resources under conditions similar to those of their origin can provide breeders and users with valuable information on effective utilization of genetic resources for the breeding programmes and utilization. Characterization of genetic resources is focused mainly on morphological characters. The evaluation consists of data on plant growth and development, characteristics of plant stand, analysis of yield elements, etc. (Dotlačil et al., 2001). First steps of evaluations after seed samples receiving, are field evaluations. The phenological and morphological evaluation such as length of vegetation, plant height, length of inflorescence, colour of inflorescence, type of inflorescence, etc., is performed during vegetation. The length of vegetation is very important for amaranth cultivation in the Czech Republic, because many of the amaranths genotypes are sensitive to day-length. They remain in the vegetative period for a long time and create seeds after day-shortening (NRC, 1984). In the Czech Republic, they flower in the second half of September. Because the early frost, they cannot mature their seeds.

For genetic improvement of *Amaranthus*, germplasm collections will play a key role as well. However, only limited information is available on intra- and inter-specific genetic diversity and relationships within *Amaranthus* germplasm collections (Chan & Sun, 1997). In spite of the fact that it has been the object of many studies, the genus *Amaranthus* is still poorly understood, being widely considered as a "difficult" genus. Currently, the taxonomic problems are far from being clarified especially because of the widespread nomenclatural disorder caused chiefly by repeated misapplication of names (Costea et al., 2001) which is shown in Table 1. Due to variation of morphological characters, accurate classification of amaranth genetic resources is not always possible (Transue et al., 1994).

For preliminary identification of *Amaranthus* species, the useful tool can be the number, thickness, orientation and density of branches in inflorescences. The flowers are arranged in small and very contracted cymes, which are agglomerated, axillary and additionally arranged in racemose or spiciform terminal, large and complex synflorescences. Although extremely variable, there is usually a tendency towards a morphological "type" (Costea et al., 2001).

The colour of the seeds is commonly dark-brown to blackish, or whitish-yellowish, sometimes with reddish nuances at the species cultivated as cereals. Many cultivars of *A. caudatus* have pink cotyledons visible through the seed coat. The colour may be uniform or not, in the last case usually with the marginal zone paler. Weedy species and species used as a vegetable have mostly black or dark seeds (Costea et al., 2001; Jarošová et al., 1997; Das, 2011).

Many species of the genus are greatly affected by environmental factors (nutritional elements, water availability, light conditions, injurious factors, etc. exhibiting a great morphological variability with little taxonomic significance (Costea et al., 2001). All the above mentioned characteristics are useful for the taxonomy of the genus but difficult to use

for the current identification of taxa (Costea et al., 2001). Also it is dependent on the cultivation in the field conditions. In the case of a gene bank, when seed samples are received, it is necessary to sow them in the field conditions for the morphological and phenological evaluations. But in the case of weedy species, it would be better to know, if the samples are not harmful weeds. We need to exclude weeds from our collection.

Many different methods of identification have been used for evaluation of amaranth diversity. RAPD analysis was successful in the investigation of the relationships of four *A. hypochondriacus* varieties (Barba de la Rosa et al., 2009). AFLP markers were successfully used to determine species what demonstrated taxonomic ambiguity at the basic morphologic level (Costea et al., 2006). Other methods such as ITS, ISSR and isozyme profile were used to get exhaustive view of interrelationship and relative closeness among amaranth species (Das, 2011; Xu & Sun, 2001). Also other methods such as electrophoresis profiles of proteins have been successfully used to clarify the taxonomy of many families. There was published, that electrophoresis can also be used to characterize the seed protein profiles of species and cultivars, compare cultivars of different geographical origin, and provide taxonomically useful descriptors that are substantially free from environmental influence. This method is rapid, relatively cheap, largely unaffected by the growth environment and eliminate to grow plant to maturity (Juan et al., 2007; Jugran et al., 2010). Drzewiecki (2001) used SDS PAGE of urea-soluble proteins of amaranth seeds for distinguishing both – species and their cultivars. Samples of seven species were divided into three groups by protein patterns according to similarity. According to solubility, Osborne (1907) divided proteins into four classes: albumins soluble in water, globulins soluble in high salt concentration, prolamins soluble in aqueous alcohol and glutelins soluble in acid or alkaline solutions (Segura-Nieto et al., 1994). The division into four protein fractions brings the possibility to see the differences among seed samples more clearly. The first general characterization of the protein fraction spectra of amaranth species was performed by Gorinstein et al. (1991) and Drzewiecki et al. (2003). Finally, Dzunkova et al. (2011) set up the methodology for clear identification of the amaranth species using glutelin protein fraction. The washing off water, salt- and alcohol- soluble proteins in protein fraction separation process makes polymorphic peaks of amaranth glutelins to be distinguished very easily.

SDS PAGE has been the traditional method for analysing glutenin subunit composition of wheat, but the procedure is slow, laborious and non-quantitative. The chip microfluidic technology, based on capillary electrophoresis, provides new opportunities in analysis of wheat HMW-GSs. This procedure is rapid, simple to operate, enabling automatic and immediate quantitative interpretation. Other advantages over traditional gel electrophoresis are lower sample and reagent volume requirements and a reduced exposure to hazardous chemicals (Bradova & Matejova, 2008).

In this work, we focused on evaluation for precise determination of amaranth genetic resources in the Czech Gene Bank. One of our aims was to separate amaranth species according to protein patterns and to verify our hypothesis of different protein fraction pattern based on species and variety. We compared spectra of storage proteins and their fractions of wild weedy and cultivated species of amaranths and verified the suitability of this method for species identification in our collection.

Latin name	Synonyms
Amaranthus caudatus L.[1]	*Amaranthus caudatus* subsp. *caudatus* *Amaranthus caudatus* subsp. *mantegazzianus*
Amaranthus caudatus subsp. *caudatus*[1]	= *Amaranthus alopecurus* Hochst. ex A. Br. & Bouche = *Amaranthus abyssinicus* hort. ex L.H. Bailey. = *Amaranthus caudatus* subsp. *saueri* Jehlik = *Amaranthus caudatus* L. = *Amaranthus maximus* Mill. = *Chenopodium millmi* J.T. del Granado = *Amaranthus caudatus* var. *alopecurus* Moq.
Amaranthus caudatus subsp. *mantegazzianus*[1]	= *Amaranthus edulis* Spegazz. = *Amaranthus mantegazzianus* Passer.
Amaranthus cruentus L. [1]	= *Amaranthus caudatus*auct. = *Amaranthus paniculatus* L. = *Amaranthus hybridus* var. *cruentus* = *Amaranthus sanguineus* L. = *Amaranthus hybridus* '*paniculatus*'. = *Amaranthus speciosus*
Amaranthus retroflexus L. [1]	= *Amaranthus patulus*auct. = *Amaranthus delilei* Richter & Loret
Amaranthus hypochondriacus L. [1]	= *Amaranthus chlorostachys* var. *erythrostachys* (Moq.) Aell. = *Amaranthus leucospermus* S. Wats. = *Amaranthus leucocarpus* S. Wats. = *Amaranthus hybridus* convar. *erythrostachys* (Moq.) Thell. ex Asch. & Graebn. = *Amaranthus hybridus* subsp. *hypochondriacus* (L.) Thell. = *Amaranthus flavus* L. = *Amaranthus frumentacea* Buch.-Ham. = *Amaranthus chlorostachys* var. *leucocarpus* (S. Wats.) Aell. = *Amaranthus anardana* Buch.-Ham.
Amaranthus cannabinus (L.) J.D.Sauer[2]	*Acnida cannabina* L.
Amaranthus deflexus L. [2]	
Amaranthus tuberculatus (Moq.) J.D.Sauer[2]	*Acnida tuberculata* Moq.

[1]according to Mansfeld'sEncyclopedia of Agricultural and Horticultural Crops (Hanelt& IPGCPR, 2001)
[2]according to IPNI (2011)

Table 1. Synonyms of selected amaranth species

2. Materials and methods

2.1 Plant material

For the evaluation there were used 46 amaranth genotypes from Crop Research Institute in Prague, Czech Republic (CRI) and from USDA, ARS, NCRPIS Iowa State University. In these samples, there were 6 accessions of wild weed and 40 of the cultivated species. The acronyms used for the wild species were as follows: De - *A. deflexus*, Au - *A. australis*, Wr - *A. wrightii*, Tu - *A. tuberculatus*, Cn - *A. cannabinus*, Re - *A. retroflexus*. The cultivated samples were evaluated in the field conditions in 2008 and 2009 according to the list of descriptors for amaranths created for purposes of the Czech Gene Bank. The morphological and phenological characters are evaluated according to List of Descriptors for amaranth created in the Czech Gene Bank. Following traits were evaluated in the field conditions:

- number of days from emergence to inflorescence observation,
- number of days from emergence to flowering,
- number of days from emergence to maturity.

The first two traits were assessed when 50% of plants were in this stage. The numbers of day from emergence to inflorescence observation and the numbers of days from emergence to flowering are important characters due to fact, that certain amaranth genotypes are sensitive to day-length. Maturity was estimated when 75% of the grains were mature. Plant height was measured from the soil surface to the top of the main stem in cm. Length of inflorescence was measured from the downmost branch to the top of inflorescence of the main stem in cm. Weight of thousand seeds (WTS) was weight of thousand seeds in g.

2.2 Total seed protein content and protein fractions content determination

The measurements of total seed protein content and protein fraction content were performed by the Kjeldahl method (Czech state norm 56 0512-12) in Kjeltec automatic analyzer (Kjeltec 2300, Foss Tecator, Sweden) with the protein-nitrogen coefficient set to 6.025. Protein fractions (albumins, globulins, prolamins and total glutelins) were extracted according to the protocol developed for the wheat protein fraction separation by Dvoracek (2006) with some modifications. For the determination of protein fractions content was used 0.5 g of milled amaranth seeds. The protein fractions were extracted by adding 5 ml of solvent (distilled water for albumins, 0.5 M NaCl for globulins, cold 60% ethanol for prolamins), vortexing and centrifuging by 10 000 × g for 15 minutes (Universal 32R HettichCentrifugen, Germany). This procedure was repeated twice and the supernatants from each extraction were saved and poured together. In the case of prolamins, after first addition of solvent, tubes were vortexed and chilled to 4°C for 4 hours; after that the procedure was performed exactly as for albumins and globulins. The protein content of whole seed was also measured by milling 1g of amaranth seeds. For the boiling in the automatic digestion system (2015lift, 2020 digestor, Foss Tecator, Denmark) were used 10 ml from the obtained 15 ml of each fraction extract. Into the each 250 ml tube one catalyser tablet, 3.5 g of K_2SO_4 and $CuSO_4$ mixture and 10 ml of H_2SO_4 were added. In one tube was a blank sample. Tubes were let to boil to the temperature of 420°C for about 1 h 40 min. After cooling for about 10 min, 75 ml of distilled water was added. The content of glutelin and the residual nitrogen fraction was calculated as the difference between the content of the total seed protein and three measured fractions.

2.3 Electrophoresis of the proteins

2.3.1 Extraction of the total seed storage proteins

Five different approaches to the extraction were tested for the development of the best extraction approach:

1. single seed extracted in 18 µl of the extraction solution,
2. bulk of 10 seeds extracted in 50 µl of the extraction solution,
3. bulk of 10 seeds extracted in 100 µl of the extraction solution,
4. bulk of 100 seeds extracted in 200 µl of the extraction solution,
5. bulk of 100 seeds extracted in 400 µl of the extraction solution.

Seed samples were crushed separately and mixed with extraction solution (consisted of 0.0625 M Tris-HCl pH 8.8, 5% (w/v) 2- mercaptoethanol, 2% (w/v) SDS, 10% (w/v) glycerol, 0.01% (w/v) bromphenol blue) by vortexing (MS2 Minishaker, IKA, Germany) several times in 1.5 ml tubes. Tubes were allowed to stand at 4 °C for three hours. After this extraction time, the tubes were centrifuged at 12 000 x g for 15 min (Universal 32R HettichCentrifugen, Germany). After the replacement of the samples to the new tubes, the samples were heated in a boiling water bath for 2 min.

Ten seeds from each variety were selected randomly, crushed and put into 2 ml micro tube. The protein fractions were extracted by adding 100 µl of solvent (distilled water for albumins, 0.5 M NaCl for globulins, cold 60% ethanol for prolamins), vortexing and centrifuging by 10 000 × g for 15 minutes (Universal 32R HettichCentrifugen, Germany). This procedure was repeated twice but the supernatants of the second and third wash were always discarded. In the case of prolamins, after first addition of solvent, tubes were vortexed and chilled to 4°C for 4 hours; after that the procedure was performed as in the case of albumins and globulins. Tubes containing protein fractions extract and the seed pellets (glutelins) were freezed to -25°C. After the supernatant in the tubes became solid, the top of the tubes was perforated by a needle to form small holes what serve to prevent the loss of the sample by lyofilisation. The lyofilisation was performed by freeze dryer (Christ, Germany) during 24 h at -58°C and 0.018 mBar. The lyophilized solid samples were mixed with 100 µl extraction solution (consisted of 0.0625 M Tris-HCl pH 8.8, 5% (w/v) 2-mercaptoethanol, 2% (w/v) SDS, 10% (w/v) glycerol, 0.01% (w/v) bromphenol blue) by vortexing several times in 1.5 ml tubes. Tubes were allowed to stand at 4 °C for three hours. After this extraction time, the tubes were centrifuged at 12000 x g for 15 min. The supernatants were put into new tubes and heated in boiling water for 2 min.

2.3.2 Protein separation by SDS PAGE

The amaranth protein extracts were separated in conditions of discontinuous electrophoresis (SDS-PAGE) according to Laemmli (1970) 4% stacking gel of pH 6.8, 10% separation gel of pH 8.8 on the polyacrylamide gels of the size 180 x 160 x 0.75 mm.

On the gel was loaded:

- 15 µl of the single seed sample,
- 20 µl of the 10 seed bulk, 100 seed bulk and all the protein fraction samples,
- 7 µl of the protein marker: SigmaMarker Wide Range (MW 6,500-200,000).

The electrophoresis was performed on 90 mA (45 mA / gel) and let to run for about 4 hours. The gels were stained with a solution of 0.1% (w/v) Coomasie Brilliant Blue (CBB) R250, 50% (w/v) methanol, 10% acetic acid, 0.02% (w/v) bromphenol blue salt for 1 day and destained with a solution of 25% (w/v) denatured alcohol and 3.5% (w/v) acetic acid, what lasted also 1 day. Gels were preserved in solution: 45% (w/v) denatured alcohol, 3% (w/v) glycerol for 2 hours, then dried and stored into cellophane sheets. The whole procedure including the test of the different extraction concentrations, the protein fraction separation procedure and the electrophoresis was repeated for the control of the correct experiment performance.

2.3.3 Chip electrophoresis

All the extracted protein fraction samples were analyzed by chip capillary electrophoresis using commercial Experion Pro260 Analysis Kit for 10 Chips and the Experion automated electrophoresis system (Bio-Rad Laboratories, USA) for protein quantification according to the manufacturer's instructions. Experion automated electrophoresis station performs automatically all the steps of the gel-based electrophoresis (samples separation, staining, destaining, imaging, band detections, and data analysis).

2.4 Statistical analysis

For the statistical evaluation of morphological traits, analysis of variance (ANOVA) and the Tukey HSD test were used (software -Statistica 7.0 CZ). In the case of protein fraction proportion in accessions with different seed colour, the basic statistics of R statistics 2.10.0 software were used for calculation of mean x, standard deviation sx and p-values (adjusted by Holm correction, two sided Welch Two Sample t-test used).

The SDS-PAGE spectra of total seed storage proteins and protein fractions were compared and confronted with the spectra of the chip capillary electrophoresis. The bands in the spectra were analyzed regarding the positions of the bands and also the relative intensity of the bands. The intensity of the bands was analyzed individually for each sample considering the intensity of the internal markers of the chip electrophoresis and the general intensity of all the bands in the sample. The intensity of the bands was expressed as the relative protein concentration measured by chip capillary electrophoresis what was the multiplication of numbers 0, 1, 2, 3 used in our statistics (0– no band, 1- light band, 2 – medium intensity band, 3 – dark band). The spectra expressed as the numerical values were analyzed by R statistics 2.10.0 software. The relationships between accessions were expressed by Pearson correlation using single linkage. The hierarchical clustering dendrogram was cut at the level of correlation 0.99 to show the well defined clusters.

