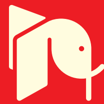
Next Generation Plant Breeding

Edited by Yelda Özden Çiftçi



NEXT GENERATION PLANT BREEDING

Edited by **Yelda Özden Çiftçi**

Next Generation Plant Breeding
http://dx.doi.org/10.5772/intechopen.71234
Edited by Yelda Özden Çiftçi

Contributors

Kaiyue Liu, Rongling Wu, Wenhao Bo, Lina Wang, Libo Jiang, Santosh Kumar Upadhyay, Shivi Tyagi, Alok Sharma, Yee-Han Chan, Kah-Yung Bernard Kah Yung Leong, Wan Muhamad Asrul Nizam Wan Abdullah, Swee Hua Erin Lim, Kok-Song Lai, Mohd Shamshad, Achla Sharma, Zhaoming Qi, Shiyu Huang, Jingyao Yu, Candong Li, Qianchun Gong, Zhengong Yin, Xinyu Wang, Hongtao Qin, Qingshan Chen, Dawei Xin, Chunyan Liu, Zhenbang Hu, Hongwei Jiang, Xinrui Mao, Huidong Qi, M. Govindaraj, Jihad Orabi

© The Editor(s) and the Author(s) 2018
The rights of the editor(s) and the author(s) have been asserted in accordance with the Copyright, Designs and Patents Act 1988. All rights to the book as a whole are reserved by INTECHOPEN LIMITED. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECHOPEN LIMITED's written permission. Enquiries concerning the use of the book should be directed to INTECHOPEN LIMITED rights and permissions department (permissions@intechopen.com).
Violations are liable to prosecution under the governing Copyright Law.

Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be foundat http://www.intechopen.com/copyright-policy.html.

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in London, United Kingdom, 2018 by IntechOpen
eBook (PDF) Published by IntechOpen, 2019
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, The Shard, 25th floor, 32 London Bridge Street
London, SE19SG – United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

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

Next Generation Plant Breeding
Edited by Yelda Özden Çiftçi

p. cm.

Print ISBN 978-1-78923-925-6

Online ISBN 978-1-78923-926-3

eBook (PDF) ISBN 978-1-83881-542-4

We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

3,750+
Open access books available

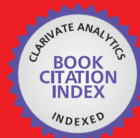
116,000+
International authors and editors

120M+
Downloads

151
Countries delivered to

Our authors are among the
Top 1%
most cited scientists

12.2%
Contributors from top 500 universities



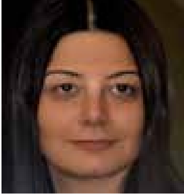
WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Meet the editor



Yelda Özden Çiftçi is a professor in the Department of Molecular Biology and Genetics at Gebze Technical University, Kocaeli, Turkey. She has gained much experience in the following areas of study: plant tissue culture and micropropagation with particular expertise in the effects of gaseous compounds (ethylene) on in vitro culture, PCR-based molecular markers, in vitro conservation and cryopreservation, genetic transformation of woody plants, abiotic stress tolerance, and plant-microbe interaction. She has authored or coauthored more than 75 papers, reviews, book chapters, books, and congress proceedings.

Contents

Preface X

Section 1 Marker Assisted Breeding 1

Chapter 1 **Marker-Assisted Breeding in Wheat 3**
Nana Vagndorf, Peter Skov Kristensen, Jeppe Reitan Andersen, Ahmed Jahoor and Jihad Orabi

Chapter 2 **Soybean Breeding on Seed Composition Trait 23**
Qi Zhaoming, Jingyao Yu, Hongtao Qin, Zhang Zhanguo, Shiyu Huang, Xinyu Wang, Mao Xinrui, Qi Huidong, Zhengong Yin, Candong Li, Xiaoxia Wu, Xin Dawei, Jiang Hongwei, Liu Chunyan, Hu Zhenbang and Chen Qingshan

Chapter 3 **Breeding Cultivars for Heat Stress Tolerance in Staple Food Crops 45**
Mahalingam Govindaraj, Santosh K. Pattanashetti, Nagesh Patne and Anand A. Kanatti

Chapter 4 **Functional Mapping of Plant Growth in Arabidopsis thaliana 75**
Kaiyue Liu, Wenhao Bo, Lina Wang, Rongling Wu and Libo Jiang

Chapter 5 **The Usage of Genomic Selection Strategy in Plant Breeding 93**
Mohd Shamshad and Achla Sharma

Section 2 RNA-seq and Gene Editing in Plants 109

Chapter 6 **Role of Next-Generation RNA-Seq Data in Discovery and Characterization of Long Non-Coding RNA in Plants 111**
Shivi Tyagi, Alok Sharma and Santosh Kumar Upadhyay

Chapter 7 **The CRISPR/Cas9 System for Crop Improvement: Progress and Prospects 129**
Kah-Yung Bernard Leong, Yee-Han Chan, Wan Muhamad Asrul Nizam Wan Abdullah, Swee-Hua Erin Lim and Kok-Song Lai

Preface

Today, the major challenges of agricultural production include unpredictable and extreme weather conditions that cause abiotic stress together with the severity and distribution of many pathogens and pests. Furthermore, climate change and decrease in the availability of arable land per person also cause severe limitations. Thus, we need more crop production to feed an ever-growing population.

The potential solution to this problem could be the development and utilization of modern genomics and biotechnological tools, which can provide a second Green Revolution as a complementary strategy to conventional breeding. This was also pointed out by Nobel Peace Prize Laureate Dr. Norman E. Borlaug in 2000 (Plant Physiol. 124:487–490). In accordance, currently, several successful examples that apply molecular tools have effectively contributed to developing cultivars in rice, millet, maize, several legumes, and horticultural crops.

In this sense, next-generation breeding techniques, including marker-assisted selection (MAS), next-generation sequencing (NGS), gene editing techniques, together with omic technologies, including genomics, transcriptomics, proteomics, and phenomics, have great potential not only to develop new crop cultivars but also to increase crop tolerance to abiotic and biotic stresses for sustainable agriculture. Among them, MAS could be regarded as a precise and efficient selection system that allows for recessive allele selection at early-stage selection and multiple genes pyramiding without traditional phenotypic evaluation for each trait. NGS and powerful computational pipelines have appeared as powerful tools both to identify many DNA sequence polymorphism-based genetic markers and to determine gene expression levels with RNA-sequencing (RNA-seq) methods within a short time and at low cost. However, there is still a necessity to increase the number of available high-quality whole-genome reference models and crop plant sequences to accelerate gene mapping and discovery to aid the application of MAS. Hence, information on genome sequences and gene functions is also a precondition for effective genome editing, which represents a new breeding technology that enables targeted or directional breeding with its low cost, precision, and rapidness in spite of the challenges (nontarget effects and regulations) that have to be resolved. Last but not least, it is obvious that the abovementioned next-generation plant-breeding techniques will be a frontier area of plant science and business for the cultivation of crops for tomorrow. Thus, we need a broad interdisciplinary approach both to transmit the information and experience obtained with model plants to crops and to achieve acceptance from the public and policymakers.

The present book consists of six chapters devoted to marker-assisted breeding, i.e., wheat, soybean, and vegetables, together with RNA-seq and gene editing in plants. All of these chapters are written by scientists who have experience in biotechnological plant breeding.

I would like to thank all of the authors not only for making a valuable contribution to recent resources in modern plant breeding but also for enabling this "Open Acess" publication to reach many scientists, teachers, and students working in the field. Moreover, I would also like to thank Intech Publishing Company, especially Ms. Ivana Glavic, Author Service Manager, and the technical editor of the book. Finally, I would like to thank my beloved husband, Mr. Burak Çiftçi; my two lovely sons, İbrahim Tuna and Kuzey Kenan; my dear parents, Özden and Çiftçi family; and my research team for all their patience, support, and courage.

Prof. Yelda Özden Çiftçi
Gebze Technical University
Department of Molecular Biology and Genetics
Kocaeli, Turkey

Section 1

Marker Assisted Breeding

Chapter 1

Marker-Assisted Breeding in Wheat

Nana Vagndorf, Peter Skov Kristensen,
Jeppe Reitan Andersen, Ahmed Jahoor and
Jihad Orabi

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74724

Abstract

Selection is an integral component in plant breeding, which ensures the progressive values of the breeding material, in terms of yield and quality. However, selection is influenced by the environment in any given growing season. The observed phenotype is a product of the genotype (G), the environment (E), and/or genotype × environment (G×E). Therefore, phenotypic selection is not always the best predirector of the genotype. Therefore, an environment-independent method is preferred by the breeder. The development of molecular markers in plants has facilitated marker-assisted selection (MAS). MAS requires the establishment of correlation between a desired trait such as disease resistance and molecular marker(s). This can be obtained, e.g., by phenotyping a genetic mapping population followed by QTL analysis. Initially, this process was slow due to the laborious nature of the first DNA molecular marker system, such as restriction fragment length polymorphism (RFLP). Later, with the discovery of various marker systems amenable to automation and the development of genotyping techniques and instruments, MAS has become a standard procedure in plant breeding. In wheat breeding, MAS helped to accelerate the introgression of many genes that contribute to improve quality and resistance.

Keywords: wheat, marker-assisted breeding, molecular markers, wheat diseases, wheat quality

1. Introduction

Wheat is one of the most important sources of food worldwide. Data from FAOSTAT indicate that the need is still growing, indicated by the steadily increasing yield since 1961 (**Figure 1**). The need for an enhanced wheat production combined with stagnation in the area cultivated (**Figure 1**) leads to a demand for a more effective and higher yielding wheat production.

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (cc) BY

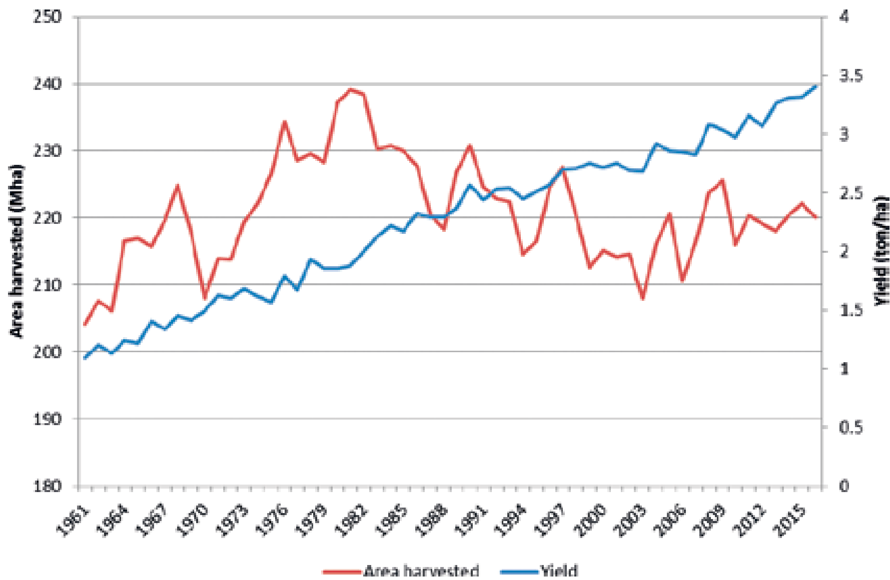


Figure 1. Development in wheat cultivation in the years 1961–2016. The primary y-axis displays the area harvested in Mha (red line) and the secondary y-axis displays the yield in tones pr. ha (blue line). Data from FAOSTAT.

Marker-assisted selection (MAS) or molecular breeding offers an opportunity to accelerate the traditional breeding. Traditional breeding is based on phenotypic selection of genotypes obtained from crosses. Genotype × environment (G×E) interaction is a common problem including time-consuming and costly procedures of phenotyping. By employing molecular markers, desirable genes can be fixated in early generations of the breeding program. In addition, molecular markers are unaffected by environmental conditions and are detectable in all stages of the plant growth. Scientists and breeders across the world implement MAS in breeding programs [1].

2. Linkage and molecular markers

MAS is based on the concept of genetic linkage between loci. This describes the tendency of loci located closely together on the same chromosome being more likely to be inherited together in a recombination event during meiosis. Thus, two alleles located very close on the chromosome will almost always be inherited together.

Molecular markers are used in MAS to highlight a place on the chromosome close to or in a specific gene of interest. The technique is based on detecting different alleles (polymorphisms) between several individuals. Due to genetic linkage, the molecular marker will reveal if the linked allele is present or not in a line. Several types of molecular markers exist, depending on the type of polymorphism. In today's MAS, markers detecting single nucleotide differences are usually employed.

3. Application of MAS in breeding for disease resistance in wheat diseases

Plant diseases are a major constraint in wheat production and significant resources are allocated to control various diseases. The relatively long growing season of winter wheat renders it vulnerable to a range of diseases and breeding for disease resistance is generally thought to be the first line of defence [2]. Disease resistance is generally separated into quantitative and qualitative resistance. Qualitative resistance is most often controlled by a single gene and follows the gene-for-gene hypothesis. Thus, an R gene in the host can specifically interact with an *Avr* gene in the pathogen to induce a defence reaction in the host. Qualitative resistance often mediates a complete resistance response, whereas quantitative resistance is regarded as an incomplete or partial resistance. This type of resistance is usually mediated by several minor genes, which are designated as a quantitative trait locus (QTL) [3]. Following the development of MAS, targeted pyramiding of several resistance genes in single lines is now possible. In the following sections, three severe diseases of wheat and correlated resistance genes are described. Common to these diseases is that fungicides are becoming less efficient. Hence, more effective approaches are desired.

3.1. Septoria tritici blotch

One of the most important foliar diseases in wheat is Septoria tritici blotch (STB), caused by the fungus *Zymoseptoria tritici* formally known as *Septoria tritici* (anamorph) and *Mycosphaerella graminicola* L (teleomorph) [4]. STB is a devastating disease causing massive yield losses worldwide every year in wheat. Severe epidemics can reduce wheat yield by 35–50% [5]. Symptoms of the disease are chlorotic lesions on the leaf with black fruiting bodies containing fungal spores. Breeding strategies over the years have primarily focused on breeding for higher yield, in turn increasing the susceptibility towards STB [6]. Disease control can be performed by delaying sowing time, probably due to less time in the autumn for *Z. tritici* to infect seedlings and produce inoculum. Furthermore, the application of fungicides and implementation of resistant cultivars in breeding programs are widely used [5]. In previous years, several studies have identified STB resistance genes using molecular techniques.

Stb1, *Stb2* and *Stb3* were the first qualitative genes for STB resistance to be named [7]. Prior to that, STB resistance was thought of as a quantitative, polygenic trait. *Stb1* was mapped to the long arm of chromosome 5B in the cultivar Bulgaria88 [8]. *Stb2* was mapped to the short arm of chromosome 1B in the cultivar Veranopolis [9]. Additionally, *Stb11* was mapped to the short arm of 1B [10]. However, no studies have included an allelism test of *Stb2* and *Stb11*. *Stb3* was mapped to the short arm of 7A in Israel493 [11]. *Stb6* was mapped to the short arm of chromosome 3A in the cultivar Flame [12]. This gene is the only STB-resistant gene found to possess a gene-for-gene relationship, in which a specific R gene in the host interacts with an *Avr* gene in the pathogen. This was demonstrated in a study where Flame was found to confer specific resistance towards the *Z. tritici* isolate IPO323 [12]. This study conforms to the original gene-for-gene model proposed by Flor [13]. *Stb6* was subsequently found to be one of the most abundant STB-resistant genes in European wheat [14]. In total, 18 *Stb* genes have been

identified and mapped using various molecular markers. Additionally, several QTL have been identified conferring STB resistance [7, 15]. **Table 1** summarises major STB-resistance genes together with linked markers suitable for MAS. Additionally, several QTL have been identified conferring STB resistance [7, 15].

One of the more promising resistance genes identified in recent years is *Stb16*. This gene was identified in synthetic hexaploid wheat lines, which represent a rich source of variation [23]. *Stb16* explained a high proportion of STB disease resistance and conferred resistance at the seedling stage to all tested *Z. tritici* isolates. Moreover, 20 tested isolates were all avirulent to this gene, indicating that *Stb16* confers broad-spectrum resistance. If this is the case, *Stb16* holds promise for future breeding of efficient and durable STB resistance.

In order to obtain the most resistant wheat variety, breeders should take a number of things into account. Since qualitative resistance genes often conform to the gene-for-gene hypothesis, they are readily overcome by the pathogen. Due to the high frequency of genetic recombination of *Z. tritici,* the specific recognition of R proteins by the host is lost [26]. Furthermore, the strong

Resistance gene	Marker type	Marker name	Location	Reference
Stb1	SSR	Xbarc74, Xgwm335	5BL	[8]
Stb2	SSR	Xwmc406, Xbarc008	1BS	[9]
Stb3	SSR	Xwmc83	7AS	[11]
Stb4	SSR	Xgwm111, Xgwm44	7DS	[16]
Stb5	SSR	Xgwm44	7DS	[17]
Stb6	SSR	Xgwm369	3AS	[12]
Stb7	SSR	Xgwm160, Xwmc219, Xwmc319	4AL	[18]
Stb8	SSR	Xgwm146, Xgwm577	7BL	[19]
Stb9	SSR	Xfbb226, XksuF1b	2BL	[20]
Stb10	SSR	Xgwm848	1D	[21]
Stb11	SSR	Xbarc008	1BS	[22]
Stb12	SSR	Xwmc219, Xgw313	4AL	[21]
Stb13	SSR	Xwmc396	7BL	Wheat gene catalogue
Stb14	SSR	Xwmc500, Xwmc632	3BS	Wheat gene catalogue
Stb16	SSR	Xgwm494	3DL	[23]
Stb17	SSR	Xhbg247	5AL	[23]
Stb18	SSR	Xgpw5176, Xgpw3087	6DS	[24]
StbWW	SSR	Xbarc119b	1BS	[25]

The name of the resistance gene, marker type, marker name, the location on the genome and the reference are indicated.

Table 1. An overview of the named and mapped genes for STB resistance.

selection pressure placed on the pathogens by one major resistance gene promotes the rise of new adapted races in the pathogen population [27]. An earlier study has proved that commercial cultivation of a highly resistant cultivar can result in loss of resistance towards STB. The intensive cultivation of the variety Gene in the 1990s in Oregon, US, resulted in resistance breakdown. Gene was found to be resistant to two specific isolates, which were avirulent to *Stb6* and *Stb10* [28]. Gradually, an adaptation of *Z. tritici* to one of the resistance genes occurred and the resistance was lost [26]. In general, it appears that most *Z. tritici* isolates used in earlier studies are virulent to almost all *Stb* genes [7]. This may indicate that *Z. tritici* easily can overcome single qualitative resistance genes. In contrast, the quantitative resistance is regarded as more durable. This is due to a lower selection pressure on the pathogen as a result of smaller resistance effects of individual QTL [3]. Furthermore, since quantitative resistance is often polygenic, the mutation of one gene does not necessarily break down disease resistance completely.

3.2. Fusarium head blight

Fusarium head blight (FHB) is an important disease in all wheat growing countries. Epidemics occur frequently, especially under seasons with regular rainfall [29]. The United States Department of Agriculture (USDA) has stated that FHB is the most devastating plant disease since the rust epidemics in the 1950s. FHB contaminates the grain with mycotoxins, in turn restricting its use for both animal and human consumption [30]. The disease is caused by several species of *Fusarium*; however, the predominant causal agent is the fungus *Fusarium graminearum* (teleomorph *Gibberella zeae*). The first symptoms of FHB on wheat plants occur shortly after flowering as diseased spikelets display premature bleaching. The bleaching usually spreads to the whole spike as the pathogen grows. When conditions are optimal for the pathogen, i.e., in a warm and moist environment, light pink coloured spores, called sporodochia, appear on individual spikelets. Later during the season, black fruiting bodies will appear. These are the sexual structures of the fungus, called perithecia. Disease progression results in shrinking and wrinkling of the grain inside the spike. As with the pathogen causing STB, *F. graminearum* produces both sexual and asexual spores: ascospores and macroconidia, respectively [30]. The major toxin produced by FHB in wheat is deoxynivalenol (DON). DON is a protein synthesis inhibitor also known as vomitoxin due to its negative impact on the digestive system of pigs. Several recommendations and restrictions have been made in order to keep DON levels sufficiently low in wheat for both animal and human consumption [31].

Chemical control and crop management are not sufficient to control FHB; thus, breeding resistant varieties plays a key role. Conventional breeding involves repeated testing of breeding lines under natural or artificial inoculations. This process is time-consuming, costly, and prone to influence by environment. Thus, it is relevant to supplement phenotypic selection with MAS for FHB resistance. [32]. FHB resistance is generally divided into three types: resistance to initial infection (type I), resistance to spreading of the pathogen in infected tissue (type II) and resistance to DON accumulation (type III) [33]. Several studies have demonstrated that FHB resistance is of quantitative nature [29]. Furthermore, the expression of resistance is highly dependent on the pathogen, the environment and the host [34], in turn complicating phenotypic selection. Several QTL for FHB resistance have been identified and

located during recent years [29]. The first QTL for type II resistance was identified in the spring wheat 'Sumai 3' on chromosome 3BS. This QTL was named *Fhb1* and characterised by molecular markers [35–37]. Recently, *Fhb1* was cloned from Sumai 3 and a pore-forming toxin-like (PFT) gene was found to confer FHB resistance [38]. *Fhb1* has been found to reduce FHB disease severity tremendously and MAS is employed to incorporate the resistance in breeding programs [29]. A QTL, named *Fhb2*, on chromosome 6BS was found to confer type II FHB resistance [39, 40]. Additionally, *Fhb4* was identified and located on chromosome 4B [41]. **Table 2** lists all FHB-resistant genes identified by molecular markers. Currently, breeders are pyramiding *Fhb1*, *Fhb2* and *Fhb4* in single breeding lines to obtain optimal FHB resistance [34]. Several additional QTL have been identified and located in numerous studies [29].

Resistance gene	Marker type	Marker name	Location	Reference
Fhb1	SSR	Xgwm493, Xgwm533	3BS	[42]
Fhb2	SSR	Xgwm133, Xgwm644	6BS	[40]
Fhb4	SSR	Xhbg226, Xgwm149	4B	[41]
Fhb5	SSR	Xgwm304, Xgwm415	5A	[43]
Fhb6	KASP	Wg1s_snp1	1AS	[44]
Fhb7	SSR	XsdauK66, Xcfa2240	7DS	[45]

Table 2. Overview of the FHB-resistant genes identified in wheat using molecular markers.

3.3. Wheat stripe rust (yellow rust)

Wheat stripe rust, mostly designated as 'yellow rust' (YR), causes major yield losses every year. The disease is caused by *Puccinia striiformis,* which belongs to the family *Pucciniaceae* of rust fungi. The most devastating epidemics occur in temperate areas with cool and humid summers or in warmer areas with cool nights. The fungus is heteroecious, i.e., it requires at least two hosts in order to proliferate. *P. striiformis* uses cereals as a primary host and *Berberis* spp. as a secondary host for sexual recombination. Typical, yellow stripes develop on the leaf in lesions. Spores continue to be produced as stripes spread longitudinally on the leaf. After the onset of senescence, *P. striiformis* will produce teliospores. Teliospores can infect the secondary host, *Berberis* spp., and initiate onset of pycnia infection of the *Berberis* leaf [46].

Breeding for YR resistance was initiated in 1905 by Biffen [47]. To date, more than 70 genes (*Yr* genes) conferring YR resistance have been identified [48]. Most of the catalogued genes confer seedling resistance, while relatively few confer adult plant resistance. In general, studies have shown that seedling resistance is conferred by single genes and the resistance is therefore easily overcome by the pathogen by mutations in virulence genes. Adult plant resistance is generally thought to be more durable [49]. High-temperature adult plant (HTAP) genes are expressed as the plants grow older and the weather becomes warmer [50]. HTAP genes confer a non-specific, quantitative resistance. Studies have proven that varieties with HTAP genes display resistance

Resistance gene	Marker type	Marker name	Location	Reference
Yr5	SSR	Xgwm501	2BL	[55]
Yr7	SSR	Xgwm526	2BL	[56]
Yr15	SSR	Xbarc8, Xgwm493	1BS	[57]
Yr18	CAPS	Cssfr6	7D	[58]
Yr36	SSR	Xgwm508, Xbarc136	6BS	[54]
Yr60	SSR	Xwmc776	4AL	[59]
Yr76	SSR	Xwmc11, Xwmc532		[60]
Yr78	SNP	IWA7257	6BS	[61]

Table 3. A selection of the genes conferring YR resistance identified by molecular markers.

to YR even after having been cultivated for 60 years [51]. Additionally, several studies have mapped QTL to all wheat chromosomes except chromosome 1D and 3A [49]. Commonly used resistance genes employed in wheat breeding programs include *Yr*18, *Yr*29 and *Yr*36 [52–54]. *Yr36* is tightly linked to *Gpc-B*1, a high-protein gene, rendering varieties with *Yr36* and *Gpc-B*1 useful in breeding for YR resistance and improved quality. **Table 3** lists a selection of *Yr* genes that have been characterised and mapped with molecular markers suitable for MAS.

Several incidences have been reported where *Yr* genes have been classified as ineffective. Some of the most widely used resistance genes including *Yr17* [62], *Yr27* [63] and *Yr31* [64] have recently lost resistance towards YR.

4. Marker-assisted wheat breeding for improving quality traits

Wheat is grown in large parts of the world and is used for animal feed or for a wide range of products such as pasta, biscuits, cakes and bread. The end-use quality differs greatly between wheat cultivars and is influenced by several traits, e.g., grain hardness, grain protein content, gluten content and composition and starch properties. Quality should therefore be an important focus in wheat breeding programs. However, wheat quality cannot be easily determined phenotypically, and different methods are preferred in different countries and industries. Methods for testing quality are typically time-consuming and costly and require relatively large amounts of grain, which is typically not available until late stages of breeding programs. Thus, markers for wheat quality traits can be very useful, as they enable screening of a high number of lines and can be used early in breeding programs [65, 66].

4.1. Grain hardness

Grain hardness influences milling, flour and end-use properties of wheat. Flour from grain with hard endosperm texture has higher water absorption than flour from soft grain and is therefore preferred for bread-making. A soft endosperm texture leads to less starch granule damage

during the milling and consequently to lower water absorption, which is preferred in the production of biscuits and cakes. Grain hardness is primarily controlled by the *Hardness* locus on chromosome 5DS. This locus consists of three small genes: *Pina-D1*, *Pinb-D1* (*Puroindoline a/b*) and *grain softness protein-1 (Gsp-1*). Wheat varieties with the wild-type alleles *Pina-D1a* and *Pinb-D1a* normally have soft grain, while deletions or other loss-of-function mutations in one or both *Pin* genes cause harder grain (**Table 4**) [67, 68]. *Pinb-D1* mutations are positively associated with many quality traits, but the alleles are not equally useful in breeding for improved quality. *Pinb-D1d* has been reported to have a lower effect on gluten quality and loaf volume than the *b*- or *c*-allele [69]. Alleles of *Pinb-D1* can be detected using PCR primers that target a specific mutation (*Pinb-D1b*), using a restriction enzyme on the amplified *Pinb-D1* gene (*Pinb-D1c*), or by sequencing the amplified gene (*Pinb-D1d-g*) [67, 70, 71].

Allele	Change in protein	Primer sequences, 5′–3′	PCR product	References
Pina-D1a	Wild-type	F: ATGAAGGCCCTCTTCCTCA R: TCACCAGTAATAGCCAATAGTG	448 bp	[73, 74]
Pina-D1b	Large deletion	F: ATGAAGGCCCTCTTCCTCA R: TCACCAGTAATAGCCAATAGTG	Null (0 bp)	[73, 74]
Pinb-D1a	Wild-type	F: ATGAAGACCTTATTCCTCCTA R: CTCATGCTCACAGCCGC**C**	240 bp	[70, 73]
Pinb-D1b	Gly to Ser pos. 46	F: ATGAAGACCTTATTCCTCCTA R: CTCATGCTCACAGCCGC**T**	240 bp	[70, 73]
Pinb-D1c	Leu to Pro pos. 60	F: ATGAAGACCTTATTCCTCCTA R: TCACCAGTAATAGCCACTAGGGAA	448 bp*	[67, 73]
Pinb-D1d	Trp to Arg pos. 44	F: TGCAAGGATTACGTGATGGA R: TCACCAGTAATAGCCACTAGGGAA	300 bp for pyrosequencing	[67, 71]
Pinb-D1e	Trp to stop codon pos. 39	F: TGCAAGGATTACGTGATGGA R: TCACCAGTAATAGCCACTAGGGAA	300 bp for pyrosequencing	[71, 75]
Pinb-D1f	Trp to stop codon pos. 44	F: TGCAAGGATTACGTGATGGA R: TCACCAGTAATAGCCACTAGGGAA	300 bp for pyrosequencing	[71, 75]
Pinb-D1g	Cys to stop codon pos. 56	F: TGCAAGGATTACGTGATGGA R: TCACCAGTAATAGCCACTAGGGAA	300 bp for pyrosequencing	[71, 75]

Wild type alleles confer soft endosperm; mutations confer hard endosperm. For additional alleles, see reviews [68, 72].*Digest with restriction enzyme *Pvu*II to cut other alleles into 264 bp and 184 bp. *Pinb-D1c* is not cut.

Table 4. Alleles of *Pina-D1* and *Pinb-D1* and the change in amino acid sequence of the encoded protein.