3. Results and Discussion

3.1 Morphological and phenological evaluation

Mean data of morphological and phenological evaluations of amaranth are shown in table 2 and 3. From our long-term observations, genotypes with number of days from emergence to flowering higher than 100 days likely does not mature before early frost in autumn. The vegetation period in evaluated collection ranged from 92±0.00 to 163.00±0.00 days. Also height of plants in maturity and length of inflorescence is a very useful character. Both are

	From emergence to inflorescence observation (days)	From emergence to flowering (days)	From emergence to maturity (days)
Genotype	Mean±SD	Mean±SD	Mean±SD
6	56.00±1.41^{abc}	72.00±7.07^{bcdef}	107.00±16.97^{ab}
11	51.50±7.78^{abc}	67.00±4.24^{abcdef}	120.50±7.78^{ab}
12	47.50±4.50^{abc}	64.00±11.37^{abcdef}	101.00±16.97^{ab}
21	46.50±3.54^{abc}	64.00±8.49^{abcdef}	108.50±24.75^{ab}
23	51.00±2.83^{abc}	64.00±9.90^{abcdef}	109.00±25.46^{ab}
24	46.00±0.00^{abc}	64.00±9.90^{abcdef}	102.00±25.46^{ab}
35	51.50±2.12^{abc}	64.50±7.78^{abcdef}	100.50±26.16^{ab}
43	64.00±0.00^{abc}	86.00±0.00^{ef}	122.00±0.00^{ab}
44	62.00±5.66^{abc}	105.00±0.00^{abcdef}	131.00±36.77^{ab}
45	44.00±0.00^{abc}	63.00±0.00^{abcdef}	145.00±0.00^{ab}
51	49.00±4.24^{abc}	67.00±5.66^{abcdef}	107.00±26.87^{ab}
62	45.00±1.41^{abc}	64.50±9.19^{abcdef}	126.00±0.00^{ab}
70	52.00±1.41^{abc}	68.50±3.54^{abcdef}	112.50±19.09^{ab}
71	45.50±4.50^{abc}	65.00±4.24^{abcdef}	111.50±17.68^{ab}
72	50.00±7.07^{abc}	68.00±2.83^{abcdef}	100.00±26.87^{ab}
73	41.00±0.00^{ab}	56.50±0.71^{abcd}	113.50±0.71^{ab}
75	52.00±2.83^{abc}	75.00±0.00^{cdef}	104.50±16.26^{ab}
76	47.50±0.71^{abc}	61.00±4.24^{abcdef}	114.50±3.54^{ab}
80	44.00±0.00^{abc}	69.00±0.00^{bcdef}	124.00±0.00^{ab}
92	75.00±7.07^{abc}	98.00±0.00^{abcdef}	163.00±0.00^{b}
95	34.50±19.09^{ab}	51.00±7.07^{abcd}	111.00±1.41^{ab}
96	36.50±12.02^{ab}	60.50±10.61^{abcdef}	106.00±16.97^{ab}
98	47.50±12.02^{abc}	78.00±0.00^{def}	121.00±0.00^{ab}
99	45.00±0.00^{abc}	63.00±0.00^{abcdef}	111.00±0.00^{ab}
101	42.50±0.71^{abc}	55.50±3.54^{abcd}	110.00±1.41^{ab}
104	43.50±7.78^{abc}	54.00±4.24^{abcd}	114.50±20.51^{ab}
107	45.00±0.00^{abc}	65.00±0.00^{abcdef}	111.00±0.00^{ab}
109	48.00±0.00^{abc}	88.00±0.00^{f}	101.50±14.85^{ab}
110	49.50±2.12^{abc}	64.50±4.50^{abcdef}	118.00±11.31^{ab}
111	52.50±2.12^{abc}	63.00±1.41^{abcdef}	116.50±9.19^{ab}
112	55.50±7.78^{abc}	70.50±10.61^{bcdef}	120.00±14.14^{ab}
120	51.50±0.71^{abc}	70.00±0.00^{bcdef}	116.00±22.63^{ab}
121	31.50±23.33^{ab}	57.00±5.66^{abcd}	92.00±0.00^{a}
123	41.50±14.85^{ab}	53.00±15.56^{abcd}	114.00±25.46^{ab}
124	47.50±3.54^{abc}	60.00±2.83^{abcde}	115.00±11.31^{ab}
125	49.5±0.71^{abc}	58.00±0.00^{abcd}	105.50±7.78^{ab}
132	35.00±0.00^{ab}	47.00±1.41^{ab}	97.00±15.56^{ab}
134	35.00±2.83^{ab}	48.50±0.71^{abc}	97.50±16.26^{ab}
136	27.50±19.09^{a}	41.00±16.97^{a}	103.00±22.62^{ab}
143	41.50±3.54^{ab}	60.00±4.24^{abcde}	98.50±17.68^{ab}
Year			
2008	45.44±11.67^{a}	66.21±13.60^{a}	115.68±15.42^{a}
2009	48.17±8.78^{a}	63.83±13.06^{a}	108.27±19.29^{b}

SD-standard deviation
Analysis of variance (ANOVA) and the Tukey HSD test were used for statistical evaluation (software - Statistica 7.0 CZ).
Different letters in the same row are statistically significant at $p > 0.05$.

Table 2. Phenological evaluation of amaranths

	Inflorescence length (cm)	Plant height (cm)	WTS (g)	Colour of seed
Genotype	Mean±SD	Mean±SD	**Mean±SD**	
6	54.00±1.41^{de}	137.50±3.54^{bcd}	0.75±0.01^{cdefghijk}	pale
11	56.50±0.71^{de}	137.50±45.96^{bcd}	0.69±0.08^{bcdefghij}	pale
12	29.50±0.71^{bc}	117.50±3.54^{abcd}	0.68±0.04^{bcdefghij}	pale
21	34.00±1.41^{bcd}	90.00±7.07^{abc}	0.58±0.03^{abcdefg}	pink
23	44.50±0.71^{cde}	127.50±3.54^{bcd}	0.76±0.08^{cdefghijk}	black
24	66.00±1.41^{e}	152.50±3.54^{cd}	0.75±0.07^{cdefghijk}	black
35	30.50±0.71^{bc}	167.50±3.54^{d}	0.90±0.00^{ijk}	pale
43	36.00±0.00^{bcd}	150.00±0.00^{cd}	0.74±0.04^{cdefghijk}	pale
44	24.50±0.71^{a}	147.50±3.54^{cd}	0.74±0.00^{cdefghijk}	pale
45	29.00±0.00^{bc}	100.00±0.00^{abc}	0.88±0.00^{hijk}	pale
51	29.00±0.00^{bc}	142.50±3.54^{bcd}	0.85±0.00^{hijk}	pale
62	37.50±0.71^{bcd}	137.50±3.54^{bcd}	0.66±0.06^{abcdefghi}	black
70	45.50±0.71^{cde}	102.50±3.54^{abcd}	0.78±0.05^{efghijk}	pale
71	52.50±0.71^{cde}	132.50±3.54^{bcd}	0.86±0.15^{hijk}	pale
72	46.50±0.71^{cde}	122.50±3.54^{bcd}	0.93±0.10^{jk}	pale
73	41.50±0.71^{cde}	130.00±8.49^{bcd}	0.71±0.01^{cdefghijk}	pale
75	60.00±0.00^{e}	132.50±3.54^{bcd}	0.91±0.10^{ijk}	pale
76	36.50±3.54^{bcd}	109.00±4.24^{abcd}	0.84±0.08^{ghijk}	pale
80	47.00±0.00^{cde}	125.00±0.00^{bcd}	0.50±0.00^{abcd}	black
92	51.00±1.41^{de}	142.50±3.54^{bcd}	0.84±0.00^{ghijk}	pale
95	35.50±0.71^{bcd}	92.50±3.54^{abc}	0.63±0.11^{abcdefgh}	black
96	34.00±1.41^{bcd}	92.60±3.54^{abc}	0.40±0.04^{a}	black
98	43.50±0.71^{cde}	152.50±3.54^{cd}	0.96±0.00^{k}	pale
99	44.00±0.00^{cde}	110.00±0.00^{abcd}	0.70±0.00^{bcdefghijk}	black
101	38.00±0.00^{bcd}	127.50±36.06^{bcd}	0.5±0.02^{abc}	pink
104	42.50±0.71^{cde}	92.50±3.54^{abc}	0.63±0.04^{abcdefgh}	black
107	22.00±0.00^{a}	53.00±0.00^{a}	0.52±0.00^{abcde}	black
109	34.50±0.71^{bcd}	77.50±3.54^{ab}	0.44±0.04^{ab}	black
110	49.50±0.71^{cde}	155.00±7.07^{cd}	0.70±0.00^{bcdefghijk}	black
111	52.00±1.41^{de}	137.50±3.54^{bcd}	0.52±0.11^{abcde}	black
112	53.00±2.83^{de}	127.50±3.54^{bcd}	0.84±0.01^{ghijk}	pale
120	51.50±2.12^{de}	95.00±42.43^{abc}	0.73±0.03^{cdefghijk}	pale
121	36.50±2.12^{bcd}	115.00±21.21^{abcd}	0.54±0.17^{abcdef}	pale
123	47.00±0.00^{cde}	102.50±3.54^{abcd}	0.80±0.00^{fghijk}	pale
124	54.50±0.71^{de}	137.50±3.54^{bcd}	0.82±0.06^{ghijk}	pale
125	51.50±0.71^{de}	127.50±3.54^{bcd}	0.75±0.14^{cdefghijk}	pale
132	46.50±4.50^{cde}	114.50±23.33^{abcd}	0.77±0.01^{efghijk}	pale
134	44.50±3.54^{cde}	119.50±17.68^{bcd}	0.81±0.08^{fghijk}	pale
136	45.50±3.54^{cde}	92.50±38.89^{abc}	0.77±0.01^{defghijk}	pale
143	39.50±2.12^{bcd}	107.50±38.89^{abcd}	0.79±0.02^{efghijk}	pale
Year				
2008	42.71±9.90^{a}	122.85±26.21a	0.72±0.15^{a}	
2009	43.22±9.87^{a}	119.02±26.36^{a}	0.73±0.14^{a}	

Different letters in the same row are statistically significant at $p > 0.05$.
SD-standard deviation
Analysis of variance (ANOVA) and the Tukey HSD test were used for statistical evaluation (software - Statistica 7.0 CZ).

Table 3. Morphological evaluation of amaranth

important for mechanized harvest by combine harvester. Lower plants with mean inflorescence are better for grain production and mechanized harvest. From our collection it is for example accession '120' with 95.00±42.43 cm height and 51.50±2.12 cm length of inflorescence. Taller genotypes are useful to develop varieties for feed utilization (Wu eta l., 2000). On the other hand, plant height could be influenced by increasing of number of plant per m^2 (Jarošová et al., 1997). The value of weight of thousand seeds (WTS) is shown in table 3. In the relation with seed colour is clear, that the biggest WTS was observed in pale seeded samples. The seed size of the genera ranges from 0.37 to 1.21 g per 1000 seed weight according to Espitia-Rangel (1994). He noted that the low value corresponding to wild and weedy species and the high values to cultivated grain species. In our experiments the WTS ranged from 0.39 to 0.96 g.

3.2 Protein content and content of protein fractions

The results of the protein content analysis showed that the highest protein content (17.32 ± 0.82%) had *A. cruentus*accessions followed by *A. caudatus* (17.24±0.65%) and *A. hypochondriacus* (16.89±0.80%). It corresponds with other published data. Segura-Nieto et al. (1994) published, that the range of protein content is following: *A. cruentus* 13.2 – 18.2%, *A. hypochondriacus* 17.9% and *A. caudatus* 17.6 – 18.4%. The range of the total protein content into our collection (12.43 – 17.33%) was similar to the results of other authors investigating various amaranth genotypes (Barba de la Rosa et al., 2009). The amaranth albumins, globulins and prolamins formed 9.2 – 14.65%, 9.78 – 13.81% and 1.76 – 3.3% of total seed protein, respectively (Table 4). The glutelins with the residual nitrogen were the most abundant. It was in accordance with the results of Bressani & Garcia-Vela (1990) and Bejosano & Corke (1999a). The very low content of prolamins (1.76 – 3.3%) confirmed the results of several authors (Gorinstein et al., 1991a; Bejosano & Corke, 1999a; Petr et al., 2003). However, another group of authors reported several times more prolamins (Correa et al., 1986; Zheleznov et al., 1997; Vasco-Mendez & Paredes-Lopez, 1995). The differences between the results of these two groups of authors might be due to the different extraction methods (Fidantsi & Doxastakis, 2001). Significant differences between black, pale and pink coloured seeds in the content of albumins were detected. Content of albumins of the black seeded group (9.64 ± 0.40%) was significantly lower (p-value 4.10^{-3}) than of the pale seeded group (13.21 ± 1.45%) and also lower than of the pink seeded group (11.39 ± 0.00; p-value 2.10^{-2}). Bresani & Garcia-Vela (1990) did not observed any differences in the protein fractions distribution among species or cultivars of the same species, independent of the fractionation sequence used. However, our results showed that the black seeded varieties had the lowest albumin content. No significant differences in other protein fractions were detected.

	Seed colour			
	black	**pale**	**pink**	**range**
WTS (g)	0.60 ± 0.12	0.79 ± 0.09	0.54 ± 0.04	0.39 - 0.96
Protein content in %	15.69 ± 0.60	16.69 ± 0.78	16.04 ± 0.00	12.43 – 17.33
Albumins	9.64 ± 0.40	13.21± 1.45	11.39 ± 0.00	9.2 – 14.65
Globulins	10.92 ± 0.78 1	11.76 ± 1.72	10.75 ± 0.00	9.78 – 13.81
Prolamins	2.37 ± 0.82	2.68 ± 0.44	2.00 ± 0.00	1.76 – 3.3
Glutelins + residual nitrogen	77.07 ± 0.33	72.35 ± 2.67	75.86 ± 0.00	69.13 – 77.44

Table 4. Total seed protein content and protein fraction content (in % of DW) of investigated accessions with respect the seed colour.

3.3 Methodical approach to protein extraction

According to our results, the chip capillary electrophoresis could replace the standard SDS-PAGE procedure, because it produced comparable results and what is more it could be performed routinely also in small laboratories thanks to its rapid performance. On the other hand, the chip capillary electrophoresis showed wider range of proteins spectra (up to 260 kDa).

The test of different concentrations was used for selection of the best extraction approach for chip and SDS-PAGE electrophoresis. By the chip capillary electrophoresis, the bulked samples of 100 seeds in 400 µl of extraction buffer were also tested. The chip capillary electrophoresis showed the high sensitivity and therefore the high concentration of the protein in the main bands resulted in their illegility. The protocol of chip electrophoresis does not provide many possibilities to chase the loaded amount of the sample. The satisfactory results of the chip electrophoresis brought the use of the single seeds.

For the SDS PAGE there were used single seed samples, bulked samples of 10 seeds extracted in 50 and 100 µl and bulked samples of 100 seeds extracted in 200 and 400 µl of extraction solution were used. The protein patterns of the samples extracted from the single seeds did not show the intensity required for the analysis of all the bands in the spectra (Figure 1). On the other hand, samples obtained by extraction of 10 seeds in 50 µl and 100 seeds extracted in 200 µl of extraction solution did not show clearly separated bands, what resulted in their illegility. In comparison with the spectra of the less concentrated samples (single seeds, 10 seeds in 100 µl of extraction solution), the main bands of the more concentrated samples were thick and joined together. The bands, which were in the less concentrated samples less intensive, were expressed so intensively that formed dark background what resulted in the impossibility of identification of the individual bands in the protein spectra. The protein spectra of the samples obtained by the extraction of 100 seeds in 400 µl were also over expressed, but the less intensive bands did not form the background, so the mayor bands were more easily identified, but several mayor bands joined together.

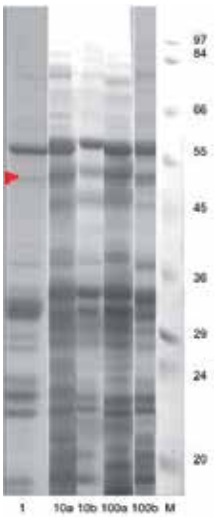
As the best approach for the total seed storage protein extraction for classical SDS-PAGE we selected bulked samples of 10 seed extracted in 100 µl of extraction solution. The bulked samples of 10 seeds extracted in 100 µl to be the most suitable tools, because of their clear expression of protein patterns and moreover they can be used when samples with higher number of seeds are not available. This selected approach differed from methodology selected by Drzewiecki (2001) who used 50 µl or by Gorinstein et al. (2005) who used 62.5 µl of extraction solution for 10 seeds bulked samples. The need for using more extraction solution in our study might be to consequence of higher protein extraction as a result of the proper seed crushing performed in our study which was not mentioned in the methodology description of other authors (Drzewiecki, 2001; Gorinstein et al., 2005).

When using total seed storage protein spectra for accessions identification by chip electrophoresis the single seed samples with several repetitions showed up as the best approach. These results were with accordance with Bradova & Matejova (2008) that compared whole seed storage proteins of wheat.