4.2. Gluten

The characteristic viscoelastic properties of wheat dough are due to a network of gluten proteins that is formed when flour is mixed with water. Thus, gluten is a major factor contributing to wheat quality. High grain protein content is typically associated with high quality, since roughly 80% of the grain protein is gluten [76]. However, both the amount and the composition of gluten affect wheat quality. Gluten consists of two types of proteins: polymeric glutenins and monomeric gliadins. Glutenins can be classified as low or high molecular weight (LMW or HMW) subunits, while gliadins can be classified as α, β, γ or ω types [77, 78]. The most important HMW glutenins, LMW glutenins, and gliadins are encoded by the *Glu-1*, *Glu-3* and *Gli-1* loci, respectively (**Table 5**). HMW glutenins generally have the largest impact on wheat quality. Each of the three *Glu-1* loci comprises two genes that can encode an x- or a y-type HMW subunit. In hexaploid wheat, only three to five of the HMW subunits are expressed (zero to two from *Glu-A1*, one to two from *Glu-B1*, and two from *Glu-D1*) [79]. The *Glu-1* alleles with the largest positive effect on baking quality are *Glu-D1d*, *Glu-A1a* or *Glu-A1b* and *Glu-B1al* [80, 81]. SDS-PAGE electrophoresis can be used to screen varieties for their HMW glutenin proteins. DNA markers have also been developed to discriminate between different alleles of *Glu-1*, *Glu-3* and *Gli-1* loci [82, 83]. For *Glu-A1* and *Glu-D1*, KASP markers are available that can be used to select varieties with the optimal alleles [84]. Each of the *Glu-3* loci (*Glu-A3*, *Glu-B3* and *Glu-D3*) contains several linked genes, and many alleles have been found for all three loci [85–89]. Markers are available for individual alleles of *Glu-A3* and *Glu-B3*, and multiplex PCR can be used to screen for certain combinations of alleles simultaneously [87]. However, the alleles of *Glu-3* loci with the largest effects are not consistent across studies [90–92]. The exact effects of the individual alleles on wheat quality traits are challenging to determine, since they can be influenced by genetic background, environment and G×E interactions [91, 93]. Furthermore, the alleles can have both additive effects and epistatic interactions [94, 95]. Ref. [93] showed that the *d*-allele of *Glu-B3* might increase the positive effects of the HMW loci *Glu-B1i* and *Glu-D1d*. The *Glu-A3b* or *d*-allele and *Glu-B3b*, *d*- or *g*-allele can possibly be used for improving dough strength and extensibility [90–92]. *Glu-B3i* has been reported to be positively associated with wheat quality in some lines and negatively associated in other lines. This discrepancy is possibly due to linkage with different *Gli-B1* alleles [90]. The *Gli-1* loci encode γ and ω gliadins and are linked to the *Glu-3* loci [96], while *Gli-2* loci encode α and β gliadins and are located on chromosome 6AS, 6BS and 6DS [78]. Overview of markers (including primer sequences) for more alleles of *Glu* loci and other quality genes can be found in [82].

4.3. Wheat-rye translocation and falling number

The wheat-rye translocation 1BL.1RS has been employed in many breeding programs as it carries resistance genes against powdery mildew and rusts. Markers for the resistance genes can be used to test for the absence or presence of the translocation in wheat varieties [100]. Alternatively, markers for *Glu-B3* or *Gli-B1* might be used (**Table 6**), since many wheat varieties with the 1BL.1RS translocation do not have these two loci, but instead can have the rye secalin locus *Sec-1* [96]. Therefore, wheat quality can be negatively affected by the translocation [101]. Additionally, the 1BL.1RS translocation can have a negative effect on falling number. Falling number is an indirect measure of α-amylase enzyme activity. The α-amylases are encoded by

Locus	Chr. arm	Primer sequences, 5′–3′	PCR product	References
Glu-A1	1AL	FAM: AAGTGTAACTTCTCCGCAACA VIC: AAGTGTAACTTCTCCGCAACG Common: GGCCTGGATAGTATGAAACC	FAM: ***Glu-A1a*** or ***Glu-A1b*** VIC: *Glu-A1c*	[84, 97]
Glu-B1	1BL	F: ACGTGTCCAAGCTTTGGTTC R: GATTGGTGGGTGGATACAGG	***Glu-B1al***: 447 bp Others: 0 bp	[98]
Glu-D1	1DL	FAM: ATAGTATGAAACCTGCTGCGGAG VIC: ATAGTATGAAACCTGCTGCGGAC Common: TACTAAAAAGGTATTACCCAAGTGTAACTT	FAM: *Glu-D1a* or others VIC: ***Glu-D1d***	[84, 99]
Glu-A3	1AS	F: TTCAGATGCAGCCAAACAA R: GCTGTGCTTGGATGATACTCTA	***Glu-A3b***: 894 bp Others: 0 bp	[86, 87]
		F: TTCAGATGCAGCCAAACAA R: TGGGGTTGGGAGACACATA	***Glu-A3d:*** 967 bp Others: 0 bp	
Glu-B3	1BS	F: ATCAGGTGTAAAAGTGATAG R: TGCTACATCGACATATCCA	**Glu-B3b***:* 1549 bp Others: 0 bp	[88]
		F: CACCATGAAGACCTTCCTCA R: CACCATGAAGACCTTCCTCA	***Glu-B3d***: 662 bp Others: 0 bp	
		F: CCAAGAAATACTAGTTAACACTAGTC R: GTTGGGGTTGGGAAACA	***Glu-B3g***: 853 bp Others: 0 bp	

Favourable alleles are marked in bold.

Table 5. Important HWM and LMW glutenin loci, their chromosomal location and primer sequences for detection of alleles with positive effects on wheat quality.

Locus	Primer sequences, 5′–3′	PCR product	References
Gli-B1 *(1BL.1RS)*	F: TGATCTGGCCACAAAGGGA R: CATTGGCCACCAATTCCTGT	***Gli-B1.1:*** 369 bp *Gli-B1.2* or 1BL.1RS: 0 bp	[96]
	F: TGATCTGGCCACAAAGGGC R: CATTGGCCACCAATTCCTGT	***Gli-B1.2:*** 397 bp *Gli-B1.1* or 1BL.1RS: 0 bp	
Rht-D1	F: GGTAGGGAGGCGAGAGGCGAG R: CATCCCCATGGCCATCTCGAGCTA	**Rht-D1b***:* 237 bp *Rht-D1a:* 0 bp	[106]
wbm	F: CCGTCACAAGATTTACAGGGTTG R: TTATGGATCTCTTTATGTCTGTGT	**High *wbm* expression**: 961 bp Others: 0 bp	[103]
Gpc-B1	F: TCTCCAAGAGGGGAGAGACA R: TTCCTCTACCCATGAATCTAGCA	***Gpc-B1***: 122 bp No *Gpc-B1*: 126 bp	[105]

Favourable alleles are marked in bold.

Table 6. Additional loci influencing wheat quality traits.

the loci *α-Amy-1*, *α-Amy-2* and *α-Amy-3* located on the homoeologous chromosome groups 6, 7 and 5, respectively. High falling number reduces the risk of pre-harvest sprouting, which has a considerable negative impact on quality. Environmental conditions around the time of harvest influence falling number, but it is also influenced genetically. The *b*-allele of the *Rht-D1* (*reduced height*) gene on chromosome 4D is correlated with increased falling number [102].

4.4. Other genes for improving quality

Ref. [103] identified a gene, *wheat bread making (wbm*), that was highly expressed in developing seeds of wheat varieties with good bread-making quality. Polymorphisms in the promoter region sequence were identified between good- and poor-quality varieties. The allele identified in the good quality varieties was positively associated with gluten and bread-making quality in CIMMYT (The International Maize and Wheat Improvement Center) germplasm [104].

Genes from wild wheat relatives might also be used for improving quality in modern cultivars. Backcrossing can be used to transfer the genes into breeding material. In this case, MAS is useful since offspring containing the desired genes easily can be detected, and linkage drag can be reduced. One example of such a gene is *Gpc-B1* (*grain protein content*), which was found in wild emmer (*Triticum turgidum* L. ssp. *dicoccoides*). This gene has been used for increasing grain protein content in both durum and common wheat [105]. Markers tightly linked to *Gpc-B1* were identified, but require digestion with restriction enzymes. Therefore, [105] recommends the use of the marker shown in **Table 6** for MAS, although it is not completely linked to *Gpc-B1*.

5. Conclusion and perspectives

Trait-linked DNA markers have been identified for numerous traits in wheat, including disease resistance and grain quality. Employing such markers in MAS offers several advantages to wheat breeding compared to conventional phenotypic selection and laborious analysis of grain quality. These advantages include the fixation of desirable traits at an early stage of the breeding program and marker-assisted backcrossing in order to transfer agronomically important genes from wild relatives to cultivated wheat.

In addition, DNA markers are neutral to both environment and tissue type. Thus, they can be employed at any plant developmental stage and independent on environmental conditions during selection. This is particularly relevant for selection for disease resistance. DNA markers further offer the possibility for targeted pyramiding of several resistance genes, a task impossible by phenotypic selection due to complex host-pathogen interactions. To secure durable resistance, it is important to combine qualitative and quantitative resistance in a given line. Here, molecular markers can be used to combine both resistances.

As DNA markers have been correlated to numerous traits, they can be employed to combine, e.g., resistance and grain quality in the early generations. Consequently, DNA markers are being employed in early generations to select for several traits, in turn reducing the number of lines entering replicated, multi-location trials. Similarly, the number of samples for laboratory analysis of grain quality can be reduced. In effect, the application of MAS can lead to an

optimisation of resources demanded by any given breeding program, allowing the breeder to focus phenotypic selection on highly multi-genic traits, difficult to handle with MAS, e.g., yield.

Following developments in technologies and statistical genetics, the application of DNA markers in breeding is rapidly changing. While MAS has been employed to select for traits controlled by one/few genes, genomic selection will allow accurate selection for traits affected by numerous genes.

Once genomic selection has been validated in breeding programs, it can be implemented in combination with MAS. This will further improve selection efficiency and accuracy for disease resistance and quality parameters as well as for multi-genic traits such as yield.

Author details

Nana Vagndorf[1,2], Peter Skov Kristensen[1,3], Jeppe Reitan Andersen[1], Ahmed Jahoor[1,4] and Jihad Orabi[1]*

*Address all correspondence to: jior@nordicseed.com

1 Nordic Seed A/S, Odder, Denmark

2 Department of Agroecology, Flakkebjerg Research Centre, Aarhus University, Slagelse, Denmark

3 Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark

4 Department of Plant Breeding, The Swedish University of Agricultural Sciences, Alnarp, Sweden

References

[1] Dubcovsky J. Marker Assisted Selection in Wheat. http://maswheat.ucdavis.edu/

[2] Oerke EC. Crop losses to pests. The Journal of Agricultural Science. 2006;**144**(1):31

[3] Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. Shades of gray: The world of quantitative disease resistance. Trends in Plant Science. 2009;**14**:21-29

[4] Eyal Z, Scharen AL, Prescott JM, Van Ginkel M. The Septoria Diseases of Wheat: Concepts and Methods of Disease Management. Mexico: CIMMYT; 1987. 54 p

[5] Ponomarenko A, Goodwin SB, Kema GHJ. Septoria tritici blotch (STB) of wheat. Plant Health Instructor. 2011. http://www.apsnet.org/edcenter/intropp/lessons/fungi/ascomycetes/Pages/Septoria.aspx

[6] Torriani SFF, Melichar JPE, Mills C, Pain N, Sierotzki H, Courbot M. Zymoseptoria tritici: A major threat to wheat production, integrated approaches to control. Fungal Genetics and Biology. 2015;**79**:8-12

[7] Brown J, Chartrain L, Lasserre-Zuber P, Saintenac C. Genetics of resistance to Zymoseptoria tritici and applications to wheat breeding. Fungal Genetics and Biology. 2015;**79**: 33-41

[8] Adhikari TB, Yang X, Cavaletto JR, Hu X, Buechley G, Ohm HW, et al. Molecular mapping of Stb1, a potentially durable gene for resistance to septoria tritici blotch in wheat. Theoretical and Applied Genetics. 2004;**109**(5):944-953. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15490099

[9] Liu Y, Zhang L, Thompson IA, Goodwin SB, Ohm HW. Molecular mapping re-locates the Stb2 gene for resistance to Septoria tritici blotch derived from cultivar Veranopolis on wheat chromosome 1BS. Euphytica [Internet]. 2012;**190**(1):145-156. Available from: http://link.springer.com/10.1007/s10681-012-0796-8

[10] Arraiano LS, Chartrain L, Bossolini E, Slatter HN, Keller B, JKM B. A gene in European wheat cultivars for resistance to an African isolate of Mycosphaerella graminicola. Plant Pathology [Internet]. 2007;**56**(1):73-78. Available from:. DOI: 10.1111/j.1365-3059.2006.01499.x

[11] Goodwin SB, Cavaletto JR, Hale IL, Thompson I, Xu SS, Adhikari TB, et al. A new map location of gene *Stb3* for resistance to Septoria tritici blotch in wheat. Crop Science [Internet]. 2015;**55**(1):35-43. Available from: https://www.crops.org/publications/cs/abstracts/55/1/35

[12] Brading P, Verstappen E, Kema G, Brown J. A gene-for-gene relationship between wheat and mycosphaerella graminicola, the septoria tritici blotch pathogen. Phytopathology [Internet]. 2002;**92**(4):439-445. Available from http://www.ncbi.nlm.nih.gov/pubmed/18942957

[13] Flor HH. Current status of the gene-for-gene concept. Annual Review of Phytopathology. 1971;**9**:275-296

[14] Arraiano LS, Brown J. Identification of isolate-specific and partial resistance to septoria tritici blotch in 238 European wheat cultivars and breeding lines. Plant Pathology. 2006;**55**(6):726-738

[15] Vagndorf N, Nielsen NH, Edriss V, Andersen JR, Orabi J, Jørgensen LN, et al. Genomewide association study reveals novel quantitative trait loci associated with resistance towards Septoria tritici blotch in North European winter wheat. Plant Breeding. 2017;**136**:474-482

[16] Adhikari TB, Cavaletto JR, Dubcovsky J, Gieco JO, Schlatter AR, Goodwin SB. Molecular mapping of the Stb4 gene for resistance to septoria tritici blotch in wheat. Phytopathology [Internet]. 2004;**94**(11):1198-1206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18944455

[17] Arraiano LS, Worland AJ, Ellerbrook C, Brown JKM. Chromosomal location of a gene for resistance to septoria tritici blotch (*Mycosphaerella graminicola*) in the hexaploid wheat " Synthetic 6x". Theoretical and Applied Genetics. 2001;**103**:758-764

[18] McCartney CA, Brûlé-Babel AL, Lamari L, Somers DJ. Chromosomal location of a race-specific resistance gene to Mycosphaerella graminicola in the spring wheat ST6. Theoretical and Applied Genetics [Internet]. 2003;**107**(7):1181-1186. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12898022

[19] Adhikari TB, Anderson JM, Goodwin SB. Identification and Molecular Mapping of a Gene in Wheat Conferring Resistance to *Mycosphaerella graminicola*. Phytopathology [Internet]. 2003;**93**(9):1158-1164. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18944101

[20] Chartrain L, Sourdille P, Bernard M, Brown JKM. Identification and location of Stb9, a gene for resistance to septoria tritici blotch in wheat cultivars Courtot and Tonic. Plant Pathology [Internet]. 2009;**58**(3):547-555. Available from: http://doi.wiley.com/10.1111/j.1365-3059.2008.02013.x

[21] Chartrain L, Berry S, Brown J. Resistance of wheat line kavkaz-k4500 L.6.a.4 to septoria tritici blotch controlled by isolate-specific resistance genes. Phytopathology [Internet]. 2005;**95**(6):664-671. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18943783

[22] Chartrain L, Joaquim P, Berry S, Arraiano L, Azanza F, Brown J. Genetics of resistance to septoria tritici blotch in the Portuguese wheat breeding line TE 9111. Theoretical and Applied Genetics [Internet]. 2005;**110**(6):1138-1144. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15759105

[23] Ghaffary SMT, Faris JD, Friesen TL, Visser RGF, van der Lee TJ, Robert O, et al. New broad-spectrum resistance to septoria tritici blotch derived from synthetic hexaploid wheat. Theoretical and Applied Genetics [Internet]. 2012;**124**(1):125-142. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3249545&tool=pmcentrez&rendertype=abstract

[24] Ghaffary SMT, Robert O, Laurent V, Lonnet P, Margalé E, van der Lee TJ, et al. Genetic analysis of resistance to septoria tritici blotch in the French winter wheat cultivars Balance and Apache. Theoretical and Applied Genetics [Internet]. 2011;**123**(5):741-754. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3155673&tool=pmcentrez&rendertype=abstract

[25] Raman R, Milgate AW, Imtiaz M, Tan MK, Raman H, Lisle C, et al. Molecular mapping and physical location of major gene conferring seedling resistance to Septoria Tritici blotch in wheat. Molecular Breeding. 2009;**24**(2):153-164

[26] Cowger C, Hoffer ME, Mundt CC. Specific adaptation by *Mycosphaerella graminicola* to a resistant wheat cultivar. Plant Pathology. 2000;**49**(4):445-451

[27] Orton ES, Deller S, Brown JKM. *Mycosphaerella graminicola*: From genomics to disease control. Molecular Plant Pathology. 2011;**12**(5):413-424

[28] Chartrain L, Brading P, Makepeace J, Brown J. Sources of resistance to septoria tritici blotch and implications for wheat breeding. Plant Pathology. 2004;**53**:454-460

[29] Buerstmayr H, Ban T, Anderson JA. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: A review. Plant Breeding. 2009;**128**:1-26

[30] Schmale DG, Bergstrom GC. Fusarium Head Blight (FHB) or Scab. APSNET. The Plant Health Instructor; 2010. https://www.apsnet.org/edcenter/intropp/lessons/fungi/ascomycetes/Pages/Fusarium.aspx

[31] Anonymous. Commision Regulation (EC) No 856/2005 of 6 June 2005 amending regulation (EC) no 466/2001as regards Fusarium toxins. 2005

[32] Anderson JA. Marker-assisted selection for Fusarium head blight resistance in wheat. International Journal of Food Microbiology. 2007;**119**(1-2):51-53

[33] Mesterhazy A. Types and components of resistance to Fusarium head blight of wheat. Plant Breeding. 1995;**114**(5):377-386

[34] Randhawa HS, Asif M, Pozniak C, Clarke JM, Graf RJ, Fox SL, et al. Application of molecular markers to wheat breeding in Canada. Plant Breeding. 2013;**132**:458-471

[35] Liu S, Anderson JA. Targeted molecular mapping of a major wheat QTL for Fusarium head blight resistance using wheat ESTs and synteny with rice. Genome. 2003;**46**:817-823

[36] Liu S, Anderson JA. Marker assisted evaluation of Fusarium head blight resistant wheat germplasm. Crop Science. 2003;**43**(3):760-766

[37] Guo P-G, Bai G-H, Shaner GE. AFLP and STS tagging of a major QTL for Fusarium head blight resistance in wheat. Theoretical and Applied Genetics [Internet]. 2003;**106**(6):1011-1017. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12671748

[38] Rawat N, Pumphrey MO, Liu S, Zhang X, Tiwari VK, Ando K, et al. Wheat Fhb1 encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. Nature Genetics. 2016;**48**(12):1576-1580

[39] Waldron BL, Moreno-Sevilla B, Anderson JA, Stack RW, Frohberg RC. RFLP mapping of QTL for fusarium head blight resistance in wheat. Crop Science. 1999;**39**(3):805-811

[40] Cuthbert PA, Somers DJ, Brulé-Babel A. Mapping of Fhb2 on chromosome 6BS: A gene controlling Fusarium head blight field resistance in bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics. 2007;**114**(3):429-437

[41] Xue S, Li G, Jia H, Xu F, Lin F, Tang M, et al. Fine mapping Fhb4, a major QTL conditioning resistance to Fusarium infection in bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics. 2010;**121**(1):147-156

[42] Cuthbert PA, Somers DJ, Thomas J, Cloutier S, Brulé-Babel A. Fine mapping Fhb1, a major gene controlling fusarium head blight resistance in bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics. 2006;**112**(8):1465-1472

[43] Xue S, Xu F, Tang M, Zhou Y, Li G, An X, et al. Precise mapping Fhb5, a major QTL conditioning resistance to Fusarium infection in bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics. 2011;**123**(6):1055-1063

[44] Cainong JC, Bockus WW, Feng Y, Chen P, Qi L, Sehgal SK, et al. Chromosome engineering, mapping, and transferring of resistance to Fusarium head blight disease from *Elymus tsukushiensis* into wheat. Theoretical and Applied Genetics. 2015;**128**(6):1019-1027

[45] Guo J, Zhang X, Hou Y, Cai J, Shen X, Zhou T, et al. High-density mapping of the major FHB resistance gene Fhb7 derived from *Thinopyrum ponticum* and its pyramiding with Fhb1 by marker-assisted selection. Theoretical and Applied Genetics. 2015;**128**(11):2301-2316

[46] Hovmøller MS, Sørensen CK, Walter S, Justesen AF. Diversity of *Puccinia striiformis* on cereals and grasses. Annual Review of Phytopathology [Internet]. 2011;**49**:197-217. Available from: http://www.annualreviews.org/doi/abs/10.1146/annurev-phyto-072910-095230

[47] Biffen RH. Mendel's laws of inheritance and wheat breeding. Journal of Agricultural Science [Internet]. 1905;**1**(1):4. Available from: http://www.journals.cambridge.org/abstract_S0021859600000137

[48] Yang E, Li G, Li L, Zhang Z, Yang W, Peng Y, et al. Characterization of stripe rust resistance genes in the wheat cultivar Chuanmai45. International Journal of Molecular Sciences. 2016;**17**(4):601

[49] Chen W, Wellings C, Chen X, Kang Z, Liu T. Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*. Molecular Plant Pathology. 2014;**15**(5):433-446

[50] Qayoum A. High-temperature, adult-plant resistance to stripe rust of wheat. Phytopathology [Internet]. 1985;**75**(10):1121. Available from: http://www.apsnet.org/publications/phytopathology/backissues/Documents/1985Articles/Phyto75n10_1121.pdf%5Cnhttp://www.apsnet.org/publications/phytopathology/backissues/Documents/1985Articles/Phyto75n10_1121.PDF%5Cnhttp://www.apsnet.org/publications/phytop

[51] Chen X. High-temperature adult-plant resistance, key for sustainable control of stripe rust. Americas Journal of Plant Science and Biotechnology [Internet]. 2013;**4**(March):608-627. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:High-Temperature+Adult-Plant+Resistance+,+Key+for+Sustainable+Control+of+Stripe+Rust#0

[52] Singh RP. Genetic association of leaf rust resistance gene Lr34 with adult plant resistance to stripe rust in bread wheat. Phytopathology. 1992;**82**(8):835-838

[53] William M, Singh RP, Huerta-Espino J, Islas SO, Hoisington D. Molecular marker mapping of leaf rust resistance gene lr46 and its association with stripe rust resistance gene yr29 in wheat. Phytopathology. 2003;**93**(2):153-159

[54] Uauy C, Brevis JC, Chen X, Khan I, Jackson L, Chicaiza O, et al. High-temperature adult-plant (HTAP) stripe rust resistance gene Yr36 from *Triticum turgidum* ssp. dicoccoides is closely linked to the grain protein content locus Gpc-B1. Theoretical and Applied Genetics. 2005;**112**(1):97-105

[55] Sun Q, Wei Y, Ni Z, Xie C, Yang T. Microsatellite marker for yellow rust resistance gene Yr5 in wheat introgressed from spelt wheat. Plant Breeding. 2002;**121**(6):539-541

[56] Yan GP, Chen XM, Line RF, Wellings CR. Resistance gene-analog polymorphism markers co-segregating with the YR5 gene for resistance to wheat stripe rust. Theoretical and Applied Genetics [Internet]. 2003;**106**(4):636-643. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12595992

[57] Yaniv E, Raats D, Ronin Y, Korol AB, Grama A, Bariana H, et al. Evaluation of marker-assisted selection for the stripe rust resistance gene Yr15, introgressed from wild emmer wheat. Molecular Breeding. 2015;**35**(1):1-12

[58] Lagudah ES, Krattinger SG, Herrera-Foessel S, Singh RP, Huerta-Espino J, Spielmeyer W, et al. Gene-specific markers for the wheat gene Lr34/Yr18/Pm38 which confers resistance to multiple fungal pathogens. Theoretical and Applied Genetics. 2009;**119**(5):889-898

[59] Herrera-Foessel SA, Singh RP, Lan CX, Huerta-Espino J, Calvo-Salazar V, Bansal UK, et al. Yr60, a gene conferring moderate resistance to stripe rust in wheat. Plant Disease [Internet]. 2015;**99**(4):508-511. Available from: http://apsjournals.apsnet.org/doi/10.1094/PDIS-08-14-0796-RE

[60] Xiang C, Feng J, Wang M, Chen X, See D, Wan A, et al. Molecular mapping of stripe rust resistance gene Yr76 in winter club wheat cultivar tyee. Phytopathology. 2016;**106**(10):1186-1193

[61] Dong Z, Hegarty JM, Zhang J, Zhang W, Chao S, Chen X, et al. Validation and characterization of a QTL for adult plant resistance to stripe rust on wheat chromosome arm 6BS (Yr78). Theoretical and Applied Genetics. 2017;**130**(10):2127-2137

[62] Wan A, Zhao Z, Chen X, He Z, Jin S, Jia Q, et al. Wheat stripe rust epidemic and virulence of *Puccinia striiformis* f. sp. *tritici* in China in 2002. Plant Disease [Internet]. 2004;**88**(8):896-904. Available from: http://apsjournals.apsnet.org/doi/abs/10.1094/PDIS.2004.88.8.896

[63] Singh RP, Huerta-Espino J, Pfeiffer W, Figueroa-Lopez P. Occurrence and impact of a new leaf rust race on durum wheat in northwestern Mexico from 2001 to 2003. Plant Disease. 2004;**88**(7):703-708

[64] Rosewarne GM, Singh RP, Huerta-Espino J, Herrera-Foessel SA, Forrest KL, Hayden MJ, et al. Analysis of leaf and stripe rust severities reveals pathotype changes and multiple minor QTLs associated with resistance in an avocet × pastor wheat population. Theoretical and Applied Genetics. 2012;**124**(7):1283-1294

[65] Peña RJ. Wheat for bread and other foods. In: Curtis BC, Rajaram S, Gómez Macpherson H, editors. BREAD WHEAT Improvement and Production. Rome: FAO; 2002. pp. 483-542

[66] Gale KR. Diagnostic DNA markers for quality traits in wheat. Journal of Cereal Science. 2005;**41**(2):181-192

[67] Lillemo M, Morris CF. A leucine to proline mutation in puroindoline B is frequently present in hard wheats from northern Europe. Theoretical and Applied Genetics [Internet]. 2000;**100**(7):1100-1107. Available from:. DOI: 10.1038/srep10446

[68] Bhave M, Morris CF. Molecular genetics of puroindolines and related genes: Allelic diversity in wheat and other grasses. Plant Molecular Biology. 2008;**66**(3):205-219

[69] Mohler V, Schmolke M, Paladey E, Seling S, Hartl L. Association analysis of Puroindoline-D1 and Puroindoline b-2 loci with 13 quality traits in European winter wheat (*Triticum aestivum* L.). Journal of Cereal Science [Internet]. 2012;**56**(3):623-628. DOI: 10.1016/j.jcs.2012.06.010

[70] Giroux MJ, Morris CF. A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. Theoretical and Applied Genetics. 1997;**95**(5-6):857-864

[71] Huang XQ, Röder MS. Development of SNP assays for genotyping the puroindoline b gene for grain hardness in wheat using pyrosequencing. Journal of Agricultural and Food Chemistry. 2005;**53**(6):2070-2075

[72] Morris CF, Bhave M. Reconciliation of D-genome puroindoline allele designations with current DNA sequence data. Journal of Cereal Science. 2008;**48**(2):277-287

[73] Gautier MF, Aleman ME, Guirao A, Marion D, Joudrier P. *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Molecular Biology. 1994;**25**(0167-4412 SB–IM):43-57

[74] Giroux MJ, Morris CF. Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. Proceedings of the National Academy of Sciences of the United States of America [Internet]. 1998;**95**(11):6262-6266. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9600953&retmode=ref&cmd=prlinks%5Cnpapers3://publication/uuid/C38930D2-518C-4497-A4FA-DEAFDB1BD7D9%5Cnhttp://www.pnas.org/cgi/doi/10.1073/pnas.95.11.6262

[75] Morris CF, Lillemo M, Simeone MC, Giroux MJ, Babb SL, Kidwell KK. Prevalence of puroindoline grain hardness genotypes among historically significant North American spring and winter wheats. Crop Science. 2001;**41**(1):218-228

[76] Shewry PR. Wheat. Journal of Experimental Botany. 2009;**60**(6):1537-1553

[77] Payne PI. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annual Review of Plant Biology. 1987;**38**:141-153

[78] Barak S, Mudgil D, Khatkar BS. Biochemical and functional properties of wheat Gliadins: A review. Critical Reviews in Food Science and Nutrition. 2015;**55**(3):357-368

[79] Shewry PR, Halford NG, Belton PS, Tatham AS. The structure and properties of gluten: An elastic protein from wheat grain. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2002;**357**(1418):133-142

[80] Payne PI, Nightingale MA, Krattiger AF, Holt LM. The relationship between HMW glutenin subunit composition and the bread making quality of British grown wheat varieties. Journal of the Science of Food and Agriculture. 1987;**40**(1):51-65

[81] Butow BJ, Ma W, Gale KR, Cornish GB, Rampling L, Larroque O, et al. Molecular discrimination of Bx7 alleles demonstrates that a highly expressed high-molecular-weight glutenin allele has a major impact on wheat flour dough strength. Theoretical and Applied Genetics. 2003;**107**(8):1524-1532

[82] Liu Y, He Z, Appels R, Xia X. Functional markers in wheat: current status and future prospects. Theoretical and Applied Genetics [Internet]. 2012;**125**(1):1-10. Available from http://www.ncbi.nlm.nih.gov/pubmed/22366867

[83] Liu S, Rudd JC, Bai G, Haley SD, Ibrahim AMH, Xue Q, et al. Molecular markers linked to important genes in hard winter wheat. Crop Science [Internet]. 2014;**54**(4):1304. Available from: https://dl.sciencesocieties.org/publications/cs/abstracts/54/4/1304

[84] Pfluger L. MASWheat [Internet]. 2017. Available from: http://maswheat.ucdavis.edu/protocols/Gluten/index.htm

[85] McIntosh RA, Yamazaki Y, Dubcovsky J, Rogers J, Morris C, Appels R, et al. Catalogue of Gene Symbols for Wheat [Internet]. 2017. Available from: https://shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp

[86] Zhang W, Gianibelli MC, Rampling LR, Gale KR. Characterisation and marker development for low molecular weight glutenin genes from Glu-A3 alleles of bread wheat (*Triticum aestivum* L). Theoretical and Applied Genetics [Internet]. 2004;**108**(7):1409-1419. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14727031

[87] Wang L, Li G, Peña RJ, Xia X, He Z. Development of STS markers and establishment of multiplex PCR for Glu-A3 alleles in common wheat (*Triticum aestivum* L.). Journal of Cereal Science. 2010;**51**(3):305-312