3.4 Polymorphism of the glutelins

The electrophoresis of the glutelin fraction is widely used for crop varieties identification. There were published several articles about wheat (Matejova&Bradova, 2008; Dutta et al.,

2011), rice (Gorinstein et al., 2003), barley (Smith & Simpson, 1983), lupine (Vaz et al., 2004) etc. varieties identification based on glutelin patterns. Similarly amaranth glutelins showed polymorphism not only in position of bands but also in their intensity.



1 - a single seed extracted in 18 µl,
10a - bulk of 10 seeds extracted in 50 µl,
10b - bulk of 10 seeds extracted in 100 µl,
100b - bulk of 100 seeds extracted in 400 µl of the extraction solution.
M - wide range protein marker (bands in kDa).

Fig. 1. SDS – PAGE spectra of total seed storage proteins of sample obtained by different extraction approaches.

In the cluster dendrogram (Figure2), there were clearly separated the grain and the wild monoecious and the wild dioecious accessions. All investigated amaranth species had in common three major bands of the MW 21 – 23 kDa, but remarkable differences in the rest of the spectra were the reason for the segregation into three main clusters. The glutelin spectra of the grain amaranth varieties were very similar to the total seed storage protein patterns, but the main polymorphic bands were better distinguished because of the washing off the first three fractions during fraction separation procedure which probably formed the "background" of the spectra. The principal polymorphism was detected in following band positions 38, 39, 54, 58, 60, 64 and 65 kDa with three intensity levels (1-3). The amaranth glutelins showed up as the most abundant protein fraction by SDS-PAGE analysis also in the study of Bejosano&Corke (1999). The division of the grain amaranth glutelins into three major groups reported also Gorinstein et al. (2004) and Barba de la Rosa et al. (2009).

Figure 2 indicated three well defined clusters: grain species, monoecious wild species and dioecious wild species. The grain species *A. cruentus*, *A. hypochondriacus*, *A. caudatus* closely matched together with one sample *A. mantegazzianus*. There were clearly segregated clusters with the wild monoecious species (*A.wrightii*, *A. delfexus* and *A. retroflexus*) and the wild dioecous species (*A. australis*, *A. cannabinus* and *A. tuberculatus*).

A. caudatus group presentedwas two accessions '21' and '101' characterized by the dark band 60 kDa and the light band 39 kDa in their glutelin spectra. The *A. cruentus* cluster was clearly separated in the dendrogram of hierarchical distancing by the presence of the dark band of 58 kDa and of the light band in the position of 39 kDa. *A. hypochondriacus* accessions were characterized by the lack of any band in the position 58 kDa and by the presence of the dark band 54 kDa and the light band 38 kDa. The typical band (in the position 54 kDa) used for *A. hypochondriacus* recognition was qualified as characteristic for *A. hypochondriacus* by several authors (Drzewiecki, 2001; Marcone, 2002; Gorinstein et al., 2005), but its position was determined differently: as 55 kDa (Marcone, 2002) or 52 kDa (Drzewiecki, 2001) or in the case of protein fractions as 55 kDa, too (Thanapornpoonpong et al., 2008). The characteristic presence of the band 58 kDa in *A. cruentus* spectra and of the band 54 kDa in *A. hypochondriacus* spectra was confirmed by the results of Thanapornpoonpong et al. (2008).

Some of the accessions possessed extra light band of 65 kDa and were aggregated close to the *A. hypochondriacus* cluster. Their similarity to the other *A. hypochondriacus* varieties was expressed by very high correlation 0.987.

The dark band of 54 kDa, the dark band of 64 kDa and the light band in the position 65 kDa showed up in the glutelin spectra of the accession '134'. The accession '80' had the same glutelin spectra, but its band of 54 kDa was of medium intensity. These two varieties might be the hybrids of *A. hypochondriacus* and other unknown species which could have dark band of 64 kDa and light band of 39 kDa or they might be *A. hypochondriacus* varieties with some special properties that were not considered in our study. The accessions '132' with the dark band of 60 kDa typical for *A. caudatus* accessions was also present in the spectra and therefore the correlation between these accessions and the *A. caudatus* accessions was as high as 0.911. These accessions also showed the light band of 38 kDa and the medium intensity band of 54 kDa (typical marker for *A. hypochondriacus* spectra).

The dioecious wild species *A. australis*, *A. cannabinus* and *A. tuberculatus* formed a totally distinct cluster. They possessed several major dark bands of lower molecular weight 32 - 50 kDa. From this group, *A. cannabinus* and *A. australis* were the most similar, their correlation was 0.675. The monoecious wild species (*A. wrightii*, *A. deflexus* and *A. retroflexus*) and the dioecious wild species had in common one light band in the position of 65 kDa. The major dark bands of the monoecious wild species were of MW 29 - 66 kDa. The spectra of the monoecious wild species had some similarities with the spectra of the grain species. The grain species spectra were characterized by the two bands of MW 31 and 33 kDa while in the spectra of *A. retroflexus* these bands were just "shifted up" to MW 32 and 34 kDa. Protein fractions spectra of the wild species had not been published yet by other researchers. The results indicated the high correlation of the spectra of *A. retroflexus* and *A. wrightii* what confirmed the similarity observed by the first morphological descriptions made by Watson (1877).

Accessions possessing several bands of different intensities in the polymorphic area were qualified as the hybrid accessions. The accession '99' had in its spectra several bands in the polymorphic area: the dark band of 54 kDa, light band of 58 kDa, medium intensity band in the position of 60 kDa and the light band of 65 kDa. Its similarity with *A. hypochondriacus* was expressed as correlation 0.901 and to the accession '95'. The accession '95' differed from the accession '99' just in the intensity of the bands of 58 kDa and 60 kDa (correlation 0.971). Varieties '62' and '110' were designated as hybrid varieties. They had the both bands of 54 kDa (marker for *A. cruentus*) and 58 kDa (marker for *A. hypochondriacus*) of medium

intensity. Moreover, they possessed the light band of 39 kDa. The presence of the light band 39 kDa (typical marker for *A. cruentus*) was the reason for their higher correlation with *A. cruentus* group (0.920) than with A. *hypochondriacus* group (0.892). The variety '111' was exceptional. Moreover, it had higher correlation with *A. hypochondriacus* varieties (0.960).

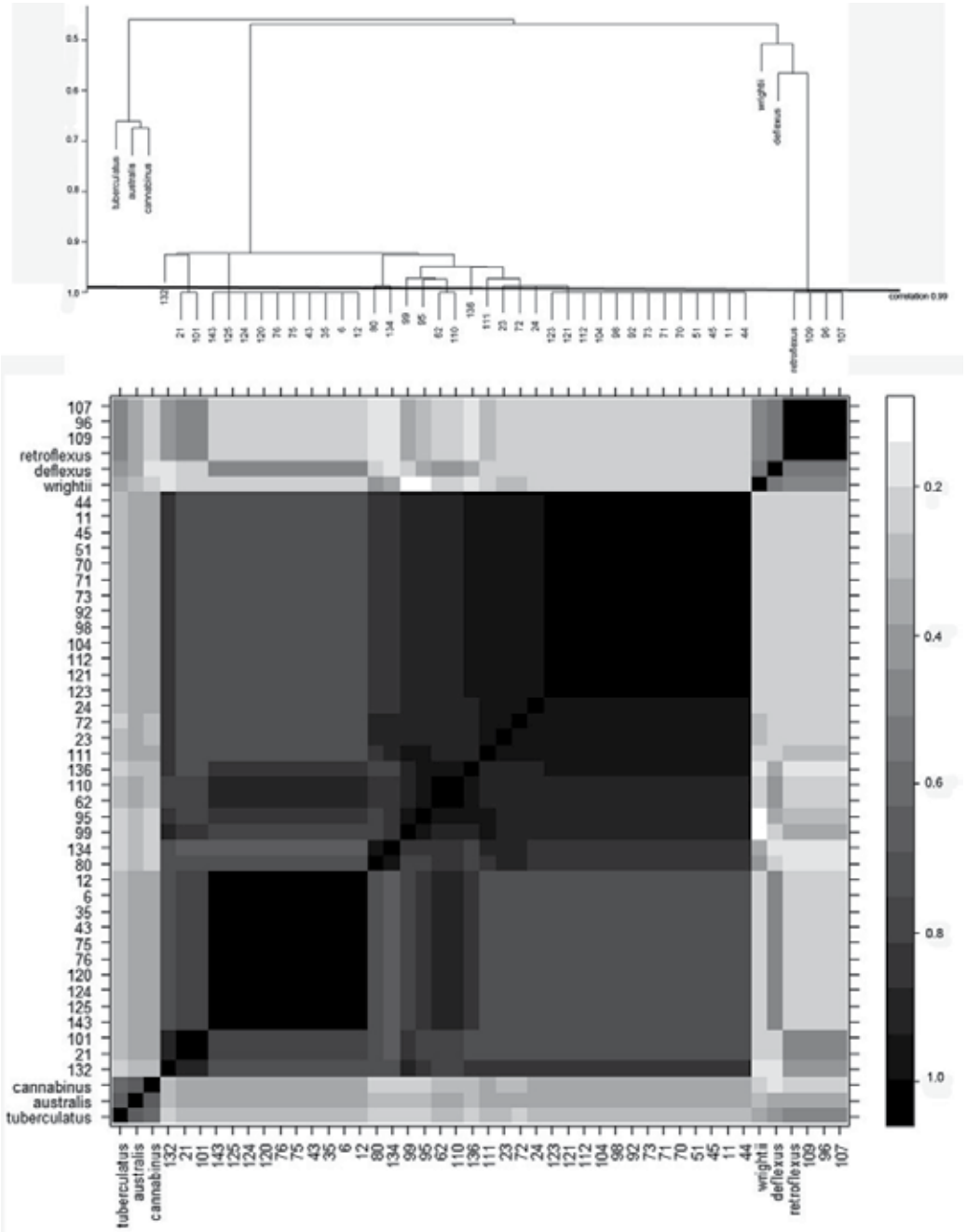


Fig. 2. Relations among amaranth samples expressed by Pearson correlation in dendrogram

4. Conclusion

Amaranth is mostly named as a crop of the future. Due to very good contents of protein, oil and many components with positive effects to humans, it is one of the promising crops. In the Czech Republic, there was interest of amaranth growing in the fields and the consumption of amaranth products is increasing as well. Most of grain raw material is imported to the Czech Republic from other countries, but there is increasing demand of Czech amaranth production. For amaranth cultivation it is necessary to know, what species could be grown. Because amaranth is not native in Europe, we have to receive seeds from other sides. In Czech legislation act about invasive weeds exists. Several amaranth species are included in this Act. In order to avoid cultivation of weedy amaranths, it is necessary to know the characteristics of the cultivated species and do not confuse them. Due to vegetable and weedy amaranth have black seed colour, it is impossible to use this trait as a marker. Amaranth glutelins were the best tool for the amaranth species identification, because they showed high polymorphism not only in position of bands but also in their intensity. The method used here was based on the data concerning the relative intensity and the position of the bands in the glutelin spectra obtained by the chip capillary electrophoresis what resulted in the exact similarity calculation of the protein fraction spectra and thus in the segregation of the cultivated grain species, the monoecious wild species and the dioecious wild species into three separate clusters. Each of the grain amaranth species was characterized by one dark band in the polymorphic region (54 – 65 kDa), while the hybrids possessed more bands of different relative intensity. The study brought several new contributions to the amaranth genetic research and is a very useful tool for species identification before cultivation in the field conditions. Unfortunately, this method is not so sensitive for individual amaranth genotype identification. We work on it in our current tasks.

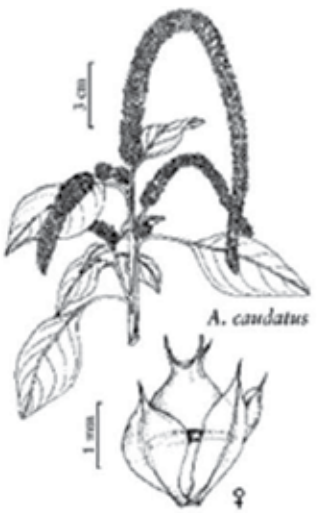


Fig. 3. *A. caudatus* (Standley, 1949)

Fig. 4. *A. hypochondriacus* (NRC, 1984)

Fig. 5. *A. cruentus* (Bojian et al., 2003)

Fig. 6. *A. cannabinus* (Standley, 1949)

Fig. 7. *A. australis* (Standley, 1949)

Fig. 8. *A. retroflexus* (Standley, 1949)

Fig. 9. *A. tuberculatus* (Standley, 1949)

5. Acknowledgement

This work was supported by the Czech Ministry of Agriculture QH92111.

6. References

Alvarez-Jubete, L.; Wijngaard, H.; Arendt, E.K. & Gallagher, E. (2010). Polyphenol composition and in vitro antioxidant activity of amaranth, quinoa, buckwheat and wheat as affected by sprouting and Baking. *Food Chemistry* 119, pp. 770–778, ISSN 0308-8146

Barba de la Rosa, A.P.; Fomsgaard, I. S.; Larsen, B.; Mortensen, A. G.; Olvera-Martınez, L.; Silva-Sanchez, C.; Mendoza-Herrera, A.; Gonzalez-Castaneda J. & De Leon-Rodrıguez A. (2009). Amaranth (*Amaranthus hypochondriacus*) as an alternative crop for sustainable food production: Phenolic acids and flavonoids with potentialimpact on its nutraceutical quality. *Journal of Cereal Science* 49, pp. 117–121, ISSN 0733-5210

Becker, R. (1994). Amaranth Oil: Composition, Processing, and Nutritional Qualities. In: *Amaranth: Biology, Chemistry and Technology*. Paredes-Lopez O. (Ed.) 133-141. CRC, ISBN 0-8493-5374-2, Boca Raton, USA

Bejosano, F. P. & Corke, H. (1999). Properties of protein concentrates and hydrolysates from *Amaranthus* and buckwheat. *Industrial Crops and Products,* 10 (3), pp. 175-183, ISSN 09266690

Bojian, B.; Clemants, S. E. & Borsch T. (2003). *Amaranthaceae*. Flora of China 9: 415-429.Available on internet http://hua.huh.harvard.edu/china/mss/welcome.htm [accessed 15. August 2011]

Brenner, D. M.; Baltensperger, D. D.; Kulakow, P. A.; Lehmann, J. W.; Myers, R. L.; Slabbert, M. M. & Sleugh, B. B. (2000). Genetic resources and breeding of Amaranthus. *Plant Breeding Reviews* 19, pp. 227-285. ISBN 0-471-38787-8, John Wiley & sons Inc.