[88] Wang LH, Zhao XL, He ZH, Ma W, Appels R, Peña RJ, et al. Characterization of low-molecular-weight glutenin subunit Glu-B3 genes and development of STS markers in common wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics. 2009;**118**(3):525-539

[89] Zhao XL, Xia XC, He ZH, Lei ZS, Appels R, Yang Y, et al. Novel DNA variations to characterize low molecular weight glutenin Glu-D3 genes and develop STS markers in common wheat. Theoretical and Applied Genetics. 2007;**114**(3):451-460

[90] Bonafede MD, Tranquilli G, Pflüger LA, Peña RJ, Dubcovsky J. Effect of allelic variation at the Glu-3/Gli-1 loci on breadmaking quality parameters in hexaploid wheat (*Triticum aestivum* L.). Journal of Cereal Science. 2015;**62**:143-150

[91] Gupta RB, Bekes F, Wrigley CW. Prediction of Physical dough properties from glutenin subunit composition in bread wheats: Correlation studies. Cereal Chemistry. 1991;**68**:328-33

[92] He ZH, Liu L, Xia XC, Liu JJ, Pena RJ. Composition of HMW and LMW glutenin subunits and their effects on dough properties, pan bread, and noodle quality of Chinese bread Wheats. Cereal Chemistry. 2005;**82**(4):345-350

[93] Ma W, Appels R, Bekes F, Larroque O, Morell MK, Gale KR. Genetic characterisation of dough rheological properties in a wheat doubled haploid population: Additive genetic effects and epistatic interactions. Theoretical and Applied Genetics. 2005;**111**(3):410-422

[94] Würschum T, Leiser WL, Kazman E, Longin CFH. Genetic control of protein content and sedimentation volume in European winter wheat cultivars. Theoretical and Applied Genetics. 2016;**129**(9):1685-96

[95] Krystkowiak K, Langner M, Adamski T, Salmanowicz BP, Kaczmarek Z, Krajewski P, et al. Interactions between Glu-1 and Glu-3 loci and associations of selected molecular markers with quality traits in winter wheat (*Triticum aestivum* L.) DH lines. Journal of Applied Genetics. 2017;**58**(1):37-48

[96] Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR. Identification of SNPs and development of allele-specific PCR markers for γ-gliadin alleles in *Triticum aestivum*. Theoretical and Applied Genetics [Internet]. 2003;**107**(1):130-138. Available from: http://link.springer.com/10.1007/s00122-003-1223-2

[97] Liu S, Chao S, Anderson JA. New DNA markers for high molecular weight glutenin subunits in wheat. Theoretical and Applied Genetics. 2008;**118**(1):177-183

[98] Ragupathy R, Naeem HA, Reimer E, Lukow OM, Sapirstein HD, Cloutier S. Evolutionary origin of the segmental duplication encompassing the wheat GLU-B1 locus encoding the overexpressed Bx7 (Bx7OE) high molecular weight glutenin subunit. Theoretical and Applied Genetics. 2008;**116**(2):283-296

[99] Ishikawa G, Nakamura T. A new co-dominant PCR-based marker to identify the high-molecular-weight glutenin subunit combination. Wheat Information Service. 2007;**103**:1-3

[100] Mago R, Spielmeyer W, Lawrence GJ, Lagudah ES, Ellis JG, Pryor A. Identification and mapping of molecular markers linked to rust resistance genes located on chromosome 1RS of rye using wheat-rye translocation lines. Theoretical and Applied Genetics. 2002;**104**(8):1317-1324

[101] Graybosch RA. Uneasy unions: Quality effects of rye chromatin transfers to wheat. Journal of Cereal Science. 2001;**33**:3-16

[102] Mohler V, Albrecht T, Mrva K, Schweizer G, Hartl L. Genetic analysis of falling number in three bi-parental common winter wheat populations. Plant Breeding. 2014;**133**(4): 448-453

[103] Furtado A, Bundock PC, Banks PM, Fox G, Yin X, Henry RJ. A novel highly differentially expressed gene in wheat endosperm associated with bread quality. Scientific Reports. 2015;**5**(April):1-14

[104] Guzmán C, Xiao Y, Crossa J, González-Santoyo H, Huerta J, Singh R, et al. Sources of the highly expressed wheat bread making (wbm) gene in CIMMYT spring wheat germplasm and its effect on processing and bread-making quality. Euphytica. 2016;**209**: 689-692

[105] Distelfeld A, Uauy C, Fahima T, Dubcovsky J. Physical map of the wheat high-grain protein content gene Gpc-B1 and development of a high-throughput molecular marker. The New Phytologist. 2006;**169**(4):753-763

[106] Ellis MH, Spielmeyer W, Gale KR, Rebetzke GJ, Richards RA. "Perfect" markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. Theoretical and Applied Genetics. 2002;**105**(6-7):1038-1042

Chapter 2

Soybean Breeding on Seed Composition Trait

Qi Zhaoming, Jingyao Yu, Hongtao Qin,
Zhang Zhanguo, Shiyu Huang, Xinyu Wang,
Mao Xinrui, Qi Huidong, Zhengong Yin, Candong Li,
Xiaoxia Wu, Xin Dawei, Jiang Hongwei,
Liu Chunyan, Hu Zhenbang and Chen Qingshan

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74353

Abstract

Soybean is a most important crop providing edible oil and plant protein source for human beings, in addition to animal feed because of high protein and oil content. This review summarized the progresses in the QTL mapping, candidate gene cloning and functional analysis and also the regulation of soybean oil and seed storage protein accumulation. Furthermore, as soybean genome has been sequenced and released, prospects of multiple omics and advanced biotechnology should be combined and applied for further refine research and high-quality breeding.

Keywords: soybean, seed oil content, seed storage protein

1. Introduction

Soybean (*Glycine max* [L.] Merr.) accounts around 60% of the world's oilseed consumption and also 68% of world protein meal consumption (http://www.soystats.com), which plays an important role year by year. In addition, during oil purification, protein-rich soybean meal is produced, which also provided around 75% of protein meal for animal feed worldwide [1]. Thus, improvement of soybean quality is important for worldwide commercial production, and it is also a key target for soybean breeding.

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (cc) BY

1.1. Soybean protein and oil content QTL analysis

Soybean oil and protein content were quantitative trait and effected by multiple genes and environments factors [2, 3]; there were over 312 soybean oil QTLs and 231 soybean protein QTLs having been detected by different population and environments (SoyBase, http://www.soybase.org), with the main mapping methods including the analysis of variance (ANOVA; [4]), interval mapping (IM; [5–7]), composite interval mapping (CIM; [8, 9]), multiple interval mapping (MIM; [10]) and inclusive composite interval mapping (ICIM; [11]). Among the published soybean oil content QTLs, some of them showed 'hot regions' that have been identified four or more times at the same or similar intervals in different studies, which include Gm05: 35.2–40.8 Mb, Gm09: 40.3–46.8 Mb, Gm12: 34.1–40.6 Mb, Gm14: 33.8–49.2 Mb, Gm15: 0.8–13.9 Mb, Gm18: 51.6–59.8 Mb, Gm19: 32.9–48.0 Mb and Gm20: 23.5–34.6 Mb [12]. For soybean protein content, there were also some 'hot regions' included Gm04: 43.6–47.7 Mb, Gm05: 39.7–41.4 Mb, Gm07: 4.2–9.6 Mb, Gm08: 5.8–10.2 Mb, Gm14: 4.8–9.6 Mb, Gm15: 0.0–7.5 Mb, Gm18: 47.9–54.0 Mb, Gm19: 35.5–42.1 Mb and Gm20: 2.1–34.2 Mb [13, 14]. Meta-analysis is a statistical method that could combine results from different sources in a single study [15]; it can increase QTL precision and validity by using mathematical models to refine the integration of QTLs [16] and have been performed in maize [17] and soybean [18] at the beginning of application. Meta-analysis method has also been employed to analyze the soybean oil and protein content separately by Qi et al. [19, 20].

However, soybean oil and protein content always showed the opposite relationship [21, 22], with the observation and data collections from many classical genetic analysis, the high oil variety with lower protein content and high protein variety with lower oil content [23]. And also, many classical genetic and breeding books or data noted the opposite relationship for soybean oil and protein content [2, 24–34]. Although it was very hard to find the locus which could increase soybean oil and protein content at the same time [35], based on the big amounts of QTL mapping results, few regions showed the same direction of contribution to soybean oil and protein content in the same genetic population. Orf et al. [36] mapped the additive QTL affected the soybean oil content at 39.5–41.2 Mb of Gm05 with the population crossed by Minsoy and Noir1, the results implied Minsoy bring the positive alleles for increasing soybean oil and protein content, however, Specht et al. [37] identified the similar region with the opposite results that Noir1 bring the positive alleles. Hyten et al. [38] identified a QTL at 4.8–8.7 Mb of Gm07 and the parent Williams bring the positive alleles for both traits. Reinprecht et al. [39] also demonstrated that the variety OX948 bring the positive alleles. Mao et al. [40] identified the additive QTLs affected the soybean oil and content at 51.2–56.3 Mb of Gm01, 1.0–2.3 Mb of Gm09 and 39.4–46.1 Mb of Gm19 in the cross population of Hefeng47 and Heinong37, which indicated that the soybean variety Heinong37 bring the positive alleles of those regions that could increase the soybean oil and protein content at the same time. Heinong37 was the only one Chinese variety, which may bring the positive alleles for both traits based on published data.

1.2. Soybean fatty acid composition biosynthesis and transcriptional regulation

The accumulation of starch, lipid and protein supplied the raw materials and energy for soybean seed growth and maturity. Lipid was one of the three significant raw materials, although

the biochemical pathway about synthesis of lipid has been studied thoroughly, the regulation mechanism is unclear till now [41–47]. De novo synthesis of fatty acid mainly started in plant plastid. Acetyl -CoA is a precursor of soybean seed fatty acid synthesis. It is an important intermediate of many cellular metabolisms, and it synthesizes a lot in plant cell and then acetyl-CoA carboxylase (ACCase) catalyzes the first committed step of fatty acid synthesis, acetyl-CoA carboxylate to malonyl-CoA [48]. After that, malonyl-CoA has been catalyzed by fatty acid synthase complex (FAS) and proceeding of continuous polymerization reaction based on the acyl carbon chains synthesized with a frequency of two carbons per cycle. The growing acyl carbon chain binds to acyl-carried proteins (ACP) and termination with the acyl-ACP thioesterase or acyltransferase form into acyl ACP. Furthermore, different lengths of acyl ACP synthesized the acyl-CoA with acyl-CoA synthetase and transferred from the plasmids to the endoplasmic reticulum or the cytoplasm. At last, fatty acids were attached to glycerol to synthesize triacylglycerides (TAGs) with three different acyltransferases respectively [49–52]. Till now, seed oil content can be increased by changing the expression levels of individual enzymes involved in oil metabolism [53–59]. However, the key enzyme responsible for TAG assembly is encoded by diacylglycerol acyltransferase 1 (*DGAT1*) [59–61], and expression of *DGAT1* can be used to draw fatty acids into TAG; overexpression of *DGAT1* could increase both seed oil content (by 9–12%) and seed weight (40–100%) in *Arabidopsis* [55]. Overexpression of *TmDGAT1a* and *TmDGAT1b* could increase soybean seed oil content [62]. *SiDGAT1* encoding acyl-CoA could also increase soybean seed oil content [63]. When expressing *VgDGAT1A*, (from *Vernonia galamensis*) it could make soybean oil content increase obviously [64]. Furthermore, the speed limit of fatty acid biosynthesis enzyme in dicotyledonous plants is biotin carboxylase (BC), which is a vital subunit of acetyl-CoA. Li et al. [65] cloned four genes encoding *BC* from *Brassica napus* and elucidated the evolution and the regulation of ACCase in the *Brassica*. The cytosolic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPC) catalyzes a key reaction in glycolysis, whose levels are directly correlated with seed oil accumulation [66].

Fatty acid composition were determined mainly by five fatty acids, palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2) and linolenic (C18:3) [67, 68]. Most palmitic acid (16:0) produced by the type II synthase is elongated to stearic acid (18:0) [67, 69]. In recent decades, there were many reports about the QTLs of each components of fatty acid, and there were also some 'hot regions' for soybean seed linoleic included Gm05 39.36–40.87 Mb and Gm18 48.35–50.78 Mb (with the original QTLs from Diers and Shoemaker [70]; Bachlava et al. [71]; Li et al. [65]; Xie et al. [72]); for soybean seed linolenic included Gm02 17.07–34.9 Mb, Gm09 34.56–37.74 Mb, Gm14 17.08–39.5 Mb and 45.68–46.78 Mb, Gm15 6.7–7.71 Mb, 13.07–25.6 Mb and Gm19 35.75–37.38 Mb (with the original QTLs from Li et al. [65], Bachlava et al. [71]; Diers and Shoemaker [70]; Spencer et al. [73]; Reinprecht et al. [39]; Xie et al. [72]; Shibata et al. [74]; Hyten et al. [38]); for soybean seed oleic included Gm05 39.07–40.80 Mb and Gm18 49.24–51.95 Mb (with the original QTLs from Diers and Shoemaker [70]; Reinprecht et al. [39]; Xie et al. [72]); for soybean seed palmitic included Gm05 2.84–3.92 Mb, Gm09 7.74–11.83 Mb and 34.59–38.73 Mb, Gm15 9.13–13.16 Mb, Gm17 7.60–9.45 Mb and Gm18 38.38–41.09 Mb (with the original QTLs from Li et al. [75]; Wang et al. [76]; Xie et al. [72]; Hyten et al. [38]; Li et al. [65]; Kim et al. [77], Reinprecht et al. [39]). In soybean, stearoyl-acyl carrier protein desaturase

(SAD) catalyzes the first step in seed oil biosynthesis, converting stearoyl-ACP to oleoyl-ACP, which plays a key role in determining the ratio of total saturated to unsaturated fatty acid in plants [35, 78, 79].Then, microsomal oleate desaturase (FAD2) and linoleoyl desaturase (FAD3) catalyze oleic to linoleic acid mainly in the sn-2 position, and then, fatty acid elongase converts fatty acids into a long-chain fatty acid [80]. The *FAD2* gene family of soybean was consisted of at least five members in four genome regions and was responsible for the conversion of oleic acid to linoleic acid [81–84]. The *FAD3* enzyme contributes to the synthesis of α-linolenic acids (18:3) in the polyunsaturated fatty acid pathway. To improve soybean oil quality, we aim at reducing the percentage of α-linolenic acids. *GmFAD3* mutant can reduce α-linolenic acid content in soybean seed oil, which has been verified in many studies [58, 85–87].