Bressani, R. & Garcia-Vela, L. A. (1990). Protein Fractions in Amaranth Grain and Their Chemical Characterization. *Journal of Agricultural and Food Chemistry*, 38 (5), pp. 1205-1209, ISSN 0021-8561

Bradova, J. & Matejova, E. (2008).Comparison of the Results of SDS PAGE and Chip Electrophoresis of Wheat Storage Proteins. *Chromatographia* 67, pp. S83–S88, ISSN 0009-5893

Chan, K. F. & Sun, M. (1997). Genetic diversity and relationships detected by isozyme and RAPD analysis of crop and wild species of Amaranthus. *Theoretical and applied genetics*. 95: 865-873. ISSN: 1432-2242

Correa, A. D.; Jokl, L. & Carlsson, R. (1986). Amino-acid composition of some *Amaranthus* sp. grain proteins and of its fractions. *Archivos Latinoamericanos de Nutricion* 36 (3), pp. 466-476,ISSN 0004-0622

Costea, M.; Sanders, A. & Waines, G.J. (2001). Preliminary results towards revisions of the *Amaranthus hybridus* species complex (Amaranthaceae). SIDA 19 (4), pp. 931 – 974, ISSN 0833-1475

Das, S. (2011). Systematics and taxonomic delimitation of vegetable, grain and weed amaranths: a morphological and biochemical approach. *Genetic Resources and Crop Evolution*, 58, pp. 1-15, ISSN 1573-5109

Dotlačil, L.; Stehno, Z.; Michalová, A. & Faberová, I. (2001). Plant genetic resources and agri-biodiversity in the Czech Republic. In: *Agriculture and Biodiversity: Developing Indicators for Policy Analysis Proceedings from an OECD Expert Meeting Zurich*, November 2001. 66-79, Switzerland

Drzewiecki J. (2001). Similarities and Differences between *Amaranthus* Species and Cultivars and Estimation of Outcrossing Rate on the Basis of Electrophoretic Separations of Urea-Soluble Seed Proteins. *Euphytica,* 119 (3), pp. 279-287, ISSN 1573-5060

Drzewiecki, J.; Delgado-Licon, E.; Haruenkit, R.; Pawelzik, E.; Martin-Belloso, O.; Park, Y. S.; Jung, S. T.; Trakhtenberg, S. & Gorinstein, S. (2003). Identification and Differences of Total Proteins and Their Soluble Fractions in Some Pseudocereals Based on Electrophoretic Patterns. *Journal of Agricultural and Food Chemistry* 51 (26), pp. 7798-7804, ISSN 0021-8561

Dvoracek V. (2006). Optimalizace Osbornovy metody kvantifikace bílkovinných frakcí zrna pšenice ozimé (*Triticum aestivum* L.). 10p. ISBN 80-86555-81-X – Available on: http://www.vurv.cz/files/Publications/OptimalizaceOsborn.pdf [cited: 2.9.2011]

Dutta, T.; Kaur, H.; Singh, S.; Mishra, A.; Tripathi, J. K.; Singh, N.; Pareek, A. & Singh, P. (2011). Developmental changes in storage proteins and peptidylprolyl*cis–trans*isomerase activity in grains of different wheat cultivars. *Food Chemistry*. 12, pp. 450-457, ISSN 0308-8146

Dzunkova, M.; Janovska, D.; Hlasna-Cepkova, P.; Prohaskova, A. & Kolar, M. (2011). Glutelin protein fraction as a tool for clear identification of Amaranth accessions. *Journal of Cereal Science*, 53, 2, pp. 198 – 205, ISSN 0733-5210

EC (2007). *Genetic resources in agriculture. A summary of the projects co-financed under Council Regulation (EC) No 1467/94 Community programme 1994–99*, ISBN: 92-79-03599-1, Rome, Italy

Espitia-Rangel, E. (1994). Breeding of Grain Amaranth. In: *Amaranth: Biology, Chemistry and Technology*. Paredes-Lopez O. (Ed.).23-38. CRC, ISBN 0-8493-5374-2, Boca Raton, USA

Fidantsi, A. & Doxastakis, G. (2001). Emulsifying and foaming properties of amaranth seed protein isolates. *Colloids and surfaces B: Biointerfaces,* 21, 1-3, pp. 119-124, ISSN 1873-4367

Girija, K.; Lakshman, K.; Chandrika, U.; Ghosh, S. S. & Divya, T. (2011). Anti-diabetic and anti-cholesterolemic activity of methanol extracts of three species of Amaranthus. *Asian Pacific Journal of Tropical Biomedicine*, pp. 133-138, ISSN 2221-1691

Gorinstein, S.; Moshe R. & Greene, L. J. (1991). Evaluation of four *Amaranthus* species through protein electrophoretic patterns and their amino acid composition.*Journal of Agricultural and Food Chemistry,* 39 (5), pp. 851-854, ISSN 0021-8561

Gorinstein, S.; Pawelzik, E.; Delgado-Licon, E.; Haruenkit, R.; Weisz, M. & Traktenberg, S. (2002). Characterization of Pseudocereals and Cereals Proteins by Protein and Amino Acid Analyses. *Journal of the Science of the Food and Agriculture* 82 (8), pp. 886-891, ISSN 1097-0010

Gorinstein, S.; Yamamoto, K.; Kobayashi, S.; Taniguchi, H.; Pawelzik, E.; Delgado-Licon, E.; Shaoxian, Y.; Hongliang, S.; Martinez-Ayala, A. L. & Trakhtenberg, S. (2003). Inter-Relationship between Electrophoretic Characteristics of Pseudocereal and Cereal Proteins and Their Microscopic Structure for Possible Substitution Based on Nutritional Evaluation. *International Journal of Food Sciences and Nutrition* 54 (6), pp. 427- 435, ISSN 1465-3478

Gorinstein, S.; Pawelzik, E.; Delgado-Licon, E.; Yamamoto, K.; Kobayashi, S.; Taniguchi, H.; Haruenkit, R.; Park, Y.; Jung, S.; Drzewiecki, J. & Trakhtenberg, S. (2004). Use of scanning electron microscopy to indicate the similarities and differences in pseudocereal and cereal proteins. *International Journal of Food Science and Technology,* 39, pp. 183–189, ISSN 0950-5423

Gorinstein, S.; Drzewiecki, J.; Delgado, E.; Pawelzik, E.; Martinez-Ayala, A. L.; Medina, O. J.; Haruenkit, R. & Trakhtenberg, S. (2005). Relationship between dicotyledone-amaranth, quinoa, fagopyrum, soybean and monocots- sorghum and rice based on protein analyses and their used as substitution of each other. *European Food Research and Technology,* 221, pp. 69-77, ISSN 1438-2377

Hanelt, P. & Institute of Plant Genetics and Crop Plant Research (IPGCPR) (Eds.) (2001). Mansfeld's Encyclopedia of Agricultural and Horticultural Crops. 1-6, 3716 pp. ISBN: 978-3-540-41017-1, Berlin, Germany

'The International Plant Names Index' (IPNI) (2008). Available on the Internet http://www.ipni.org [accessed 22 August 2011].

Iqbal, S. M.; Ghafoor, A. & Ayub N. (2005). Relationship between SDS-PAGE Markers and Ascochyta Blight in Chickpea. *Pakistan Journal of Botany* 37 (1), pp. 87-96, ISSN 0556-3321

Jaramillo, S. & Baena M. (2002). Ex situ conservation of plant genetic resources: training module. International Plant Genetic Resources Institute, ISBN 978-92-9043-751-2, Cali, Colombia.

Jarošová, J.; Michalová, A.; Vavreinová, S. & Moudrý, J. (1997). Cultivation and utilization of amaranth. UZPI Praha, ISBN 80-7271-042-7. (in Czech)

Juan, R.; Pastor, J.; Alai, M. & Vioque, J. (2007). Electrophoretic characterization of *Amaranthus* L. seed proteins and its systematic implications. *Botanical Journal of the Linnean Society*, 155, pp. 57–63, ISSN 0024-4074

Jugran, A.; Bhatt, I. D. & Rawal, R. S. (2010). Characterisation of Agro-diversity by Seed Storage Protein Electrophoresis: Focus on Rice Germplasm from Uttarakhand Himalaya, India. *Rice Science,* 17 (2), pp. 122-128, ISSN1672-6308

Kaul, H. E.; Aufhammer, W.; Laible, B; Nalborczyk, E.; Pirog, S. & Wasiak, K. (1996). The suitability of amaranth genotypes for grain and fodder use in Central Europe. *Die Bodenkultur,* 47 (3), pp. 173 – 181, ISSN 0006-5471

Laemmli U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227 (5259), pp. 608-685, ISSN 0028-0836

Marcone, M. F. (2002). A non-DNA-based marker technique to differentiate between *Amaranthus pumilus* (a federally protected, threatened plant species) from among more common amaranth cultivars. *Journal of Food Composition and Analysis,* 15, pp. 115-119, ISSN 0889-1575

Michalova, A. (1999). *Amaranthus* L. Vyziva a potravinarstvi 54, pp. 13-14 (in Czech)

Moudry, J.; Pejcha, J. & Peterka, J. (1999). The effect of genotype and farming technology on the yield of amaranth *Amaranthus* sp. Collection of Scientific Papers, Faculty of Agriculture in ČeskeBudujovice, *Series for Crop Sciences* 16, pp. 93-98. (in Czech)

National Research Council (NRC) (1984). Amaranth: Modern Prospects for an Ancient Crop. National Academy Press, Washington, DC.80 pp.

Osborne T. B. (1907). The Proteins of the Wheat Kernel. Carnegie Institution of Washington, Washington: 119 pp.

Petr, J. ; J.; Michalik, I.; Tlaskalova, H.; Capouchova, I.; Famera, O.; Urminska, D.; Tuckova, L. & Knoblochova, H. (2003). Extention of the Spectra of Plant Products for the Diet in Coeliac Disease. *Czech Journal of Food Science*, 21 (2), pp. 59-70, ISSN 1212-1800

Pisarikova, B.; Zraly, Z.; Kracmar, S.; Trckova, M. & Herzig, I. (2005). Nutritive value of amaranth grain (*Amaranthus* L.) in the diets for broiler chickens. *Czech Journal of Animal Science*, 50, pp. 568–573, ISSN 1212-1819

T. & Roy, S. C. (2009). Genetic Diversity of Amaranthus Species from the Indo-Gangetic Plains Revealed by RAPD Analysis Leading to the Development of Ecotype-Specific SCAR Marker. *Journal of Heredity* 100 (3), pp. 338–347, ISSN 1465-7333

Rayburn, A. L.; McCloskey, R.; Tatum, T. C.; Bollero, G. A.; Jeschke, M. R.; & Tranel, P. J. (2005). Genome Size Analysis of Weedy Amaranthus Species. *Crop Science,* 45, pp. 2557–2562, ISSN 1435-0653

Segura-Nieto, M.; Barba de la Rosa, A. P. & Paredes-Lopez, O. (1994). Biochemistry of Amaranth Proteins. In: *Amaranth: Biology, Chemistry and Technology*. Paredes-Lopez O. (Ed.), pp. 76-106. CRC, ISBN 0-8493-5374-2, Boca Raton, USA

Sauer, J. D. (1950). The grain amaranths: a survey of their history and classification. *Annals of the Missouri Botanical Garden,* 37, pp. 561-619, ISSN0026-6493

Sauer, J. D. (1967). The grain amaranths and their relatives: a revised taxonomic and geographic survey. *Annals of MissouriBotanical Garden,* 54, pp. 103-137, ISSN0026-6493

Sangameswaran, B. & Jayakar, B. (2008). Anti-diabetic, anti-hyperlipidemic and spermatogenic effects of Amaranthus spinosus Linn. onstreptozotocin-induced diabetic rats. *Journal of natural medicines,* 62, pp. 79–82, ISSN 1340-3443

Smith, D. B. & Simpson P. A. (1983): Relationships of barley proteins soluble in sodium dodecyl sulphate to malting quality and varietal identification. *Journal of Cereal Science*. 1, 185-197, ISSN0733-5210

Srivastava, S. & Gupta, P. S. (2002). SDS and Native Page Protein Profile for Identification and Characterization of Elite Sugarcane Genotypes. *Sugar Technology* 4 (3), pp. 143-147, ISSN 0972-1525

Standley P. C. & Steyermark J. A. (1946): Flora of Guatemala. Series: Fieldiana: Botany 24 (4). 152 -157. Chicago, USA

Tironi, V. A. & Añón, M. C. (2010). Amaranth proteins as a source of antioxidant peptides: Effect of proteolysis. *Food Research International* 43, pp. 315–322, ISSN 0963-9969

Thanapornpoonpong, S.; Vearasilp, S.; Pawelzik, E. & Gorinstein, S. (2008). Influence of Various Nitrogen Applications on Protein and Amino Acid Profiles of Amaranth and Quinoa. *Journal of Agricultural and Food Chemistry,* 56, pp. 11464–11470, ISSN 1520-5118

Transue, D. K.; Fairbanks, D. J.; Robison, L. R. & Andersen W. R. (1994). Species Identification by RAPD Analysis of Grain Amaranth Genetic Resources. *Crop Science* 34: 1385 – 1389, ISSN: 1435-0653

Vasco-Mendez, N. L. & Paredes-Lopez, O. (1995). Antigenic homology between *Amaranth* glutelins and other storage proteins. *Journal of Food Biochemistry,* 18 (4), pp. 227-238, ISSN 0145-8884

Vaz, A. C.; Pinheiro, C.; Martins, J. M. N. & Ricardo, C. P. P. (2004). Cultivar discrimination of Portuguese *Lupinusalbus* by seed protein electrophoresis: the importance of considering "glutelins" and glycoproteins. *Field Crops Research*. 87, 23-34, ISSN 0378-4290

Xu, F. & Sun, M. (2001). Comparative Analysis of Phylogenetic Relationships of Grain Amaranths and Their Wild Relatives (*Amaranthus; Amaranthaceae*) Using Internal Transcribed Spacer, Amplified Fragment Length Polymorphism, and Double-Primer Fluorescent Intersimple Sequence Repeat Markers. *Molecular Phylogenetics and Evolution* 21 (3), pp. 372–387, ISSN1055-7903

Watson, S. (1877). Contributions to American Botany. VII. Descriptions of New Species of Plants, with Revisions of Lychnis, Erigonum, and Chorizanthe. *Proceeding of American Academy of Arts and Sciences,* 12, pp. 246-278

H.; Sun, M.; Yue, S.; Sun, H.; Cai, Y.; Huang, R.; Brenner D. & Corke, H. (2000). Field Evaluation of an *Amaranthus* Genetic Resource Collection in China. *Genetic Resources and Crop Evolution* 47 (1), pp. 43-53, ISSN 1573-5109

Zheleznov, A. V.; Solonenko, L. P. & Zheleznova, N. B. (1997). Seed proteins of the wild and the cultivated *Amaranthus* species. *Euphytica,* 97 (2), pp. 177-182, ISSN 0014-2336

24

Agronomic and Biotechnological Strategies for Breeding Cultivated Garlic in Mexico

Héctor Silos Espino et al.*
Biotechnology Applied Laboratory, Instituto Tecnológico El Llano, Aguascalientes (ITEL), km. 18 Carr. Aguascalientes-San Luís Potosí México

1. Introduction

Garlic is an apomyctic diploid species (2n=2x=16) with vegetative reproduction that belongs to the *Allium* genus (Alliaceae), which includes onion (*Allium cepa*), leek (*A. ampeloprasum*) and shallot (*A. ascalonicum*) (Mc-Collum, 1987; Figliuolo et al., 2001; Ipek et al., 2003; 2005). The importance of garlic was recognized by humans at bronze era about 5000 years ago, and since these early times, has been used as food, condiment and medicine by Asians and Mediterranean (Ipek et al., 2005). World production of garlic is ranked 14th among vegetables with a total of 14.5 million ton (Trejo, 2006). In Mexico, its consumption is about 400 g *per capita* (Chávez, 2008), and the national production is considered low as compared to other countries such as China (80 % world production), India, Korea and the rest of the world (20 %) (FAOSTAT, 2011). Still, Mexico has a place as exporter of garlic produced mainly from the states: Zacatecas, Guanajuato, Aguascalientes, Baja California, Puebla and Sonora (Trejo, 2006). The areas in Mexico during winter 2008, dedicated to garlic were 5,085 ha with a total yield of 49,968 ton (SIAP, 2011).

Among other problems, Mexican garlic has a limited spectrum of high yielding cultivars for different environments and, at the same time, have good market qualities. Keeping in mind that kind of problems, this chapter was mainly focused on agriculture and biotechnology research done at four institutions. The first two sections include morphological, physiological and cytogenetic characterizations of the most common cultivars and related germplasm; the third section describes some advances on garlic micropropagation. The last

* Flora San Juan Hernández[1], Olivio Hernández Hernández[1], Darío Silva Bautista[1], Alan Roy Macías Ávila[1], Francisco Nieto Muñoz[1], Luis L. Valera Montero[1], Silvia Flores Benítez[1], Luis Martín Macías Valdez[2], Tarsicio Corona Torres[3], Mario Leonel Quezada Parga[4] and Juan Florencio Gómez Leyva[5]
[1]*Biotechnology Applied Laboratory, Instituto Tecnológico El Llano, Aguascalientes (ITEL), km. 18 Carr. Aguascalientes-San Luís Potosí, México*
[2]*Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Campo Experimental. Pabellón de Arteaga, km 32.5 carr. Aguascalientes-Zacatecas, México*
[3]*Cytogenetic Laboratory "Maestra Czeslawa Prywer Linzbarka", Colegio de Postgraduados, km. 36.5, carr. México-Texcoco Montecillo, Texcoco, Estado de México, México*
[4]*Fundación Produce Aguascalientes A. C. Av. Universidad 604 Interior 6 y 7, Fracc. Unidad Ganadera, Aguascalientes Ags., México*
[5]*Molecular Biology Laboratory, Instituto Tecnológico de Tlajomulco, Tlajomulco de Zúñiga, Jalisco, Mexico*

section describes our strategy for obtaining garlic genotypes with higher yield capability and better bulb quality characteristics according to the market demand.

2. Origin and distribution

Garlic is native to India (Central Asia), where was considered a spice with mystical implications due to its medicinal attributes. Egyptian hieroglyphs and Roman texts refer to garlic as a source for health and strength required for physical work. During the Middle Ages was used to prevent cholera. Nowadays, it is known for its antiseptic, diuretic, vermifuge and vasodilator activities. It also stimulates bile and stomach secretions, and acts against atherosclerosis and thrombosis.

Spaniard conquerors carried with them garlic; first to Cuba and, later, to the rest of the American colonies. Early reports of garlic fields in Mexico appeared at the beginning of the twentieth century, and fifty years later, the central region of Mexico (Bajío) was the main area for garlic production. The time of harvest in that region made possible to start exporting surplus, since at that time of the year the world production is low.

Garlic species are widely distributed on boreal areas having temperate climates and mountainous areas from tropical regions. Most of the species diversity is found from Mediterranean countries to Central Asia. USA is considered as diversification center for *Allium* (Lagunes, 2009).

3. Plant description

Garlic (*Allium sativum* L.) is propagated asexually, but shows a high morphological diversity among cultivars. These cultivars have a range of adaptation to different environments (Paredes et al., 2008). Like onion, garlic plants have thin tape-shaped leaves about 30 cm long. Roots reach a 50 cm depth or little more. Heads or bulbs are white-skinned, divided into sections called cloves. Each head could have from 6 to 12 cloves, which are covered with a white or reddish papery layer or "skin". Bulbs are consumed fresh, totally or partially dried, and pickled. Although the bulb consumption is more common, tender shoots sometimes are a delicatessen for sophisticated cuisine. These shoots may be prepared like asparagus.

Each clove is capable to develop a new plant, since they have an apical shoot bud that can elongate even though if they are not sown. This shoot is apparent after three months after the harvest, depending on the genotype and conservation conditions. Flowers are white, and the stem of some species also produce small bulbils. These stems produce a strong odor from two compounds: alliin and diallylsulfide.

4. Mexican genotypes

The origin of present-day genotypes in Mexico was a group of cultivars: 'Perla', 'California', 'Chileno' and 'Taiwan', which were introduced from USA and China. A short description of each is included:

'Perla'. Late cultivar (240 d), with creamy-colored bulbs; 10 to 16 cloves per bulb covered with about seven outer layers. Plant height is 40-45 cm tall, having a pale-green open

canopy. Experimental yields from this genotype usually range from 16 to 18 ton/ha. Physiological disorders are common, such as brush-like plant growth with excessive number of thinner leaves; the more severe this problem, the more the plant opens its canopy of leaves with reduced sheath. Bulbs of brush-like plants lose their covering layers, producing naked cloves. This disorder is high temperature-dependent, having the highest temperature influence on March and April; therefore, it varies in severity from year to year. Experimental observation indicates that some other factors alone or combined may be related to the induction of brush-like plants such as: early planting, excessive nitrogen fertilization, and planting density. This disorder worsens when these factors appear combined.

'California'. Late cultivar (260 d); recently introduced to Aguascalientes (Mexico). Bulbs are white, containing 18-26 cloves. Experimental yields range from 18 to 20 ton/ha. Plants are 50 cm tall on the average; leaves are pale-green with open canopy.

'Chileno'. Early cultivar (160 d), with a yield average of 7 ton/ha. Bulbs are purplish with 5-6 covering layers; containing 11-22 cloves (average = 19). Plants are about 50 cm tall; with semi-compact canopy and dark green leaves.

'Taiwan'. Early cultivar (170 d); its yield average is 7 ton/ha. Bulbs are purple in color, with 7-13 cloves (average = 9). Plant height reaches 50 cm on the average, with semi-compact canopy and dark green leaves.

5. Field performance of promissory genotypes

Besides the previously described genotypes, some more garlic accessions from the germplasm bank of INIFAP-CAEPAB (Fig. 1) were tested for their performance on the field (Aguascalientes, Mexico). These accessions have features suitable for breeding, as described below.

5.1 Bulb size and number of cloves per head

Higher values for bulb perimeter were found in 'California' (23.1 ± 1.8 cm) and 'Coreano' (20.4 ± 0.7 cm) varieties, as well as in the cultivars 'Perla' 'C-37-1/8'(21.1 ± 0.6 cm) and 'Perla' 'C-3-1/25' (20.5 ± 0.7 cm) (Table 1). Three varieties of white garlic, 'California' (112.3 ± 22.8 g), 'Perla' 'C-37-1/8' (84.3 ± 8.1 g) and 'Perla' 'C-3-1/25' (79.2 ± 7.8 g), as well as a marbled one, 'Coreano' (82.3 ± 8.0 g), showed the greatest bulb weight. Regarding the number of cloves per bulb, variety 'Español' produced 7.5 (±0.9), while cultivars 'Perla' 'C-3-1/25' and 'Perla C-37-1/8' had 10.9 (±1.3) and 11.9 (±1.9), respectively. Plants showing a smaller number of cloves per bulb appeared to have greater clove weights. 'Chino' and 'Coreano' varieties also showed a good clove weight performance. However, they are more susceptible to diseases and they require more time for bulb formation. Varieties with greater bulb weights appeared to be taller than those with smaller bulb weights (Table 1).

5.2 Days to harvest and yield

Varieties of garlic can be harvested at either 150 (early cycle), 180 (intermediate cycle) or 210 (late cycle) days after planting. Late cycle varieties showed greater bulb and clove weights [i.e., 'California' and 'Coreano' varieties, and 'Perla' cultivars (Table 1)]. Greatest bulb

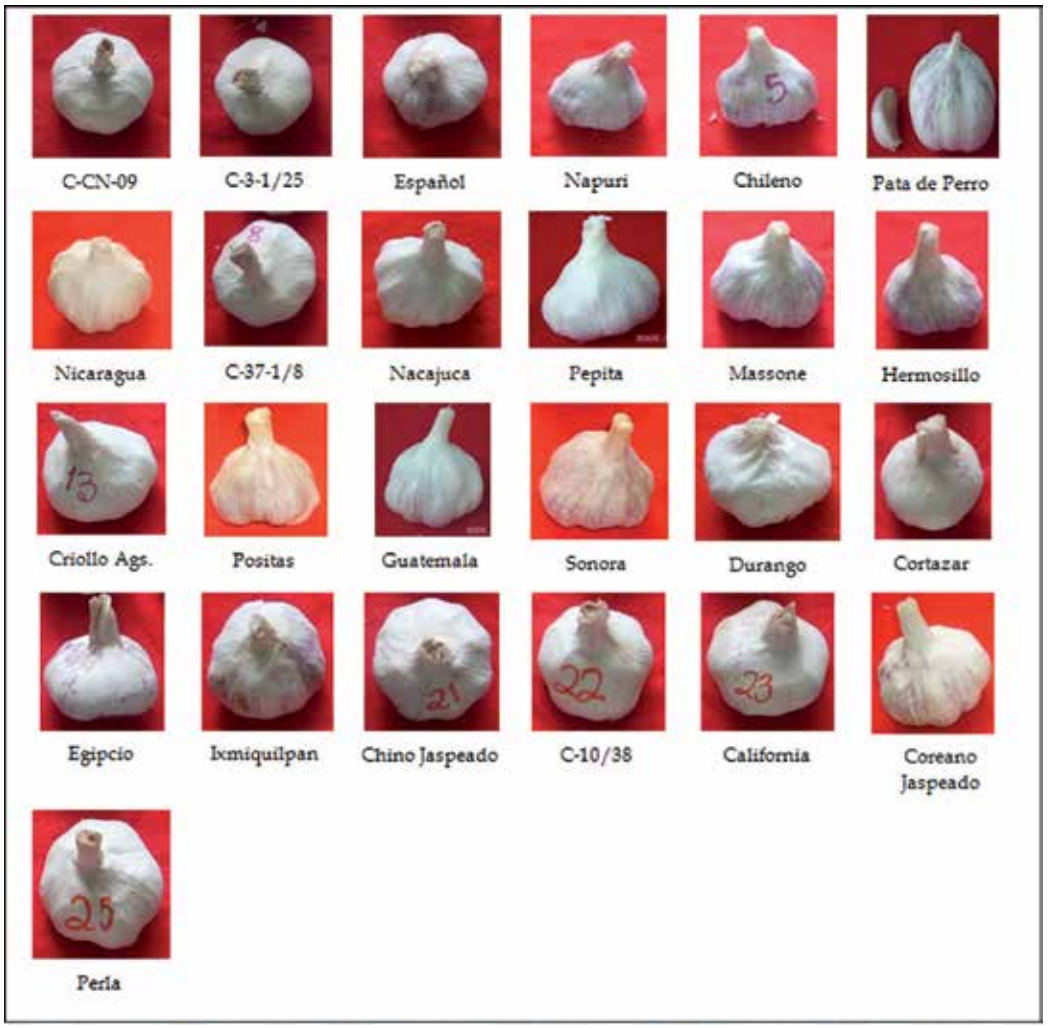


Fig. 1. Garlic genotypes from the germplasm bank at Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias - Campo Experimental Pabellón Aguascalientes (INIFAP-CAEPAB).

weight varieties (i.e., 'California') showed more than 75% greater bulb weight than lowest bulb weight varieties (i.e., 'Pata de Perro'). 'Perla' cultivars ('C-3-1/25' and 'C-37-1/8') had a better tolerance to environmental conditions (Fig. 1), their bulbs had fewer cloves (10-12) (Table 1), and their bulb and clove weights were favorably compared with those of commercial varieties (i.e., 'Chino' and 'Coreano': Table 1). 'Sonora', 'Positas', 'Hermosillo', 'Español', 'Pepita', 'Massone', 'Nicaragua', 'Nacajuca' and 'Chileno' cultivars showed very similar patterns in morphological characteristics and yields (Table 1). Correlation between bulb length and clove weight against bulb weight showed an $R^2 = 0.9668$, while clove weight against bulb weight had an $R^2 = 0.635$ (both highly significant, $p < 0.01$); in agreement with results of Heredia & Delgadillo (2002), these varieties which showed greater bulb weights also showed greater clove weights.

Genotype	Bulb length (cm)	Bulb weight (g)	Clove number/bulb	Plant length (cm)	Clove weight (g)
California 20/1	23.1 ± 1.8	112.3 ± 22.8	17.4 ± 4.6	71.1 ± 2.9	8.1 ± 2.7
Coreano	20.4 ± 0.7	82.3 ± 8.0	12.9 ± 3.5	67.6 ± 4.9	9.9 ± 2.0
Perla C-3 - 1/25	20.5 ± 0.7	79.2 ± 7.8	10.9 ± 1.3	74.3 ± 3.4	12.4 ± 1.1
Perla C-37 - 1/8	21.1 ± 0.6	84.3 ± 8.1	11.9 ± 1.9	73.9 ± 3.4	12.8 ± 3.7
Chino	18.7 ± 0.9	62.5 ± 6.7	12.2 ± 1.4	49.2 ± 3.8	7.2 ± 1.1
Ixmiquilpan	18.7 ± 1.6	60.8 ± 10.4	21.8 ± 1.3	72.9 ± 2.4	3.7 ± 1.1
Durango	19.9 ± 1.0	71.9 ± 7.7	19.9 ± 3.3	75.7 ± 3.6	4.8 ± 1.0
Criollo Ags.	17.4 ± 1.1	49.7 ± 8.3	14.5 ± 1.4	63.9 ± 2.9	5.3 ± 2.6
Cortazar	18.4 ± 0.7	57.4 ± 5.8	20.7 ± 2.3	64.7 ± 2.9	3.3 ± 0.5
Sonora	14.9 ± 0.7	36.4 ± 5.2	16.9 ± 3.5	47.7 ± 5.4	2.5 ± 0.8
Guatemala	14.8 ± 1.3	33.8 ± 8.8	14.8 ± 2.5	62.0 ± 5.4	3.0 ± 0.7
Positas	14.3 ± 0.7	34.6 ± 4.9	14.7 ± 3.5	51.5 ± 3.2	3.7 ± 0.9
Hermosillo	13.9 ± 1.3	31.9 ± 6.7	12.6 ± 3.7	60.1 ± 6.6	2.8 ± 0.8
Español	13.6 ± 1.4	24.6 ± 5.7	7.5 ± 0.9	59.9 ± 2.4	4.2 ± 1.0
Pepita	14.0 ± 0.6	32.9 ± 4.2	20.0 ± 4.0	48.3 ± 5.6	2.5 ± 0.6
Massone	13.2 ± 0.6	30.2 ± 3.6	15.0 ± 2.1	48.1 ± 4.6	2.9 ± 0.8
Nicaragua	14.1 ± 0.9	29.6 ± 6.6	13.0 ± 2.7	51.5 ± 3.2	2.7 ± 0.8
Nacajuca	14.4 ± 1.1	31.9 ± 6.7	18.4 ± 3.7	54.7 ± 7.0	2.9 ± 0.8
Chileno	14.2 ± 1.2	32.4 ± 5.8	17.6 ± 5.7	49.2 ± 3.8	2.8 ± 0.5
Napuri	13.5 ± 1.3	31.3 ± 7.1	14.8 ± 4.9	47.0 ± 2.9	3.0 ± 0.7
Pata de Perro	13.3 ± 0.7	27.7 ± 4.4	8.4 ± 1.2	55.0 ± 3.6	3.4 ± 1.3

Table 1. Size and weight of garlic varieties cultivated in Aguascalientes (Central-North Region of Mexico). Data are presented according to standard descriptors for garlic (IPGRI, 2001). Each value shows the mean ± SE.

5.3 Postharvest photosynthesis and respiratory activity of stored cloves

In order to understand some physiological events of stored garlic, analyses of six genotypes under storage were focused on the respiratory process and photosynthesis. To accomplish that goal, photosynthetic activity of stored cloves during 0, 30, 60 and 90 d were measured on three cloves selected at random from the container of each genotype. Measurements included: evaporation rate mM/s/m2/s (E), stomatal conductance mM/m2/s (G), net

photosynthesis assimilation μM/m2/s (A) and CO_2 internal concentration ppm (CI). From these measurements, it was found that for some genotypes like 'C-CN-9/2' evapotranspiration was the highest at 30 d, as opposed to 'Criollo Aguascalientes' and 'Chino Jaspeado' with the lowest value at 90 d. Stomatal conductance was high 'C-CN-9/25', mainly after 90 d. Most genotypes showed negative photosynthesis rate, and internal CO_2 showed no clear tendency within genotypes. Weight remained stable during the first 60 d, but after that period, it decreased about 1/3 of the initial values. The bulbs behavior at the final of postharvest period is show in the Fig. 2.

Fig. 2. View of stored garlic at 90 d. Sprouted heads are from short-shelf life cultivars. Heads with low number of sprouting bulbs are from 'Perla' (bottom left corner).

6. Isolation and culture of garlic protoplasts

Garlic genotypes derived from 'Perla' (C-37-1/8, C-3-1/25 and C-CN-95/2), 'Chino', 'Coreano' and 'Criollo Aguascalientes' were used for protoplast preparation. Cloves of these genotypes were peeled and disinfected (briefly: cloves were soaked in 70% ethanol 1 min and 20% commercial bleach 20 min, after three rinses with sterile distilled water the cloves were soaked again in 70% ethanol 1 min and rinsed again with sterile distilled water). Surface-sterilized cloves were cut into pieces 0.5 to 1.0 cm long, and were inoculated aseptically on MS (Murashige & Skoog, 1962) medium pH 5.8 containing 0.15 mg L^{-1} 2,4-D (Dichlorophenoxy acetic acid), 5.0 mg L^{-1} BA (Benzyl adenine), 50 g L^{-1} glucose and 3.5 g L^{-1} Phytagel. Incubation conditions were 28 ±2 °C, with a light intensity of 59 μEm²s.

Protoplasts were isolated from leaves and callus using a modification of the method from Spangenberg (1997). Briefly: 1 g of tissue was placed in petri dish together with 5 ml of enzyme mixture (2% Onozuka R-10, 1% macerozyme, 0.5% pectinase, 0.5% mannitol and 0.9% $CaCl_2$); dishes were incubated on orbital shaker at 5 rpm, about 4-6 h (Novák et al., 1983). Callus tissues were best for protoplast isolation within the range 10^6-10^8 counts/ml (Fig. 3). Protoplasts were isolated from the debris and inoculated on semi-solid MS medium. Unfortunately, whole plant regeneration remained elusive.