However, overexpression of a single gene of fatty acid synthesis does not significantly improve the fatty acid biosynthesis [88, 89]. Fatty acid synthesis is regulated by some major classical transcription factors coupling with seed development, including *WRINKLED1* (*WRI1*) *LEAFY COTYLEDON1* (*LEC1*), *LEC2*, *ABSCISIC ACID INSENSITIVE3* (*ABI3*), and *FUSCA3* (*FUS3*) [90–95] were the plant-specific B3 transcription factor family, *LEC1* was an NFY-B-type or CCAAT-binding factor-type transcription factor [96] and *WRI1* encodes a transcription factor of APETALA2-ethylene responsive element-binding protein (AP2-EREBP) family [90]. *WRI1* is a potential global regulator of *de novo* fatty acid biosynthesis that specifies the regulatory action of the direct target of *LEC2* [97]. Overexpression of the transcription factor *WRI1*, which controls the expression of genes involved in lipid metabolism, including glycolysis and fatty acid biosynthesis, increased seed oil content by 10–20% compared to the wild type [40, 90, 98–101]. *LEC1* function was partially dependent on *ABI3*, *FUS3* and *WRI1* in the regulation of fatty acid biosynthesis; both *LEC1* and *LEC1*-like genes were acted as key regulators to coordinate the expression of fatty acid biosynthetic genes [92]. *LEC2* can regulate *WRI1* directly and is necessary for the regulatory action of fatty acid metabolism [97]. Ectopic expression of *FUS3* can trigger the expression of fatty acid biosynthetic genes [41], and interaction of *FUS3* and *AKIN10* positively regulates auxin biosynthesis and indirectly regulates fatty acid biosynthesis [102]. Furthermore, few new soybean transcription factors have been identified for fatty acid biosynthesis in recent years, mainly including *GmbZIP123* regulates lipid accumulation indirectly through the sugar translocation [103]; *GmMYB73* was functioned as a repressor for negative regulator *GLABRA2 (GL2)* [104] and relieved *GL2*-inhibited expression of *PLDα1* to accelerate conversion of phosphatidylcholine to TAG [43]; *GmZF351* will improve oil accumulation by directly activating *WRI1*, *BCCP2*, *KASIII*, *TAG1* and *OLEO2* [104]; *GmNFYA* has been identified to increase seed oil content based on RNA-seq and gene coexpression networks [46] and *GmDOF4* and *GmDOF11* can increase lipid content in seeds by direct activation of lipid biosynthesis genes [41, 105]. In recent, regulatory mechanisms of seed oil content have been updated by duplicated genes in soybean [106].

In addition, other transcription factors have been identified to affect oil content in Arabidopsis, including *GL2*, *TT1*, *TT2*, *bZIP67*, *MED*, *MYB* [58, 107, 108] and *BASS2* [43, 107–112].

1.3. Soybean seed storage protein (SSP) and transcriptional regulation

Soybean seed storage proteins (SSP) have been identified and classified into four basic categories, including albumins (water-soluble), globulins (salt-soluble), prolamins (alcohol-soluble)

and glutelins (weak acid/weak base-soluble) [113, 114]. Globulin is the main component of SSP and can be classified into four groups according to different sedimentation coefficients, which are 2S (including trypsin inhibitors and cytochrome and other ingredients), 7S (β-conglycinin), 11S (glycinin) and 15S (polymer of glycinin) [115]. 7S and 11S are the main components of soybean seed storage protein, and they are accounting for 60–80% of the whole soybean seed storage protein [116–120]. Till now, about the genetic mechanisms of 7S and 11S, globulin subunits are clear in general [121–124]. β-conglycinin is accounting for roughly 30–40% of the total seed protein and is mainly composed of α-(76kD), α ′-(72kD) and β-(53kD) subunits [125–127]. Glycinin is accounting for roughly 40–60% of the total seed protein and is mainly composed of G1, G2, G3, G4 and G5 subunits (approximately 56, 54, 54, 64 and 58 kD, respectively) [113, 118, 128]. In the past several years, few QTL mapping researches were conducted for soybean seed 7S and 11S; the QTL region of 11S includes Gm09 45.6–47.6 Mb and 103.7–105.8 Mb, Gm17 79–81 Mb, Gm19 55.1–57.1 Mb, Gm19 60.3–62.35 Mb and Gm20 81.7–83.7 Mb [129]; the QTL region of 7S includes one QTL of α′-7S located on Gm08 35.7–37.7 Mb and nine QTLs of β-7S located on Gm01 65–104 Mb, Gm03 75.4–77.49 Mb, Gm17 26–81 Mb, Gm19 30–31 Mb, 100.7–115 Mb and Gm20 92–98 Mb [129, 130]. The genes of 11S and 7S have been reported, the genes of 11S subunit include *Gy1*, *Gy2*, *Gy3*, *Gy4*, *Gy5* and *Gy7* and the genes of the 7S subunit mainly include CG-alpha-1 (*7sα*), CG-alpha′-1 (*7sα′*) and CG-beta-1 (*7sβ*) [131–134]. Three genes encoding 11S, *AtCRU1*, *AtCRU2* and *AtCRU3*, have been verified in *Arabidopsis thaliana* [135]. Wang et al. [136] mapped a QTL *qBSC-1* (7S), which could regulate the SSP. Knockdown of 7S globulin subunits can change nitrogen content in transgenic soybean seeds [137]. Furthermore, the ratio of 11S to 7S is ranged from 0.5 to 1.7 among cultivar soybean and affects nutritional quality and functional properties of soybean seed storage protein directly [138, 139]. And also, it is amusing that the content of 7S and 11S are significantly negative correlation [140]. Yang et al. [141] demonstrated that the lack of 11S4A induced the compensatory accumulation of 7S globulins. By adjusting the subunit composition of soybean seed storage protein, it can remove sensitization protein efficiently; at the same time, it is an approach to improve the quality of the soy protein nutrition and production and processing [42, 103, 142, 143].

Accumulation of soybean seed storage protein is always coupling with TAGs and some key transcription factors involved in the process [144]. B3-type transcription factors can act directly on the expression of SSP genes [145]. The B3 domain, identified as the DNA-binding motif, recognizes the RY motif (CATGCA) as the target sequence [146], and RY motif (CATGCA) is a cis-acting element as a seed-specific promoter, which is the most legume seed storage protein gene that contain one or more RY repeating elements [65, 128]. Several studies have shown that the binding of the *ABI3* with the RY motif can regulate the accumulation of storage proteins in Arabidopsis seeds [147–150]. The seed-specific B3 domain transcription factors, *LEC2*, *FUS3* and *ABI3*, have been identified, and the mutations of these genes often showed the negative accumulation of seed storage proteins [151–154]. In addition of *ABI3*, *ABI4* and *LEC1* also showed the interaction to regulate the SSP [96, 155]. Some previous studies showed that these genes affect the induction of storage protein gene expression directly [156–159]. Furthermore, expression *OLEOSIN* required activation of *LEC2* and two RY elements on its promoter [146]. Both *LEC1* and *LEC2* act as positive regulators upstream of *ABI3* and *FUS3*, function analysis showed influence on the expression of seed storage protein (SSP) genes [44,

153, 158, 160, 161]. *LEC1* and *L1L* can active the promoter of *CRUCIFERIN C* (*CRC*), and *LEC1* can also regulate *CRC* and other SSP genes working with *FUS3* and *ABI3* [161]. In addition to RY motifs, the presence of G-Box elements is also proper activation of target promoters of *LEC1*, *LEC2*, *ABI3* and *FUS3* [162]. Some studies showed that *LEC2*, *ABI3* and *FUS3* collaborate with *bZIPs* TFs that interact with these G-Box elements to activate SSP genes [163, 164]. Furthermore, *GmDOF4* and *GmDOF11* can bind with the promoter of *CRA1* to regulate the expression of SSP [41]. *GmDREBL* can be upregulated by *GmABI3* and *GmABI5* and be regulated by the late stage of SSP genes [44]. *DGAT* can reduce the soluble carbohydrate content of mature seeds and increase the seed protein content at the same time [165]. Therefore, in addition to *WAR1*, *LEC1*, *LEC2*, *ABI3* and *FUS3*, transcription factors of *MYB*, *bZIP*, *MADS*, *DOF* or *AP2* families are also involved in the accumulation of storage compounds (oil and SSPs) and seed development regulatory network, as partners or direct target genes [162].

1.4. Small RNA regulation of seed composition

Small RNAs, such as miRNAs and short interfering RNAs (siRNAs), are key components of the evolutionarily conserved system of gene regulation in eukaryotes [166]. Wherein, microRNAs (miRNAs) are a class of non-coding small RNAs of 20–24 nt in length that play an important role in plant growth and development. Structurally, except for the characteristics of the segments, all miRNA precursors have well-predicted stem-loop hairpin structures, and this fold-back hairpin structure has a low degree of freedom of energy [167]. The microRNA database (http://www.mirbase.org/) is a searchable database of published miRNA sequences and annotations. According to miRBase, miRNA information of 1269 species has been collected, including 399 soybean miRNAs. For example, gma-MIR156d belongs to the MIPF0000008, MIR156 gene family, described as *Glycine max* miR156d stem-loop, annotated that microRNA (miRNA) precursor mir-156 is a family of plant non-coding RNA. This microRNA has now been predicted or experimentally confirmed in a range of plant species (MIPF0000008). The products are thought to have regulatory roles through complementarity to mRNA. SFGD is a comprehensive database of integrated genomic and transcriptome data and a comprehensive database of soy acyl lipid metabolic pathways, including a coexpression regulatory network of 23,267 genes and 1873 miRNA-target pairs as well as a set of acyl-lipid pathways containing 221 enzymes and more than 1550 genes, providing biologists with a useful toolbox [168]. In addition, SoyKB is also a website, which provides information on soybean genomics, transcriptomics, proteomics and metabolomics as well as gene function and biology annotation, including information like genes, microRNAs, metabolites and mono nucleotide polymorphisms (SNPs) [169]. Shi and Chiang used miRNA-specific forward primers and sequences complementary to poly(T) linkers as reverse primers to find a simple and effective method to determine miRNA expression. Total RNA (including miRNAs) was polyadenylated and reverse transcribed into cDNA using poly (T) linkers for real-time PCR.

There are few studies on miRNAs related to plant quality. Soybean cotyledons affect soybean seed yield and quality. Goettel et al. analyzed 304 miRNA genes expressed in soybean cotyledons and predicted their complex miRNA networks to 1910 genes. By analyzing extensive biological pathways present in soybean cotyledons, the evolutionary pathways of

soybean miR15/49 in soybean cotyledons were further demonstrated [170]. Ye et al. identified and analyzed the whole genome of miRNA endogenous target gene mimic (eTM) and the phagemid-generated siRNA (PHAS) in soybean, with a focus on lipid metabolism-related genes. Lipid metabolism was found to be regulated by a potentially complex non-coding network in soybean, of which 28 may be miRNA-regulated and nine may be further regulated [171].

2. Conclusion and perspectives

As sequencing development of soybean genome, the cultivar Williams 82 genome has been released by Schmutz et al. [172], and it update the quality of assembly of the reference genome year by year. In present version (*Glycine max Wm82.a2.v1*), 56,044 protein-coding loci and 88,647 transcripts have been predicted, and all related data have been released in Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax). At the basis of the reference genome, around 265 cultivated soybean varieties, 92 wild soybean varieties and 10 semi-wild soybean varieties have been resequenced; these information give a foundation for functional genomic analyses such as transcriptomic, proteomic, epigenomic and non-coding RNA analyses [173].

Although many genes and regulators of seed oil content and SSP have been identified and their associated regulatory networks have been well studied in Arabidopsis, there are still unclear in soybean in addition to *WAR1*, *LEC1*, *LEC2*, *ABI3* and *FUS3* due to the 75% duplication genome [172]. Combination and application of multiple omics (genomics, functional genomics, transcriptomic, proteomics and epigenomics) and advanced biotechnology (genome editing) needed to clarify the soybean seed oil content and SSP gene and regulatory network. Secondary population including recombinant heterozygous lines (RHL), chromosome segment substitution line (CSSL) and/or near isogenic lines (NIL) need to be applied to reduce the variable for analyzing the effects of single gene or transcription factors and used to identify the effective alleles and evaluate its effects and contribution. Combination of general loci could be further used for design of selection chip assay, which may lead to the foundation of high oil or high seed storage protein breeding.

Acknowledgements

This study was supported by the National Key R&D Program of China (2016 YFD0100500, 2016YFD0100300, 2016YFD0100201-21), the National Natural Science Foundation of China (31701449, 31471516, 31401465, 31400074, 31501332), the Natural Science Foundation of Heilongjiang (QC2017013), the Young Innovative Talent training plan of undergraduate colleges and universities in Heilongjiang province (UNPYSCT-2016144), special financial aid to post-doctor research fellow in Heilongjiang (To Qi Zhaoming), the Heilongjiang Funds for Distinguished Young Scientists (JC2016004) and the Outstanding Academic Leaders

Projects of Harbin, China (2015RQXXJ018), the China Post Doctoral Project (2015 M581419), the Dongnongxuezhe Project (to Chen Qingshan), the Young Talent Project (to Qi Zhaoming, 518062) of Northeast Agricultural University and SIPT project of Northeast Agriculture University (2018-171, 2018-172).

Author details

Qi Zhaoming[1]*, Jingyao Yu[1], Hongtao Qin[1], Zhang Zhanguo[1], Shiyu Huang[1], Xinyu Wang[1], Mao Xinrui[1], Qi Huidong[1], Zhengong Yin[1], Candong Li[1], Xiaoxia Wu[1], Xin Dawei[1], Jiang Hongwei[1,2], Liu Chunyan[1], Hu Zhenbang[1] and Chen Qingshan[1]

*Address all correspondence to: qizhaoming1860@126.com

1 College of Agriculture, Northeast Agricultural University, Harbin, Heilongjiang, People's Republic of China

2 Crop Research and Breeding Center of Land-Reclamation of Heilongjiang Province, Harbin, Heilongjiang, People's Republic of China

References

[1] Boland MJ, Rae AN, Vereijken JM, Meuwissen MPM, Fischer ARH, van Boekel MAJS, Rutherfurd SM, Gruppen H, Moughan PJ, Hendriks WH. The future supply of animal-derived protein for human consumption. Trends in Food Science & Technology. 2013;**29**:62-73. DOI: 10.1016/j.tifs.2012.07.002

[2] Wilcox JR. Breeding soybeans for improved oil quantity and quality. In: Shibles R, editor. Proc. 3rd World Soybean Res. 1985

[3] Burton JW. Breeding soybeans for improved protein quantity and quality. In: Shibles R, editor. Proceedings of the 3rd World Soybean Res. Conference. Westview Press, Boulder/CO. 1985. pp. 361-367

[4] Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics. 1987;**1**:174-181. DOI: 10.1016/0888-7543(87)90010-3

[5] Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989;**121**:185-199

[6] Haley CS, Knott SA. A Simple dominant inheritance of a trait lacking β-conglycinin detected in a wild soybean line le regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity. 1992;**69**:315-324

[7] Jansen RC. Interval mapping of multiple quantitative trait loci. Genetics. 1993;**135**:205-211

[8] Rodolphe F, Lefort M. A multi-marker model for detecting chromosomal segments displaying QTL activity. Genetics. 1993;**134**:1277-1288

[9] Zeng ZB. Precision mapping of quantitative trait loci. Genetics. 1994;**136**:1457-1468

[10] Kao CH, Zeng ZB, Teasdale RD. Multiple interval mapping for quantitative trait loci. Genetics. 1999;**152**:1203-1216

[11] Meng L, Li H, Zhang L, Wang J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal. 2015;**3**:269-283. DOI: 10.1016/j.cj.2015.01.001

[12] Chen QS, Mao XR, Zhang ZG, Zhu RS, Yin ZG, Leng Y, Yu HX, Jia HY, Jiang SS, Ni ZQ, Jiang HW, Han X, Liu CY, Hu ZB, Wu XX, Hu GH, Xin DW, Qi ZM. SNP-SNP interaction analysis on soybean oil content under multi-environments. PLoS One. 2016;**11**:e0163692. DOI: 10.1371/journal.pone.0163692