Genotype		Clove weight (g)		Evaporation rate (mM/s/m²/s)		Stomatal conductance (mM/m²/s)		Net photosynthesis assimilation (μM/m²/s)		CI (ppm)	
	days	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
'C-CN-9/2'	0	65.30	8.21	69	21	45	15.5	-2.1	0.5	267	49.5
	30	64.6	9.7	0.3	0	14	1	-1.6	2.9	327	16.9
	60	64.0	8.3	0.5	0.7	40	59	-5.6	7.1	1332	1938.2
	90	19.4	30.4	7.6	0.8	6618.3	2929.3	-1.6	0.4	192.7	6
'C-3-1/25'	0	68.50	7.27	0.5	0	23.7	4	-2.5	1.1	306.3	56.9
	30	61.1	4.4	0.3	0.1	13.3	3.8	0.3	3.7	247.7	223.1
	60	67.2	7.0	0.4	0.1	28.3	3.2	-3.8	1.3	372.7	66
	90	15.6	31.8	1.6	1.3	122.7	138.5	-3.5	1.3	313.7	120.5
'C-37-1/8'	0	66.35	17.22	1.3	0.6	75.7	27.1	-1.6	0.6	288.7	15.2
	30	63.2	4.3	0.2	0.2	8.7	7.6	-14.4	17.2	3849	5328
	60	65.1	16.4	41	23.4	27.7	1.5	-4.7	3.4	356	289.2
	90	21.8	28.7	4.6	3.6	1064.7	1210	7.5	15.5	295	119.8
'Chino Jaspeado'	0	89.93	12.51	0.3	0.1	20.7	10.8	-3	0.5	492.7	309.5
	30	92.3	15.7	0.6	0.3	30.3	13.5	-2.9	1.3	346.3	46.2
	60	88.5	12.0	1.1	0.2	78.3	14.6	4.4	7.9	156.3	50.3
	90	28.6	42.2	0.1	0.1	6.3	5.5	-1	2.7	0	0
'Coreano Jaspeado'	0	94.38	11.73	0.2	0.1	18.3	10.5	-1	2.1	333.3	140.8
	30	92.2	8.7	0.4	0.2	16.7	7.1	-0.2	2.6	233.7	212.7
	60	93.1	11.4	2.1	1.1	152.3	96	-8.3	2.6	311.3	131.5
	90	25.1	45.0	4.1	5.4	3371	5740	-3.8	6.1	315	103.6
'Criollo Aguascalientes'	0	48.83	7.77	0.4	0.1	23.7	5.7	-1.5	0.3	309.3	68.2
	30	51.0	7.2	0.2	0.2	7.7	6.7	-2.3	2.5	3678.7	5477.2
	60	48.0	7.5	0.4	0.2	26	12.5	-8.2	9.3	471.7	40.1
	90	15.9	23.0	0.1	0	5.3	2.9	-9.6	4.1	1899.3	1388.7

Table 2. Photosynthetic and respiration rate of cloves from six garlic genotypes after 90 d storage.

Fig. 3. Protoplast isolation from P-C-3 1/25 genotype: A,B) Bulb ('Perla' type), C) *In vitro* callus culture and D) Protoplasts at 40x magnification.

7. Karyotyping Mexican garlic genotypes

Karyotypes of C-CN-95/2, C-37 1/8, C-3 1/25 (all these 'Perla' type genotype), 'Chino', 'Coreano' y 'Criollo' were obtained from root tips. Cloves of these genotypes were placed inside petri dishes containing wet cotton wool in order to induce roots 1-2 cm long. Roots were removed and soaked with 0.05% w/v colchicine and placed in the darkness for 3:30 h at 25°C. These roots were fixed with Farmer´s solution (ethanol and glacial acetic acid, 3:1 v/v); then, they were hydrolyzed with 1N HCl at 60 °C for 10 min. Feulgen stain was applied to fixed roots before maceration with an enzymatic solution (2% pectinase, 5% celulase and citrate buffer pH 4.5) for 30 to 60 min. Roots tips were placed on microscope glass slides with a drop of 2% propionic orcein, and sandwiched with a cover glass. The slides were heated for few seconds with an alcohol burner with a very soft press so that single cells could be freed from the tissue. Then the cover glasses were gently tapped with a pencil in order to squeeze single cells for releasing and spreading the chromosomes. Observation of microphotography 100x25" allowed the following counts and measurements: chromosome number, short arms (p), long arms (q) total length, relative size of the chromosome and arm relationship (García, 1990). All of these observations were useful to classify each garlic genotype according to karyotype nomenclature and formule from Levan et al., (1964).

According to the observations, all of the genotypes tested have a chromosome number $2n$=8x=16 (Fig. 4) in agreement with other reports (Battaglia, 1963; Verma & Mital, 1978; Koul et al., 1979). The karyotype characteristics found for centromer position were as follows: 'P-CN-95/2' (1M+4m+3sm), 'P-37-1/8' (1M+6m+1st), 'P-C-3-1/25' (7m+1sm), 'Chino' (1M+3m+3sm+1st), 'Coreano' (6m+2sm) and 'Criollo' (5m+3sm); Code: M or m=metacentric, sm= submetacentric and st=subtelocentric (Table 3). The nuclear content (2C value), which is considered one of the highest among the cultivated plants (Ipek et al., 2005), is 32.7 pg. Additionally, garlic has a high karyotype variability that may be attributed to repetitive ADN (Kirk et al., 1970).

8. Genetic profile of Mexican garlic

Genetic markers are efficient tools for genetic analysis of populations and individuals. According to this concept, molecular characterization of garlic around the world has been performed either through RAPDs (Bradley & Collins, 1996; Eom & Lee, 1999; Shasany et al.,

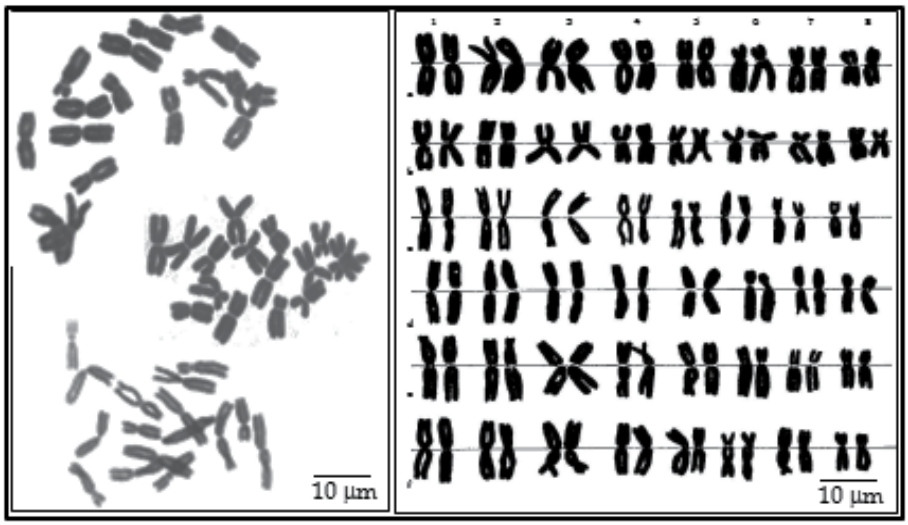


Fig. 4. Chromosomes from somatic cells (*n*=8x=16) from garlic genotypes: 'C-CN-95/2', 'C-37- 1/8', C-3 1/25, 'Chino', 'Coreano' and 'Criollo'.

2000; Peiwen et al., 2001; Ipek et al., 2003; Paredes et al., 2008; Pardo et al., 2009) or through AFLP (Rosales & Molina, 2007). Our version of this kind of analysis with 20 Mexican genotypes was as follows: Twenty garlic genotypes were subjected to RAPDs in order to construct a distance tree using clustering with the Unweighted Pair Group Method with Erithmetic Mean (UPGMA). DNA was extracted according to Doyle and Doyle (1990); DNA samples were run on agarose gel 0.8% and DNA concentration was measured with a spectrometer (model GBC Cintra 10e UV-visible). RAPD reactions were performed in a 25 ml volume, consisting of 10x buffer solution [10 mM Tris-HCl buffer (pH 8.0), 50 mM KCl_2], 2.5 mM $MgCl_2$, 2.5 units of Taq DNA polymerase (Promega), 100 µM dNTP, 50 ng genomic DNA and 0.4 µM OPB series (OPB-8, OPB-9, OPB-10, OPB-11, OPB-15 and OPB 17) primer (Operon Technologies, Alameda, CA, USA). A total of 20 µl of mineral oil was placed over the reaction mixture. Amplifications were carried out in a DNA thermocycler (Model FPR0G02Y Techne Progene, England), under the following conditions: an initial denaturalization step of 2 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 1 min at 35 °C, and 2 min at 72 °C, with a final extension step of 7 min at 72 °C. Amplification products were analyzed by electrophoresis in a 1.2% agarose gel. It was run at 100 V for 4 h, and detected by staining the gel with ethidium bromide (10 ng/100 ml of agarose solution in TBE). All visible and unambiguous fragments amplified by the chosen primers were entered under the heading of total visible fragments. Fragment data were entered on a spreadsheet to form a binary matrix, where (1) represented fragment presence and (0) fragment absence for each fragment accession combination. Cluster analysis was conducted by converting the data matrix into a similarity matrix using a simple matching coefficient. This coefficient was calculated by dividing the number of matches (0-0 and 1-1) by the total number of comparisons (Nei & Li, 1979). A cluster analysis was then conducted using the unweighted pair group method, with arithmetical averages (UPGMA) process using the S–Professional Plus 2000 program. The results obtained were compared with others studies realized by different authors, and were discusses as following: Six decamer OPB primers showing

Chrom. no.	*LBL*	*LBC*	*LT*	*LR*	*r*	*N*	*Chrom. no.*	*LBL*	*LBC*	*LT*	*LR*	*r*	*N*
'P-CN-95/2'							**Chino'**						
1	6.75	2.5	9.25	0.12	2.7	sm	1	6.75	4	10.75	0.13	1.7	sm
2	6.5	2.5	9	0.11	2.6	sm	2	6.5	6.25	12.75	0.16	1.04	m
3	6	6	12	0.15	1	M	3	6.5	2	8.5	0.10	3.25	st
4	6	5.5	11.5	0.15	1.09	m	4	6	6	12	0.15	1	M
5	6	5	11	0.14	1.2	m	5	6	5.5	11.5	0.14	1.09	m
6	5.5	4	9.5	0.12	1.3	m	6	5.5	4.75	10.25	0.13	1.2	m
7	5	4.5	9.5	0.12	1.1	m	7	5.5	3.5	9	0.11	1.6	sm
8	4.5	2.5	7	0.09	1.8	sm	8	4.5	2.5	7	0.08	1.8	sm
Formula: 1M+4m+3sm							Formula: 1M+3m+3sm+1st						
'P-37-1/8'							**'Coreano'**						
1	6.75	2	8.75	0.14	3.4	st	1	7	5.5	12.5	0.15	1.2	m
2	5	4.5	9.5	0.15	1.1	m	2	7.5	4	11.5	0.14	1.8	sm
3	5	4	9	0.14	1.2	m	3	7	2.5	9.5	0.11	2.8	sm
4	4	4	8	0.13	1	M	4	6.5	6	12.5	0.15	1.08	m
5	4.5	2.75	7.25	0.12	1.6	m	5	6.5	5	11.5	0.14	1.3	m
6	4	3.25	7.25	0.12	1.6	m	6	5	4.5	9.5	0.11	1.1	m
7	4	3	7	0.11	1.3	m	7	5	4.5	9.5	0.11	1.1	m
8	3.5	2.25	5.75	0.09	1.5	m	8	4.5	3	7.5	0.09	1.5	m
Formula: 1M+6m+1st							Formula: 6m+2sm						
'P-C-3-1/25'							**'Criollo'**						
1	6.75	4	10.7	0.14	1.7	m	1	7	5.5	12.5	0.15	1.2	m
2	6.5	5.5	12	0.15	1.2	m	2	7.5	4	11.5	0.13	1.8	sm
3	6	5	11	0.14	1.2	m	3	6.5	6	12.5	0.15	1.0	m
4	5.75	5.5	11.25	0.15	1.04	m	4	6.5	6	12.5	0.15	1.0	m
5	5.75	2.75	8.5	0.11	2.0	sm	5	6.5	4.5	11	0.13	1.4	m
6	5	4	9	0.12	1.3	m	6	6.5	2.5	9	0.10	2.6	sm
7	5	3.5	8.5	0.11	1.4	m	7	5.5	3.25	8.75	0.10	1.7	sm
8	4	2.5	6.5	0.08	1.6	m	8	4.5	3.5	8	0.09	1.3	m
Formula: 7m+1sm							Formula: 5m+3sm						

Table 3. Chromosome morphological description from Mexican garlic genotypes. LBL=Long arm, LBC=Short arm, LT=Total length, LR=Relative length, r=Arm relationship and N=Centromere nomenclature.

distinct polymorphic fingerprint were selected to reveal the genetic variation among the garlic samples. In almost all varieties, it was possible to identify around 10 bands. A dendrogram was generated from the binary matrix of measured data (Fig. 5), and two groups were identified. The first group was formed by eight varieties ('Durango', 'Nicaragua', 'Cortazar', 'Hermosillo', 'Massone', 'Pepita', 'Sonora' and 'Napuri') that are characterized by a lower production (smaller clove weight and/or greater number of cloves: Fig. 6), required more days to dormancy (6 months) and need fewer days (150) to harvest (data not shown). The second group was constituted by white, colored and marbled garlic ('Coreano', 'Positas', and 'Perla' cultivars, 'Criollo Aguascalientes', 'Español', 'Chileno', 'Ixmiquilpan', 'California', 'Chino', 'Pata de Perro' and 'Guatemala'). These are characterized by better bulb and clove weights, lower clove numbers/bulb, fewer days to dormancy (5-6 months), and between 180-210 days to harvest. In general, garlic varieties were clustered according to yield level, clove and bulb weights, number of cloves/bulb and dormancy period (Table 1). These results agree with those of García et al., (2003) using the AFLP technique. The most productive variety ('California') has the inconvenient of having a larger number of cloves/bulb and requires lower temperatures to achieve complete bulb formation. Dissimilarity among the two groups was 0.33. The lowest dissimilarity (0.0) corresponded to the most related varieties ('Sonora'-'Napuri', 'Criollo Aguascalientes'-'Español', 'California'-'Chino' and 'Pata de Perro'-'Guatemala'). The highest dissimilarity (0.70) was between the variety 'Cortazar' and the varieties 'Ixmiquilpan', 'Pata de Perro', and 'Guatemala'. (Choi et al., 2003) reported a dissimilarity of 0.4 between two large groups from a total of 75 garlic varieties using the "M" or affinity coefficient. Using the Jaccard coefficient, Ipek et al., (2003) obtained a lowest dissimilarity of 0.0 and a highest dissimilarity of 0.75 between two large groups of garlic. These results are similar to those of Al-Zaim et

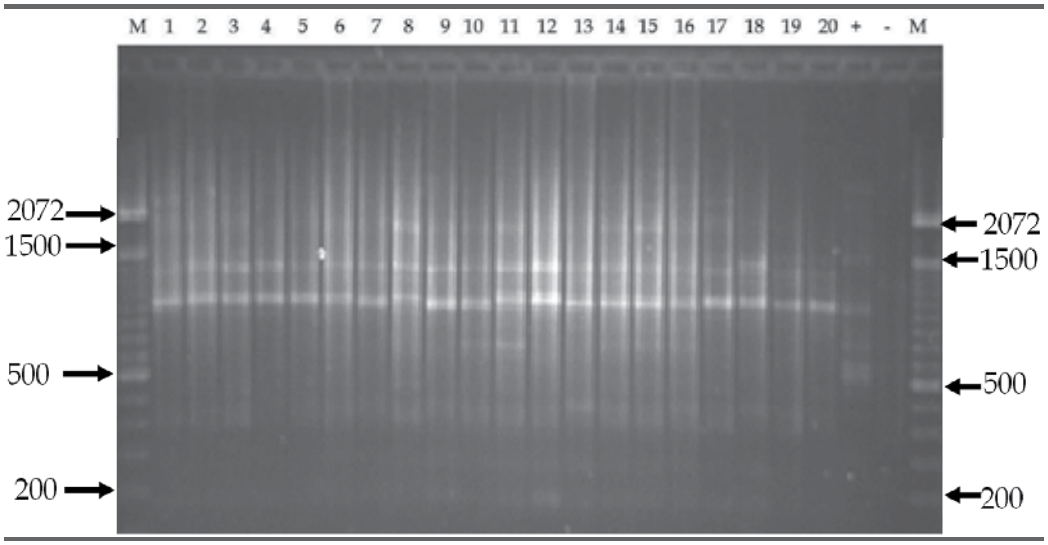


Fig. 5. RAPD from Mexican genotypes. Lanes: 1: C-3-1/25, 2: C-37-1/8, 3: 'Coreano', 4: 'California', 5: 'Chino', 6: 'Criollo Aguascalientes', 7: 'Español', 8: 'Cortazar', 9: 'Positas', 10: 'Pepita', 11: 'Massone', 11: 'Durango', 13: 'Chileno', 14: 'Hermosillo', 15: 'Sonora', 16: 'Nápuri', 17: 'Nicaragua', 18: 'Ixmiquilpan', 19: 'Pata de perro', 20: 'Guatemala', (+): positive control, (-): negative control and M: Molecular weight marker.