[13] Zhang YH, Liu MF, He JB, Wang YF, Xing GN, Li Y, Yang SP, Zhao TJ, Gai JY. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.]. Theoretical and Applied Genetics. 2015;**128**:1061-1072. DOI: 10.1007/s00122-015-2490-4

[14] Qi ZM, Pan JB, Han X, Qi HD, Xin DW, Li W, Mao XR, Wang ZY, Jiang HW, Liu CY, Hu ZB, Hu GH, Zhu RS, Chen QS. Identification of major QTLs and epistatic interactions for seed protein concentration in soybean under multiple environments based on a high-density map. Molecular Breeding. 2016;**36**:5. DOI: 10.1007/s11032-016-0475-x

[15] Primary GGV. Secondary, and meta-analysis of research. Educational Researcher. 1976;**5**:3-8

[16] Goffinet B, Gerber S. Quantitative trait loci: A meta-analysis. Genetics. 2000;**155**:463-473

[17] Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A. Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics. 2004;**168**:2169-2185. DOI: 10.1534/genetics.104.032375

[18] Guo B, Sleper DA, Lu P, Shannon JG, Nguyen HT, Arelli PR. QTLs associated with resistance to soybean cyst nematode in soybean: Meta-analysis of QTL location. Crop Science. 2006;**46**:595-602. DOI: 10.2135/cropsci2005.04-0036-2

[19] Qi ZM, Wu Q, Han X, Sun YN, Du XY, Liu CY, Jiang LH, Hu GH, Chen QS. Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. Euphytica. 2011;**179**:499-514. DOI: 10.1007/s10681-011-0386-1

[20] Qi ZM, Sun YN, Wu Q, Liu CY, Hu GH, Chen QS. A meta-analysis of seed protein concentration QTL in soybean. Canadian Journal of Plant Science. 2011;**91**:221-230. DOI: 10.1139/CJPS09193

[21] Johnson HW, Bernard RL. Soybean genetics and breeding. Advances in Agronomy. 1962;**14**:149-221. DOI: 10.1016/S0065-2113(08)60438-1

[22] Kwon SH, Torrie JH. Heritability of and interrelationships among traits of two soybean populations. Crop Science. 1989;**4**(2):482-488

[23] Burton JW. Quantitative genetics: Results relevant to soybean breeding. Agronomy (USA). 1987:45-64

[24] Johnson HW, Robinson HF, Comstock RE. Genotypic and phenotypic correlations in soybeans and their implications in selection. Agronomy Journal. 1955;**47**:477-483. DOI: 10.2134

[25] Hanson WD, Leffel RC, Howell RW. Genetic analysis of energy production in the soybean. Crop Science. 1961;**1**:121-126

[26] Smith RR, Weber CR. Mass selection by specific gravity for protein and oil in soybean populations. Crop Science. 1968;**8**:373-377. DOI: 10.2135/cropsci1968.0011183X000800030033x

[27] Hartwig EE, Hinson K. Association between chemical composition of seed and seed yield of soybeans. Crop Science. 1972;**12**:829-830. DOI: 10.2135/cropsci1972.0011183X001200060034x

[28] Shannon JG, Wilcox JR, Probst AH. Estimated gains from selection for protein and yield in the F4 generation of six soybean populations. Crop Science. 1972;**12**:824-826. DOI: 10.2135/cropsci1972.0011183X001200060032x

[29] Brim CA, Caldwell BE, Howell RW. Soybeans: Improvement, production and uses. American Society of Agronom. 1973:681 p

[30] Brim CA, Burton JW. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Science. 1979;**19**:494-498. DOI: 10.2135/cropsci1979.0011183X001900040016x

[31] Sebern NA, Lambert JW. Effect of stratification for percent protein in two soybean populations. Crop Science. 1984;**24**:225-228. DOI: 10.2135/cropsci1984.0011183X002400020003x

[32] Wehrmann VK, Fehr WR, Cianzio SR, Cavins JF. Transfer of high seed protein to high-yielding soybean cultivars. Crop Science. 1987;**27**:927-931. DOI: 10.2135/cropsci1987.0011183X002700050020x

[33] Wilcox JR, Cavins JF. Backcrossing high seed protein to a soybean cultivar. Crop Science. 1995;**35**:1036-1041. DOI: 10.2135/cropsci1995.0011183X003500040019x

[34] Cober ER, Voldeng DH. Developing high-protein, high-yield soybean populations and lines. Crop Science. 2000;**40**:39-42

[35] Thompson GA, Scherer DE, Foxall-Van Aken S, Kenney JW, Young HL, Shintani DK, Kridl JC, Knauf VC. Primary structures of the precursor and mature forms of stearoyl-acyl

carrier protein desaturase from safflower embryos and requirement of ferredoxin for enzyme activity. Proceedings of the National Academy of Sciences. 1991;**88**:2578-2582. DOI: 10.1073/pnas.88.6.2578

[36] Orf, JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, Lark KG. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Science. 1999;**39**:1642-1651. DOI: 10.2135/cropsci1999.3961642x

[37] Specht JE, Chase K, Macrander M, Graef GL, Chung J, Markwell JP, Germann M, Orf JH, Lark KG. Soybean response to water. Crop Science. 2001;**41**:493-509. DOI: 10.1104/pp.124.4.1752

[38] Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME. Seed quality QTL in a prominent soybean population. Theoretical and Applied Genetics. 2004;**109**:552-561. DOI: 10.1007/s00122-004-1661-5

[39] Reinprecht Y, Poysa VW, Yu K, Rajcan I, Ablett GR, Pauls KP. Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. Genome. 2006;**49**:1510-1527. DOI: 10.1139/g06-112

[40] Mao T, Jiang Z, Han Y, Teng W, Zhao X, Li W. Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. Plant Breeding. 2013;**132**:630-641. DOI: 10.1111/pbr.12091

[41] Wang H, Guo J, Lambert K, Lin Y. Developmental control of Arabidopsis seed oil biosynthesis. Planta. 2007;**226**:773-783. DOI: 10.1007/s00425-007-0524-0

[42] Song B, Shen LW, Liu SS. Marker-assisted backcrossing of a null allele of the α-subunit of soybean (*Glycine max*) b-conglycinin with a Chinese soybean cultivar. Plant Breeding. 2014;**133**:638-648. DOI: 10.1111/pbr.12203

[43] Liu YF, Li QT, Lu X, Song QX, Lam SM, Zhang WK, Ma B, Lin Q, Man WQ, Du WG, Shui GH, Chen SY, Zhang JS. Soybean GmMYB73 promotes lipid accumulation in transgenic plants. BMC Plant Biology. 2014;**14**:73. DOI: 10.1186/1471-2229-14-73

[44] Zhang YQ, Lu X, Zhao FY, Li QT, Niu SL, Wei W, Zhang WK, Ma B, Chen SY, Zhang JS. Soybean GmDREBL increases lipid content in seeds of transgenic Arabidopsis. Scientific Reports. 2016;**6**:34307. DOI: 10.1038/srep34307

[45] Zhang YQ, Lu X, Li QT, Chen SY, Zhang JS. Recent advances in identification and functional analysis of genes responsible for soybean nutritional quality. Scientia Agricultura Sinica. 2016;**49**:4299-4309

[46] Lu X, Li QT, Xiong Q, Li W, Bi YD, Lai YC, Liu XL, Man WQ, Zhang WK, Ma B, Chen SY, Zhang JS. The transcriptomic signature of developing soybean seeds reveals genetic basis of seed trait adaptation during domestication. The. Plant Journal. 2016;**86**:530-544. DOI: 10.1111/tpj.13181

[47] Li QT, Lu X, Song Q, Chen HW, Wei W, Tao JJ, Bi YD, Bian XH, Shen M, Ma B, Zhang WK, Li W, Lai YC, Lam SM, Shui GH, Chen SY, Zhang JS. Selection for a zinc-finger

protein contributes to seed oil increase during soybean domestication. Plant Physiology. 2017;**173**:01610. DOI: 10.1104/pp.16.01610

[48] Sasaki Y, Nagano Y. Plant acetyl-CoA carboxylase: Structure, biosynthesis, regulation, and gene manipulation for plant breeding. Bioscience, Biotechnology, and Biochemistry. 2004;**68**:1175-1184. DOI: 10.1271/bbb.68.1175

[49] Slabas AR, Fawcett T. The biochemistry and molecular biology of plant lipid biosynthesis. Plant Molecular Biology. 1992;**19**:169-191. DOI: 10.1007/978-94-011-2656-4_11

[50] Ohlrogge J, Browse J. Lipid biosynthesis. The Plant Cell. 1995;**7**:957. DOI: 10.1105/tpc.7.7.957

[51] Harwood JL. Recent advances in the biosynthesis of plant fatty acids. Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism. 1996;**1301**:7-56. DOI: 10.1016/0005-2760(95)00242-1

[52] Thelen JJ, Ohlrogge JB. Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering. 2002;**4**:12-21. DOI: 10.1006/mben.2001.0204

[53] Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J. Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiology. 1997;**113**:75-81. DOI: 10.1104/pp.113.1.75

[54] Zou J, Katavic V, Giblin EM, Barton DL, MacKenzie SL, Keller WA, Hu X, Taylor DC. Modification of seed oil content and acyl composition in the Brassicaceae by expression of a yeast sn-2 acyltransferase gene. The Plant Cell. 1997;**9**:909-923. DOI: 10.1105/tpc.9.6.909

[55] Jako C, Kumar A, Wei Y, Zou J, Barton DL, Giblin EM, Covello PS, Taylor DC. Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiology. 2001;**126**:861-874. DOI: 10.1104/pp.126.2.861

[56] Vigeolas H, Waldeck P, Zank T, Geigenberger P. Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnology Journal. 2007;**5**:431-441. DOI: 10.1111/j.1467-7652.2007.00252.x

[57] Kelly AA, Shaw E, Powers SJ, Kurup S, Eastmond PJ. Suppression of the SUGAR-DEPENDENT1 triacylglycerol lipase family during seed development enhances oil yield in oilseed rape (*Brassica napus* L.). Plant Biotechnology Journal. 2013;**11**:355-361. DOI: 10.1111/pbi.12021

[58] Kim M, Song JT, Bilyeu KD, Lee JD. A new low linolenic acid allele of GmFAD3A gene in soybean PE1690. Molecular Breeding. 2015;**35**:155. DOI: 10.1007/s11032-015-0352-z

[59] Van Erp H, Kelly AA, Menard G, Peter JE. Multigene engineering of triacylglycerol metabolism boosts seed oil content in Arabidopsis. Plant Physiology. 2014;**165**:30-36. DOI: 10.1104/pp.114.236430

[60] Katavic V, Reed DW, Taylor DC, Giblin EM, Barton DL, Zou J, MacKenzie SL, Covello PS, Kunst L. Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity. Plant Physiology. 1995;**108**:399-409. DOI: 10.1104/pp.108.1.399

[61] Zhang M, Fan J, Taylor DC, Ohlrogge JB. DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. The Plant Cell. 2009;**21**:3885-3901. DOI: 10.1105/tpc.109.071795

[62] Li M, Zhao M, Wu H, Wu W, Xu Y. Cloning, characterization and functional analysis of two type 1 diacylglycerol acyltransferases (DGAT1s) from *Tetraena mongolica*. Journal of Integrative Plant Biology. 2013;**55**:490-503. DOI: 10.1111/jipb.12046

[63] Wang ZK, Huang WJ, Chang JM, Sebastian A, Li YG, Li HY, Wu XX, Zhang BB, Meng FL, Li WB. Overexpression of SiDGAT1, a gene encoding acyl-CoA: Diacylglycerol acyltransferase from *Sesamum indicum* L. increases oil content in transgenic Arabidopsis and soybean. The Plant Cell, Tissue and Organ Culture (PCTOC). 2014;**119**:399-410. DOI: 10.1007/s11240-014-0543-z

[64] Hatanaka T, Serson W, Li R, Armstrong P, Yu K, Pfeiffer T, Li X, Hildebrand D. A Vernonia Diacylglycerol Acyltransferase can increase renewable oil production. Journal of Agricultural and Food Chemistry. 2016;**64**:7188-7194. DOI: 10.1021/acs.jafc.6b02498

[65] Li H, Zhao T, Wang Y, Yu D, Chen S, Zhou R, Gai J. Genetic structure composed of additive QTL, epistatic QTL pairs and collective unmapped minor QTL conferring oil content and fatty acid components of soybeans. Euphytica. 2011;**182**:117. DOI: 10.1007/s10681-011-0524-9

[66] Guo L, Ma F, Wei F, Fanella B, Allen DK, Wang X. Cytosolic phosphorylating glyceraldehyde-3-phosphate dehydrogenases affect Arabidopsis cellular metabolism and promote seed oil accumulation. The Plant Cell. 2014;**26**:3023-3035. DOI: 10.1105/tpc.114.126946

[67] Wilson RF. Seed composition. In: Boerma HR, Specht JE, editors. Soybeans: Improvement. Production and Uses. 2004. pp. 621-677

[68] Zeng QY, Yang CY, Ma QB, Li XP, Dong WW, Nian H. Identification of wild soybean miRNAs and their target genes responsive to aluminum stress. BMC Plant Biology. 2012;**12**:182. DOI: 10.1186/1471-2229-12-182

[69] Panthee DR, Pantalone VR, Saxton AM. Modifier QTL for fatty acid composition in soybean oil. Euphytica. 2006;**152**:67-73. DOI: 10.1007/s10681-006-9179-3

[70] Diers BW, Shoemaker RC. Restriction fragment length polymorphism analysis of soybean fatty acid content. Journal of the American Oil Chemists Society. 1992;**69**:1242-1244. DOI: 10.1007/BF02637690

[71] Bachlava E, Dewey RE, Burton JW, Cardinal AJ. Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. Crop Science. 2009;**49**:433-442

[72] Xie D, Han Y, Zeng Y, Chang W, Teng W, Li W. SSR-and SNP-related QTL underlying linolenic acid and other fatty acid contents in soybean seeds across multiple environments. Molecular Breeding. 2012;**30**:169-179. DOI: 10.1007/s11032-011-9607-5

[73] Spencer MM, Landau-Ellis D, Meyer EJ, Pantalone VR. Molecular markers associated with linolenic acid content in soybean. Journal of the American Oil Chemists' Society. 2004;**81**:559-562. DOI: 10.2135/cropsci2001.412493x

[74] Shibata M, Takayama K, Ujiie A, Yamada T, Abe J, Kitamura K. Genetic relationship between lipid content and linolenic acid concentration in soybean seeds. Breeding Science. 2008;**58**:361-366. DOI: 10.1270/jsbbs.58.361

[75] Li Z, Wilson RF, Rayford WE, Boerma HR. Molecular mapping genes conditioning reduced palmitic acid content in N87-2122-4 soybean. Crop Science. 2002;**42**:373-378. DOI: 10.2135/cropsci2002.3730

[76] Wang X, Jiang GL, Green M, Scott RA, Hyten DL, Cregan PB. Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. Molecular Breeding. 2012;**30**:1163-1179. DOI: 10.1007/s11032-012-9704-0

[77] Kim HK, Kim YC, Kim ST, Son BG, Choi YW, Kang JS, Choi IS, Park YH, Cho YS. Analysis of quantitative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean. Journal of Life Science. 2010;**20**:1186-1192. DOI: 10.5352/JLS.2010.20.8.1186