al., (1997). Evaluating diversity and genetic relationships among the progenitor *A. longicuspis* and 27 garlic varieties collected from different regions of the world, these authors found a dissimilarity of 0 between two samples, and a highest dissimilarity of 0.82 between two large groups. Through the RAPD technique used in this work, the two 'Perla' cultivars were grouped with the best production varieties, where the two selections presented a dissimilarity of 0.1. However, the 'Perla' C-3-1/25 cultivar showed a band of 2100 bp, and could thus be identified as a possible molecular marker. Our results allowed to identify highly related garlic varieties ('Sonora'-'Napuri', 'Criollo Aguascalientes'-'Español', 'California'-'Chileno' and 'Pata de Perro'-'Guatemala'), and separate them from varieties that are characterized by a lower yield (i.e., 'Pata de Perro' and 'Napuri'), and from mixed garlic that has been generated from introduced commercial varieties ('Criollo

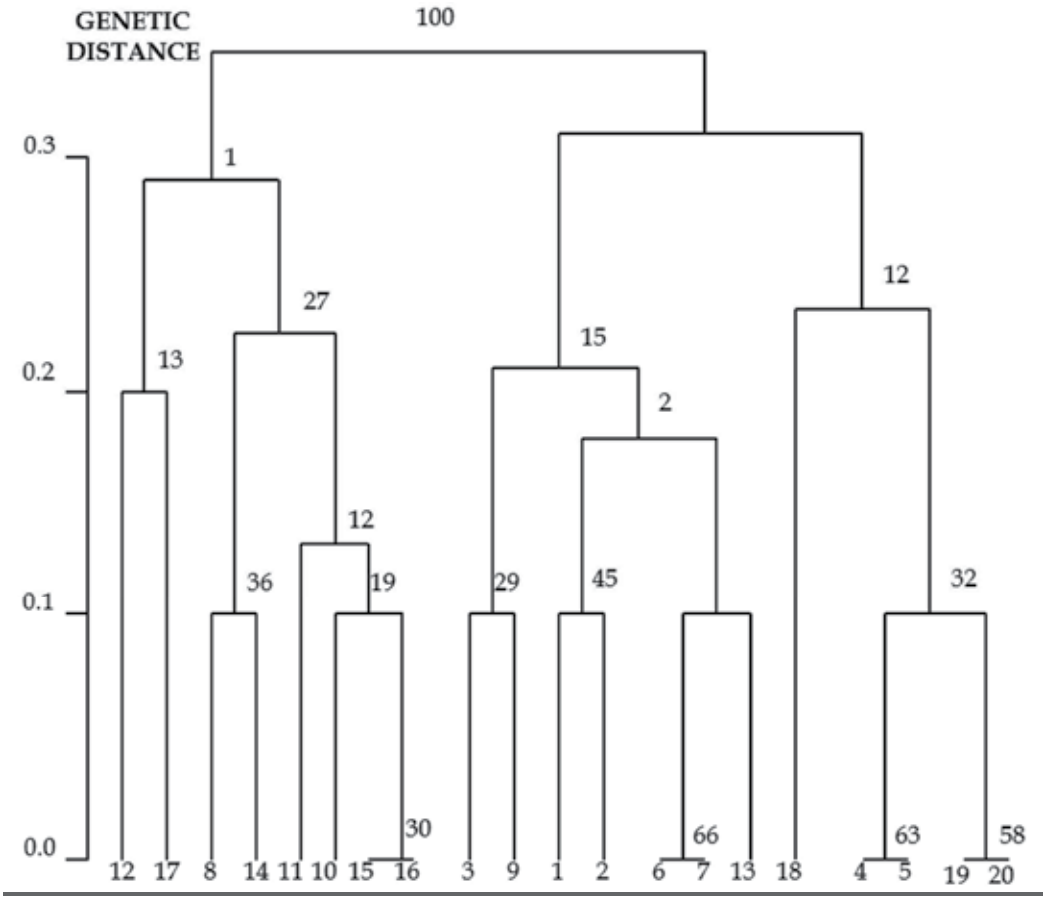


Fig. 6. Dendrogram obtained by the RAPD technique and general description according to physiological, morphological and genetic characteristics in garlic (*Allium sativum* L.) varieties cultivated in the Central Region of Mexico. Arms or branches of the dendrogram: 1. 'Perla' C-3-1/25, 2. 'Perla' C-37-1/8, 3. 'Coreano', 4. 'California', 5. 'Chino', 6. 'Criollo Aguascalientes', 7. 'Español', 8. 'Cortazar', 9. 'Positas', 10. 'Pepita', 11. 'Massone', 12. 'Durango', 13. 'Chileno', 14. 'Hermosillo', 15. 'Sonora', 16. 'Napuri', 17. 'Nicaragua', 18. 'Ixmiquilpan', 19. 'Pata de Perro' 20. 'Guatemala'.

Aguascalientes'); these have lost their potential yield because they have not been subjected to an appropriate selection process. Because we do not know the source of most garlic varieties and cultivars used in this work, we cannot establish relationships among their origins.

9. *In vitro* propagation

Most Mexican garlic cultivars are somewhat susceptible to pests and diseases. Furthermore, garlic is a seedless plant that could carry diseases to the next generation through vegetative propagation. "Seed cloves" per hectare range from 200,000 to 250,000 for typical plant density; therefore, totally clean vegetative material may be too difficult to be generated. The solution for this kind of problems is considered under biotechnological view such as *in vitro* culture so that we can obtain homogeneous healthy plants. *In vitro* bulbils after four years may produce, depending on the cultivar, from 1,400 to 10,700 bulbils ready to be used as "seed" (Burba, 1993). A recent report mentioned plants produced from cloves cultivated *in vitro* with 1 mg L^{-1} TDZ-1 (Thidiazuron), 1 mg L^{-1} GA_3 (Gibberelic acid) and coconut milk (Lagunes, 2009). *In vitro* garlic plants were also obtained on MS supplemented with 2.0 mg L^{-1} 2iP (2-isopentenyl adenine), 0.1 mg L^{-1} NAA (Naphthalene acetic acid) and 30 g L^{-1} sucrose (Mujica et al., 2008). Basal plate from cloves has the highest callus production as compared to leaves, stem segments, pedicels and aerial bulbils (Rabinowitch & Brewster, 1990). Another report mentioned that callus formation from 'Rojo de Cuenca' was the best on media having BA and NAA; furthermore, high BA concentration promoted adventitious shoot formation, but did not show influence on callus formation (Barandiaran et al., 1999). Callus was also obtained from leaves exposed to 0.3 a 0.5 mg L^{-1} 2,4-D (Fereol et al., 2002).

Our own results showed that explants about 5 mm^2 of clove sections were enough to regenerate *in vitro* whole plants from 'C-3-1/8' and 'C-37-1/25', 'Chino' and 'Coreano' garlic genotypes. The first step began with explants for root production, that were placed into MS medium supplemented with combinations of 0.15 mg L^{-1} 2,4-D, 5 mg L^{-1} adenine, 1.4 mg L^{-1} 4-amino-3,5,6-trichloropicolinic acid (Pichloram) and 1 mg L^{-1} 6-(γ,γ-dimetilamino) purine (2iP) (Table 4). MS medium also contained 30-50 g sucrose and 3.5 g L^{-1} Phytagel. Cloves were soaked in 70% ethanol 1min; after that, were transferred to 20% commercial bleach for 20 min. Then, cloves were rinsed thrice with dH_2O, placed back to 70% ethanol 1 min and rinsed again. Incubation conditions were: 28 ±2 °C and 16/8 h photoperiod. Adventitious roots were used to produce protocorms and protocorm-like bodies. These protocorms were cultured for 30 weeks on MS supplemented with 1 mg L^{-1} IAA and 5 mg L^{-1} adenine.

9.1 Protocorm formation

Protocorms 0.5-1 cm merged from root tips after 8 weeks on MS supplemented with 1.5 mg L^{-1} 2,4-D and 5 mg L^{-1} adenine (Table 4). Then they were placed into basal MS during 3 weeks, before protocorms were inoculated into four regeneration media at 18 ±4 °C and 16/8 h photoperiod during six weeks (Capote et al., 2000; Robledo-Paz et al., 2000; Quintana-Sierra et al., 2005). Protocorm and protocorm-like structures were both light and dark-green colored, compact and easily detachable (Fig. 7a, b and c), somehow similar to organogenic callus from *A. cepa* (Van der Valk et al., 1992; Zheng et al., 1998). 'Chino' and 'Coreano' protocorms were even more easily detached, dark-green colored as compared to 'Perla' derived genotypes; Novák et al., (1986) also found differential pigmentation among genotypes. Media supplemented with Pi (1.4 Mg/L^{-1}) and 2iP (1 Mg/L^{-1}) only promoted long roots.

9.2 Microbulbil formation

The treatment supplemented with 2,4-D showed the highest number of protocorm formation (Table 4); when the treatment included IAA and adenine, 'Chino' and 'Coreano' doubled to genotypes 'C-3-1/8' 'C-37-1/25'. Other treatments induced root formation. This different varietal response was found by Capote et al., (2000) and Quintana-Sierra et al., (2005) for *Allium cepa* and would be related to differences in sensibility to growth regulators (Fehér et al., 2003). Microbulbils placed, during three weeks, on basal MS (no regulators) increased their size (Fig. 7d), but after 30 weeks they developed into bulbils 1 cm diameter (Fig. 7e and 7f) and finally grew into whole plants.

Media (%)	Growth regulators (mg/$^{-1}$) Auxins	Cytokinins	Protocorm induction (%)	Number of observed structures ´C-3-1/8´	C-37-1/25´	Chino´	´Coreano´
Protocorms							
MS 100	2,4-D (1.5)	Adenine (5)	90	4	4	8	8
MS 100	Pi (1.4)	2iP (1)	0	----	----	----	----
Microbulbils							
MS 100	0	0	20	----	----	----	----
MS 100	IAA (1)	Adenine (5)	80	2-4	2-4	4-8	4-8
MS 100	NAA (1)	Kin (2)	0	----	----	----	----
MS 100	IAA (1)	Kin (2)	0	----	----	----	----
MS 100	Pi (1.4)	2ip (1)	0	----	----	----	----

Table 4. Effect of growth regulators on protocorm and microbulbil formation of four garlic 12 genotypes. IAA: Indole acetic acid, NAA: Naphthalene acetic acid, Kin: Kinetin, 2,4-D: 2,4-13 Dichlorophenoxyacetic acid, Pic: Pichloram, 2iP: 2-isopentenyl adenine.

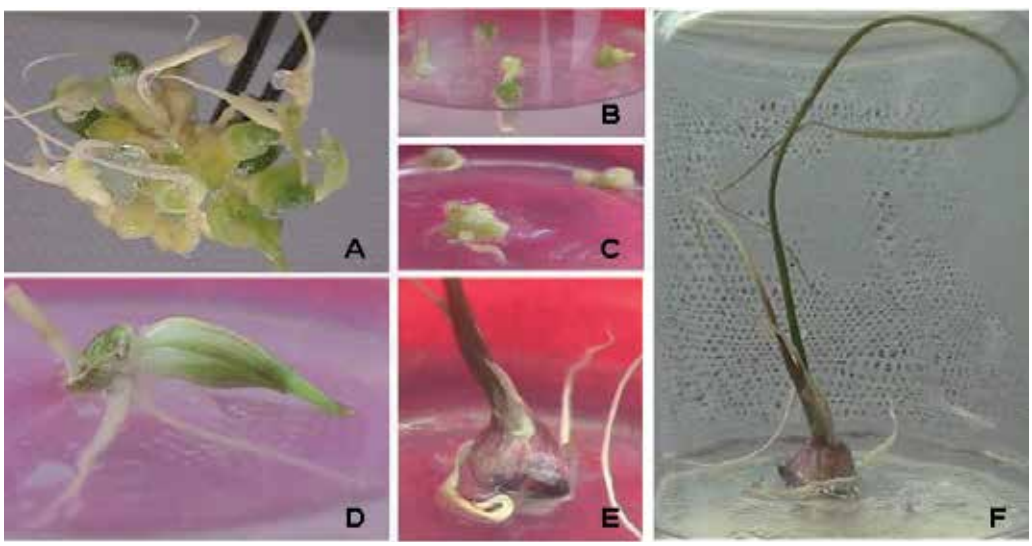


Fig. 7. *In vitro* regeneration of garlic plants (*A. sativum* L.). A) Protocorms regenerated from root tips, B) 'Chino' protocorms, C) 'Perla' protocorms, D) Microbulbil, E) and F) 'Chino' whole plant.

10. Individual selection for breeding Mexican garlic cultivars

Garlic in some cases may produce inflorescences but infertile seed; hence, crosses are not possible. Sometimes, bulbs o bulbils develop onto inflorescences (Brewster, 1994). Individual selection on best plants (yield or quality) has been used for breeding (Heredia and Heredia, 2000; González, 2006; Con, 1997). The CAEPAB group worked with individual selection from 'Perla' clones and 'Chileno' having heads with fewer cloves than the average for the original cultivar taking into account also: head size, vigor (hardiness) and plant healthiness. This initial work led to obtain two garlic cultivars: 'San Marqueño' from line 'C-37-1/8' (Macías et al., 2009) and 'Diamante' from line 'CAL-RN-11-1-1-2-4' derived from an Aguascalientes-Zacatecas collection (Macías & Maciel, 2003). In brief methodology for garlic breeding:

1. Bulb collection of promissory plants from the fields of outstanding growers of Aguascalientes and Zacatecas (May 1999).
2. Bulbs were planted in the experimental fields at CAEPAB, in order to check all of the collected material under the same growing conditions.
3. Evaluation and selection of garlic plants during 6-8 years (Table 6).
4. Storage of best clones at the germplasm bank (CAEPAB).
5. "Seed" production, enough to be transferred to growers for commercial validation.
6. After validation, best genotypes, having consistent yield results through time, are released to farmers.

Clones obtained through this kind of breeding are grown by farmers from Aguascalientes and Zacatecas (Macías et al., 2009). Nowadays, 'San Marqueño' and 'Diamante' garlic are demanded in Europe because of their high quality that makes them suitable to be exported (Fig. 8). These clones have their optimal conditions at 2000 meters over the sea level, on loamy soils, well drained, without salinity or pedregosity. Lab test are encouraged to check for soil pathogens that may reduce yield.

Evaluation	San Marqueño		Diamante	
Year	Yield (kg/ha)	SD	Yield (kg/ha)	SD
1	16,072	1,975	18,285	3,470
2	19,296	3,115	21,785	3,828
3	16,148	2,778	22,254	4,144
4	15,689	2,537	22,323	3,188
5	17,735	3,510	20,321	3,999
6	17,064	2,571	20,731	3,304
7			16,405	3,383
8			24,043	4,336

Table 6. Evaluation of 'San Marqueño' and 'Diamante' garlic cultivars.

Fig. 8. Garlic cultivars from the Mexican breeding program (CAEPAB): A) 'San Marqueño' ('Perla type'), B) 'Diamante' ('California' type) and C) Packaging box for exportation.

11. Conclusion

Horticulturists around the world look for answers from experimental stations to problems such as low yields, pests, diseases and quality defects. Similarly, garlic growers from Central Mexico have been in contact with institutions such as CAEPAB, PRODUCE-Ags and ITEL, in order to agree on agronomic and biotechnological research that may be applied to their fields. Original garlic genotypes from these growers and some other introduced to Mexico were the source for new cultivars and promissory genotypes developed by CAEPAB. Some of these that were analyzed showed a good correlation between bulb sizes and clove weight against bulb weight. It was also found that Mexican genotypes have a wide variety of on clove size and number that reflects a good genetic pool for breeding through individual selection for this seedless plant. Some other characteristics are qualitative that may have positive impact for worldwide market demands. For example, it was found for late cultivars the longest storage life.

Garlic biotechnology was also directed to characterize of Mexican cultivars. Molecular and cytogenetic characterizations for these cultivars may help to identify and to register, unambiguously, cultivated varieties. Molecular analysis such as AFLP´s or RAPDs is required in order to establish genetic uniqueness or relatedness. So far, only RAPDs have been performed but this kind of work is not concluded yet. Another biotechnological application is to produce *in vitro* pathogen-free vegetative material for massive propagation that may be released to growers. Protocorm and protocorm-like bodies were produced *in vitro* before they grew into bulbils and whole plant. Nevertheless, massive propagation of garlic has not been achieved as desired. Therefore, garlic biotechnology is going at slow pace. Finally, breeding through individual selection allowed releasing cultivars appropriate for national and international demands.

12. Acknowledgements

The authors would like to thank Consejo de Ajos de Aguascalientes A. C. (Refugio Lucio Castañeda, Gerardo Vargas Prieto, Otilio Torres Robles, and Narváes brothers) for the facilities provided (in their cultivation áreas) to carried out some research activities here showed.

13. Reference

Al-Zahim, M., Newbury, H.J. & Ford-Lloyd, B.V. (1997). Classification of genetic variation in garlic (*Allium sativum* L.) reveled by RAPD. *The American Society for Horticultural Science.*Vol. 32, No. 6, pp. 1102-1104, ISSN: 0018-5345.

Al-Zahim, M.A., Ford-Lloyd, B.V. & Newbury, H. J. (1999). Detection of somaclonal variation in garlic (*Allium sativum* L.) using RAPD and cytological analysis. *Plant Cell Reports,* Vol. 18 , pp. 473-477.

Barandiaran, X., Di Pietro, A. & Martin, J. (1998). Biolistic transfer and expression of a *uid* reporter gene in different tissue of *Allium sativum* L. *Plant Cell Reports,* Vol. 17, No. 6, pp. 737-741, ISSN: 0721-7714.

Battaglia, E. (1963). Mutazione cromosomica e cariotipo fondamental in *Allium sativum* L. Caryologia. pp.16: 1-46.

Bradley, K.F., Rieger, M.A. & Collins, G.G. (1996). Classification of Australian garlic cultivars by DNA fingerprinting. *Australian Journal of Experimental Agriculture*, Vol. 36, No. 5, (1996), pp. (613-618), ISSN: 0816-1089.

Brewster, J. L. (1994). *Onions and other vegetable alliums*. (2nd Edition). Crop production science in horticulture series. ISBN: 9781845933999. 2008.

Burba, L. (1993). Manual de producción de semillas hortícolas. Semillas de ajo. 1ª Ed., J*osé O. Valderrama, Centro de Información Tecnológica.* Vol. (9). No. 2, pp. 142-163, ISSN: 0716-8756.

Capote, R.A., Fundora, M.Z. y Pérez, D.O. (2000) Estudio de la variabilidad inducida en células y plántulas de cebolla (*Allium cepa* L.) cv Caribe-71 regeneradas *in vitro. Elfos scientiae.* Vol. 17, No. 4, (2000), pp. 241-246.

Chávez, C.M, Valenzuela, C.P., Fierros, L.G.A. & Maldonado, N.L.A. (2008). Efecto de métodos y densidades de siembra en la producción de dos variedades de ajo jaspeado en la sierra baja de sonora. XI Congreso Internacional en Ciencias Agrícolas. Universidad Autónoma de Baja California, Mexicali, Baja California. p. 381-385.

Choi, H.S., Kim, K.T., Ahn, Y.K., Kim, D.S. & Woo, J.G. (2003). Analysis of genetic relationships in garlic germoplasm and fertile garlic by RAPD. *Journal of the Korean Society for Horticultural Science.* Vol. 44, No 5, pp. 595-600, ISSN: 0253-6498

Con, L. (2009). Garlic breeding in Cuba. De: http://www.actahort.org/books/433/433_25.htm, 24 sep.

Doyle, J. J. & J. L. Doyle (1990) A rapid total DNA preparation procedure for fresh plant tissue. Focus 12:13-15.

Eom, E. & Lee, D. (1999). Characterization of chromosomal DNA polymorphism in Korean cultivars of *Allium sativum* L. *Journal of Plant Biology.*Vol.42, No.2, pp: 159-167.

FAOSTAT. (2011). FAOSTAT – Crop Statistics. *In:* FAOSTAT URL. Access: 09-12-2011. http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor

Fehér, A., Pasternak, T.P. & Dusits, D. (2003). Transition of somatic plant cells to an embryogenic state. *Plant Cell Tissue and Organ Culture.* Vol. 74, pp. 201-228. ISSN.

Fereol, L., Chovelon, V., Causse, S., Michaux-Ferriere, N & Kahane, R.R. (2002). Evidence of a somatic embryogenesis process for plant regeneration in garlic (*Allium sativum* L.*) Plant Cell Reports.* Vol. 21.

Ferrer, E., Linares, C. & González, J.M. (2000). Efficient transient expression of the β-glucoronidase reporter gene in garlic (*Allium sativum* L.). Vol. 20, No. 8., pp. 869-874, DOI: 10.1051/Agro:200101.

Figliuolo, G., Candido, V., Logozzo, G., Miccolis, V. & Spagnoletti-Zeuli, P.L. (2001). Genetic evaluation of cultivated garlic germplasm (*Allium sativum* L. and *A. ampeloprasum* L.). *Kluwer Academic Publishers.* Vol. 121, pp. 325-234.

García-Lampasana, S., Martínez, L. & Burba, J.L. (2003). Genetic diversity among Argentinean garlic clones (*Allium sativum* L.) using AFLP (Amplified Fragment Length Polymorphism). *Euphytica: Biomedical and Life sciences.*Vol. 132. No. 1, pp. 115-119, DOI: 10.1023/A:1024606004596.

González, A. (2006). "AKUKELI" una nueva variedad de ajo rosado. *Agricultura Técnica*, Vol.66, No.2, pp.210-215.

Haque, M.S., Wada, T. & Hattori, K. (1997) High frequency shoot regeneration and plantlet formation root tip of garlic. *Plant Cell Tissue and Organ Culture*, Vol.51, No.3, pp. 233, DOI:10.1023/A:1005753624853.

Heredia, G.E. & Delgadillo, S. (2002). El ajo en México. Celaya, Gto; México. 37-45 p.

Heredia, Z.A. & Heredia, G.E. (2000). Mejoramiento genético de ajo en el INIFAP. *En: El ajo en México, origen, mejoramiento genético, tecnología de producción*. Libro técnico Núm. 3. México: Secretaría de Agricultura, Ganadería y Desarrollo Rural. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Centro de Investigación Regional Centro. Campo Experimental Bajío, pág. 29-32.

Ipek, M., Ipek, A. & Simon, P.W. (2003). Comparison of AFLPs, RAPD markers, and isozymes for diversity assessment of garlic and detection of putative duplicates in germplasm collections. *J. Amer. Soc. Hort. Sci.* Vol.128, No.2, pp. 246-252.

Ipek, M., Ipek, A. & Simon, P.W. (2006). Sequence homology of polymorphic AFLP markers in garlic (*Allium sativum* L.). *Genome,* Vol. 49, pp. 1246-1255, DOI:10.1139/G06/092

Ipek, M., Ipek, A., Almquist, S.G. & Simon, P.W. (2005). Demonstration of linkage and development of the firs low-density genetic map of garlic, based on AFLP markers. *Theorical and Applied Genetics,* Vol. 110, No.2, pp. 228-236, ISSN: 00405752.

Khar, A., Yadav, R.C., Yadav, N. & Bhutani, R.D. (2005) Transient *gus* expression studies in onion (*Allium cepa* L.) and garlic (*Allium sativum* L.). *Akdeniz Universitesi Ziraat Falultesi* Dergisi.Vol.18, No.3, pp. 301-304.

Kondo, T., Hasegawa, H. & Suzuki, M. (2000). Transformation and regeneration of garlic (*Allium sativum* L.) by *Agrobacterium*-mediated gene transfer. *Plant Cell Reports*, Vol.19, No.10, pp. 989-993. DOI: 10.1007/s002990000222.

Koul, A.K., Gohil, R.N. & Langer, A. (1979). Prospects of breeding improved garlic in the light of its genetic and breeding systems. *Euphytica,* Vol.28, No.2, pp. 457-464. DOI: 10.1007/BF00056605.

Kirk, J.T., Rees, O.H. & Evans, G. (1970). Base composition of nuclear DNA within the genus *Allium*. *Heredity*. Vol. 25, pp. 507-512, DOI: 10.1038/hdy.1970.59

Lagunes, F.E. (2009). Transformación genética de ajo (*Allium sativum* L.) mediante *Agrobacterium tumefaciens*. Tesis de maestría. Colegio de Postgraduados, Campus Montecillo, Montecillo, Texcoco, Edo. de México. p.95.

Levan, A., Fredga, K. & Sandberg, A. (1964). Nomenclature for centromeric possition on chromosomes. *Hereditas,* Vol.52, pp. 201-220, DOI: 10.1111/j.1601-5223.1964.tb01953.x.

Macías, D.R., Grijalva, C.R.L. & Robles, C.F. (2010).). Productividad y calidad de variedades de ajo (*Allium sativum* L.) bajo condiciones desérticas en Caborca, Sonora. *BIOtecnia.* Vol. 12, No.1, (2010), pp.44-54.

Macías, V.L.M. & Maciel, P.L.H. (2009). Mejoramiento genético de diferentes tipos de ajo para la región productora de Aguascalientes. *In*: Memoria X Seminario de Investigación. Universidad Autónoma de Aguascalientes. Aguascalientes, Ags. México. 214-217 p.

Macías, V.L.M., Maciel, P.L.H. & Silos E.H. (2010). Diamante: variedad de ajo tipo California para el Centro Norte de México.INIFAP, CIRNOC, Campo Experimental Pabellón, Pabellón de Arteaga, Ags. México. 40 p. (Folleto técnico No. 41). ISBN: 978-607-425-509-6.

Moriconi, D.N., Burba, J.L. & Izquierdo, J. (1991). Manual de intercambio y propagación de germoplasma de ajo a través de microbulbillos. *FAO Online Catalogues: F02-Plant propagation*. Fiche No. 322397, 45 p. (1991).

McCollum, G.D. (1987). Onion and allies, pp. 186-190. In N.W. Simmonds (ed.) Evolution of crop plants.

Mujica, H., Sanabria, M.E., Mogollón, N. & Perozo, Y. (2008). Formación *in vitro* del bulbo del ajo morado (*Allium sativum* L.). *Revista de la Facultad de Agronomía (LUZ).* Vol. 25, No. 2, pp. 197-210, ISSN: 1690-9763.

Murashige, T. & Skoog, F. (1962). A revised medium for rapid growth and bioassay with tobacco tissue cultures. *Physiologia Plantarum,* Vol. 15, No 43, pp. 437-497, DOI: 10.1111/j.1399-3054.1962.tb08052.x . Article first published online: 28 APR 2006.

Nei, M. & Li, W. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA*. Vol. 76, No. 10, pp. 5256-5273. (august 1979).

Novák, F.J., Havel, L. & Dolezel, J. (1983). *Allium In* Crop Species. Hand book of Plant Cell Cultures. Techniques and Applications. Inst. Exp. Botany. Czechoslovak Acad. Sci. Praha. 419-455 p.

Panzera, F., Gimenez-Abian, M.I., López-Saez, J.F., Jiménez-Martín, G., Cuadrado, A., Shaw, P.J., Beven, A.F., Canovas, J.L. & De la Torre, C. (1996). Nucleolar organizer expresión in *Allium cepa* L. chromosomes. *Chromosoma,* Vol. 5, No.1, pp.12-19. DOI: 10.1007/BF02510034.

Pardo, A., Hernández, A. & Méndez, N. (2009). Caracterización molecular de siete clones de ajo (*Allium sativum* L.) mediante la técnica RAPD. *Bioagro*. Vol. 21, No. 2, pp. 81-86, ISSN: 1316-3361.

Paredes C. M., Becerra V. & González A. M. I., (2008). Low genetic diversity among garlic (*Allium sativum* L.) accessions detected using random amplified polymorphic DNA (RAPD). Chilean Journal of Agricultural Research, Vol. 68, No. 1:3-12.

Peschke, V. & Phillips, R.L. (1992). Genetic implications of somaclonal variation in plants. *Advances in Genetics*, Vol. 30, pp. 41-45, doi:10.1016/S0065-2660(08)60318-1.

Quintana-Sierra, M. E., A. Robledo-Paz., A. Santacruz-Varela., M. A. Gutiérrez-Espinosa., G. Carrillo-Castañeda & J. L. Cabrera-Ponce, (2005). Regeneración *in vitro* de plantas de cebolla (*Allium cepa* L.). Agrociencia 39: 647-655.

Rabinowitch, H. D. & Brewster J. M., (1990). Onions and Allied Crops. Botany, Physiology and Genetics. Vol. I CRC press. Inc. Boca Raton, Florida. U.S.A. 273 p.

Robledo-Paz, A., Villalobos-Arámbula, V.M. & Jofre-Garfias, A.E. (2000). Efficient plant regeneration of garlic (*Allium sativum* L.) by root tip culture. *In Vitro Cellular & Developmental Biology,* Vol. 36, No. 5, pp. 416-419, DOI: 10.1007/s11627-000-0075-6

Robles P. J., Armenta C. R. A., & Valenzuela C., (2006). México en el Contexto Global de la Producción de Ajo. En Memorias, Seminario Técnico: Tecnología para la producción de ajo en la sierra de Sonora. Universidad de Sonora-INIFAP-Fundación Produce Sonora. p. 7-14.

Molina, M.L.G. & Rosales, L.F.U. (2007). Diversidad genética de poblaciones de ajo (*Allium sativum* L.) cultivadas en Guatemala, definida por marcadores de ADN. *Agronomía Mesoamericana,* Vol. 18, No. 1, pp. 85-92, ISSN: 1021-7444

Sato, S., Hizume, M. & Kawamura, S. (1980). Relationship between secondary constrictions and nucleolus organizing regions in *Allium sativum*. *Protoplasma,* Vol. 105, No. 1-2, pp. 75-85, DOI: 10.1007/BF01279851

Shasany A. K., Ahirwar O. P., Kumar S. & Khanuja S. P., (2000). RAPD analysis of phenotypic diversity in the Indian garlic collection. *Journal of Medicinal and Aromatic Plant Sciences*. Abstr. 22(1B): 586-592.

SIAP. 2011. (http://reportes.siap.gob.mx/Agricola_siap/ResumenProducto.do)

Spangenberg, G. (1997). Transformation methods and analysis of gene expression in transgenic plants. ICGEB-Course. Embrapa. Brasilia, Brazil. 59-65 p.

Trejo P. P. (2006). Presentación, II Foro Nacional del Ajo. Memorias. Gobierno de Zacatecas: INIFAP, Fundación Produce, Zacatecas, SAGARPA, FIRA, Consejo Estatal de Productores de Ajo de Zacatecas A.C. Zacatecas, Zacatecas. México. P. 9-13.

Shigyo, M., Wako, T., Kojima, A. Yamauchi, N. & Tashiro, Y. (2003). Transmission of allien chromosomes from selfed progenies of a complete set of *Allium* monosomic additions: the development of a reliable method for the maintenance of a monosomic addition set. Genome, Vol. 46, No. 6, pp. 1098-1103, DOI: 10.1139/G03-075

Van Heusden, A.W., Van Ooijen, J.W., Vrielink, R., Ginkel, V., Verbeek, W.H.J., Wietsma, W.A. & Kik, C. (2000). A genetic map interspecific cross in *Allium* based on amplified fragment length polymorphism (AFLP) markers. *TAG Theorical and Applied Genetics,* Vol. 100, No. 1, pp. 118-126, DOI: 10.1007/s001220050017

Van der Valk, P., Scholten, O. E., Verstappen, F., Jansen, R.C. & Dons, J.J.M. (1992). High requency somatic embryogenesis and plant regeneration from zygotic embryo-derived callus cultures of three *Allium* species. *Plant Cell Tissue and Organ Culture*, Vol. 30, No. 3, pp. 181-191, DOI: 10.1007/BF00040020

Verma, S. C. & R. K. Mittal, (1978). Chromosome variation in the common garlic. Cytologia. 43: 383-396.

Xu, P., Liu, H., Gao, Z., Srinives, P. & Yang, C. (2001). Genetic identification of garlic cultivars and lines by using RAPD assays. *ISHS Acta Horticulturae* 555: *II International Symposium on Edible Alliaceae,* pp. 213-220.

Zheng, S., Henken, B., Sofiari, E., Jacobsen, E., Krens, F.A. & Kik, C. (1998). Factors influencing induction, propagation and regeneration of mature zygotic embryo-derived callus from *Allium cepa*. *Plant Cell Tissue and Organ Culture, Vol.* 53, No.2, pp. 99-105, DOI: 10.1023/A: 1006034623942

Edited by Mahmut Çalışkan

Genetic diversity is of fundamental importance in the continuity of a species as it provides the necessary adaptation to the prevailing biotic and abiotic environmental conditions, and enables change in the genetic composition to cope with changes in the environment. Genetic Diversity in Plants presents chapters revealing the magnitude of genetic variation existing in plant populations. The increasing availability of PCR-based molecular markers allows the detailed analyses and evaluation of genetic diversity in plants and also, the detection of genes influencing economically important traits. The purpose of the book is to provide a glimpse into the dynamic process of genetic variation by presenting the thoughts of scientists who are engaged in the generation of new ideas and techniques employed for the assessment of genetic diversity, often from very different perspectives. The book should prove useful to students, researchers, and experts in the area of conservation biology, genetic diversity, and molecular biology.

ISBN 978-953-51-4325-3

Photo by HT-Pix / iStock

IntechOpen