[78] McKeon T, Stumpf PK. Stearoyl-acyl carrier protein desaturase from safflower seeds. Methods in Enzymology. 1981;**71**:275-281

[79] Shanklin J, Somerville C. Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proceedings of the National Academy of Sciences. 1991;**88**:2510-2514. DOI: 10.1073/pnas.88.6.2510

[80] Smith MA, Cross AR, Jones OTG, Griffiths WT, Stymne S, Stobart K. Electron-transport components of the 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine △12-desaturase (△12-desaturase) in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons. Biochemical Journal. 1990;**272**:23-29. DOI: 10.1042/bj2720023

[81] Schlueter JA, Vasylenko-Sanders IF, Deshpande S, Yi J, Siegfried M, Roe BA, Schlueter SD, Scheffler BE, Shoemaker RC. The FAD2 gene family of soybean. Crop Science. 2007;**47**:S-14. DOI: 10.2135/cropsci2006.06.0382tpg

[82] Mathis L, Voytas DF, Zhang F, Haun W. U.S. Patent Application No. 14/208,027. 2014

[83] Wong CE, Zhao YT, Wang XJ, Croft L, Wang ZH, Haerizadeh F, Mattick JS, Singh MB, Carroll BJ, Bhalla PL. MicroRNAs in the shoot apical meristem of soybean. Journal of Experimental Botany. 2011;**62**:2495-2506. DOI: 10.1093/jxb/erq437

[84] Lakhssassi N, Zhou Z, Liu S, Colantonio V, AbuGhazaleh A, Meksem K. Characterization of the FAD2 gene family reveals the limitations of gel based TILLING in genes with high copy number. Frontiers in Plant Science. 2017;**8**:324

[85] Bilyeu KD, Palavalli L, Sleper DA, Beuselinck PR. Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. Crop Science. 2003;**43**: 1833-1838. DOI: 10.1016/j.ydbio.2005.10.036

[86] Flores T, Karpova O, Su X, Zeng P, Bilyeu K, Sleper DA, Nguyen HT, Zhang ZJ. Silencing of GmFAD3 gene by siRNA leads to low α-linolenic acids (18:3) of fad3-mutant phenotype in soybean [*Glycine max* (Merr.)]. Transgenic Research. 2008;**17**:839-850. DOI: 10.1007/s11248-008-9167-6

[87] Hoshino T, Watanabe S, Takagi Y, Anai T. A novel GmFAD3-2a mutant allele developed through TILLING reduces α-linolenic acid content in soybean seed oil. Breeding Science. 2014;**64**:371-377. DOI: 10.1270/jsbbs.64.371

[88] Romero P, Rodrigo MJ, Alferez F, Ballester AR, Gonzalez-Candelas L, Zacarias L, Lafuente MT. Unravelling molecular responses to moderate dehydration in harvested fruit of sweet orange (*Citrus sinensis* L. Osbeck) using a fruit-specific ABA-deficient mutant. Journal of Experimental Botany. 2012;**63**:2753-2767. DOI: 10.1093/jxb/err461

[89] Shinozaki K, Yamaguchi-Shinozaki K. Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signaling pathways. Current Opinion in Plant Biology. 2000;**3**:217-223. DOI: 10.1016/S1369-5266(00)80068-0

[90] Cernac A, Benning C. WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. The Plant Journal. 2004;**40**:575-585. DOI: 10.1111/j.1365-313X.2004.02235.x

[91] Focks N, Benning C. Wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiology. 1998;**118**:91-101. DOI: 10.1104/pp.118.1.91

[92] Mu JY, Tan HL, Zheng Q, Fu FY, Liang Y, Zhang JA, Yang XH, Wang T, Chong K, Wang XJ, Zuo JR. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiology. 2008;**148**:1042-1054. DOI: 10.1104/pp.108.126342

[93] Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer R, Goldberg RB, Harada JJ. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proceedings of the National Academy of Science of the United States of America. 2001;**98**:11806-11811. DOI: 10.1073/pnas.201413498

[94] Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM. Isolation of the Arabidopsis ABI3 gene by positional cloning. The Plant Cell. 1992;**4**:1251-1261. DOI: 0.1105/tpc.4.10.1251

[95] Luerssen H, Kirik V, Herrmann P, Misera S. FUSCA3 encodes a protein with a conserved VP1/ABI3-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. The Plant Journal. 1998;**15**:755-764. DOI: 10.1046/j.1365-313X.1998.00259.x

[96] Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell. 1998;**93**:1195-1205. DOI: 10.1016/S0092-8674(00)81463-4

[97] Baud S, Mendoza MS, To A, Harscoët E, Lepiniec L, Dubreucq B. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. The Plant Journal. 2017;**50**:825-838. DOI: 10.1111/j.1365-313X.2007.03092.x

[98] Baud S, Wuilleme S, To A, Rochat C, Lepiniec L. Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in Arabidopsis. The Plant Journal. 2009;**60**:933-947. DOI: 10.1111/j.1365-313X.2009.04011.x

[99] An D, Kim H, Ju S, Go YS, Kim HU, Suh MC. Expression of Camelina WRINKLED1 isoforms rescue the seed phenotype of the Arabidopsis wri1 mutant and increase the triacylglycerol content in tobacco leaves. Frontiers in Plant Science. 2017;**8**:34. DOI: 10.3389/fpls.2017.00034

[100] Ivarson E, Leiva-Eriksson N, Ahlman A, Kanagarajan S, Bülow L, Zhu LH. Effects of overexpression of WRI1 and hemoglobin genes on the seed oil content of *Lepidium campestre*. Frontiers in Plant Science. 2016;**7**:2032. DOI: 10.3389/fpls.2016.02032

[101] Sun RH, Ye RJ, Gao LC, Zhang L, Wang R, Mao T, Zheng YS, Li DD, Lin Y. Characterization and ectopic expression of CoWRI1, an AP2/EREBP domain-containing transcription factor from coconut (*Cocos nucifera* L.) endosperm, changes the seeds oil content in transgenic Arabidopsis thaliana and Rice (*Oryza sativa* L.). Frontiers in Plant Science. 2017;**8**:63. DOI: 10.3389/fpls.2017.00063

[102] Tsai AYL, Gazzarrini S. AKIN10 and FUSCA3 interact to control lateral organ development and phase transitions in Arabidopsis. Plant Journal. 2012;**69**:809-821. DOI: 10.1007/BF00398664

[103] Yagasaki K, Sakamoto H, Seki K, Yamada N, Takamatsu M, Taniguchi T, Takahashi K. Breeding of a new soybean cultivar "Nanahomare". Hokuriku. Crop Science. 2010;**45**:61-64

[104] Shen B, Sinkevicius KW, Selinger DA, Tarczynski MC. The homeobox gene GLABRA2 affects seed oil content in Arabidopsis. Plant Molecular Biology. 2006;**60**:377-387. DOI: 10.1007/s11103-005-4110-1

[105] Zhang J, Hao Q, Bai L, Xu J, Yin W, Song L, Xu L, Guo X, Fan C, Chen Y, Ruan J, Hao S, Li Y, Wang RR, Hu Z. Overexpression of the soybean transcription factor GmDof4 significantly enhances the lipid content of Chlorella ellipsoidea. Biotechnology for Biofuels. 2014;**7**:128. DOI: 10.1186/s13068-014-0128-4

[106] Zhang D, Zhao M, Li S, Sun L, Wang W, Cai C, Dierking EC, Ma J. Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the palaeopolyploid soybean. The Plant Journal. 2017;**90**:1120-1133. DOI: 0.1111/tpj.13533

[107] Li D, Jin C, Duan S, Zhu Y, Qi S, Liu KG, Gao CH, Ma H L, Zhang M, Liao YC, Chen M. MYB89 transcription factor represses seed oil accumulation. Plant Physiology. 2016;**173**:01634. DOI: 10.1104/pp.16.01634

[108] Duan S, Jin C, Li D, Gao C, Qi S, Liu K, Hai JB, Ma HL, Chen M. MYB76 inhibits seed fatty acid accumulation in Arabidopsis. Frontiers in Plant Science. 2017;**8**:226. DOI: 10.3389/fpls.2017.00226

[109] Chen M, Wang Z, Zhu Y, Li Z, Hussain N, Xuan L, Guo W, Zhang G, Jiang L. The effect of transparent TESTA2 on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in Arabidopsis. Plant Physiology. 2012;**160**:1023-1036. DOI: 10.1104/pp.112.202945

[110] Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, Eastmond PJ. bZIP 67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. The Plant Cell. 2013;**25**:3104-3116. DOI: 10.1105/tpc.113.116343

[111] Chen M, Zhang B, Li C, Kulaveerasingam H, Chew FT, Yu H. TRANSPARENT TESTA GLABRA1 regulates the accumulation of seed storage reserves in Arabidopsis. Plant Physiology. 2015;**169**:391-402

[112] Lee EJ, Oh M, Hwang JU, Li-Beisson Y, Nishida I, Lee Y. Seed-specific overexpression of the pyruvate transporter BASS2 increases oil content in Arabidopsis seeds. Frontiers in Plant Science. 2017;**8**:194. DOI: 10.3389/fpls.2017.00194

[113] Natarajan SS, Xu C, Bae H, Bailey BA, Cregan P, Caperna TJ, Garrette WM, Luthriaf D. Proteomic genetic analysis of glycinin subunits of sixteen soybean genotypes. Plant Physiology and Biochemistry. 2007;**45**:4366-4444. DOI: 10.1016/j.plaphy.2007.03.031

[114] Ferreira RB, Franco E, Teixeira AR. Calcium and magnesium-dependent aggregation of legume seed storage proteins. Journal of Agricultural and Food Chemistry. 1999;**47**:3009-3015. DOI: 10.1021/jf981151c

[115] Peng IC, Quass DW, Dayton WR, Allen CE. Physicochemical and functional properties of soybean 11S globulin – A review. Cereal Chemistry. 1984;**61**:480-490. DOI: 没找到

[116] Derbyshire E, Wright DB, Boulter D. Legumin and vicilin, storage proteins of legume seeds. Phytochemistry. 1976;**15**:3-24. DOI: 10.1016/S0031-9422(00)89046-9

[117] Hill JE, Breidenbach RW. Proteins of soybean seeds I. Isolation and characterization of the major components. Plant Physiology. 1974;**53**:742-746. DOI: 10.1104/pp.53.5.742

[118] Biochemistry KHB. Molecular biology of soybean seed storage proteins. Journal of New Seeds. 2000;**2**:1-25. DOI: 10.1300/J153v02n03_01

[119] Krishnan HB, Oehrle NW, Natarajan SS. A rapid and simple procedure for the depletion of abundant storage proteins from legume seeds to advance proteome analysis: A case study using *Glycine max*. Proteomics. 2009;**9**:3174-3188. DOI: 10.1002/pmic.200800875

[120] Thanh VH, Shibasaki K. Major proteins of soybean seeds. A straight forward fractionation and their characterization. Journal of Agricultural and Food Chemistry. 1976;**24**:1117-1121. DOI: 10.1021/jf60208a030

[121] Kitamura K, Davies CS, Nielsen NC. Inheritance of alleles for Cgy 1 and Gy 4 storage protein genes in soybean. TAG Theoretical and Applied Genetics. 1984;**68**:253-257. DOI: 10.1007/BF00266899

[122] Ladin BF, Doyle J, Beachy RN. Molecular characterization of a deletion mutation affecting the α'-subunit of β-conglycinin of soybean. Journal of Molecular and Applied Genetics. 1984;**2**:372-380

[123] Yagasaki K, Kaizuma N, Kitamura K. Inheritance of glycinin subunits and characterization of glycinin molecules lacking the subunits in soybean (*Glycine max* (L.)Merr.). Japanese Journal of Breeding. 1996;**46**:11-15. DOI: 10.1270/jsbbs1951.46.11

[124] Hajika M, Takahashi M, Sakai S, Matsunaga R. Dominant inheritance of a trait lacking β-conglycinin detected in a wild soybean line. Japanese Journal of Breeding. 1998;**48**:383-386. DOI: 10.1270/jsbbs1951.48.38366

[125] Koshiyama I, Gottschalk W, Muller HP, editors. Seed Proteins Biochemistry, Genetics, Nutritive Value (Advances in Agricultural Biotechnology). Dordrecht: Kluwer Academic Publishers; 1983. pp. 427-450

[126] Renkema JMS, Knabben JHM, Vliet T. Gel formation by b-conglycinin and glycinin and their mixtures. Food Hydrocolloids. 2001;**15**:407-414. DOI: 10.1016/S0268-005X(01)00051-0

[127] Tierney ML, Bray EA, Allen RD, Ma Y, Drong RF, Slightom J, Beachy RN. Isolation and characterization of a genomic clone encoding the b-subunit of b-conglycinin. Planta. 1987;**172**:356-363. DOI: 10.1007/BF00398664

[128] Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB. Characterization of the glycinin gene family in soybean. The Plant Cell. 1989;**1**:313-328. DOI: 10.1105/tpc.1.3.313

[129] Panthee DR, Kwanyuen P, Sams CE, West DR, Saxton AM, Pantalone VR. Quantitative trait loci for β-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. Journal of the American Oil Chemists' Society. 2004;**81**:1005-1012. DOI: 10.1007/s11746-004-1014-4

[130] Ma Y, Kan G, Zhang X, Wang Y, Zhang W, Du H, Yu D. Quantitative trait loci (QTL) mapping for glycinin and β-conglycinin contents in soybean (*Glycine max* L. Merr.). Journal of Agricultural and Food Chemistry. 2016;**64**:3473-3483. DOI: 10.1021/acs.jafc.6b00167

[131] Nielsen NC, Bassüner R, Beaman T. The biochemistry and cell biology of embryo storage proteins. In: Brian AL, Indra KV, editors. Cellular and Molecular Biology of Plant Seed Development. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1997. pp. 151-220. DOI: 10.1007/978-94-015-8909-3_5

[132] Harada JJ, Barker SJ, Goldberg RB. Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. The Plant Cell. 1989;**1**:415-425. DOI: 10.1105/tpc.1.4.415

[133] Beilinson V, Chen Z, Shoemaker R, Fischer R, Goldberg R, Nielsen N. Genomic organization of glycinin genes in soybean. Theoretical and Applied Genetics. 2002;**104**:1132-1140. DOI: 10.1007/s00122-002-0884-6

[134] Asakura T, Tamura T, Terauchi K, Narikawa T, Yagasaki K, Ishimaru Y, Abe K. Global gene expression profiles in developing soybean seeds. Plant Physiology and Biochemistry. 2012;**52**:147-153. DOI: 10.1016/j.plaphy.2011.12.007

[135] Withana-Gamage TS, Hegedus DD, Qiu X, Yu PQ, May T, Lydiate D, Wanasundara JPD. Characterization of Arabidopsis thaliana lines with altered seed storage protein profiles using synchrotron-powered FT-IR Spectromicroscopy. Journal of Agriculture and Food Chemistry. 2013;**6**:901-912. DOI: 10.1021/jf304328n

[136] Wang J, Liu L, Guo Y, Wang YH, Zhang L, Jin LG, Guan RX, Liu ZX, Wang LL, Chang RZ, Qiu LJ. A dominant locus, qBSC-1, controls beta subunit content of seed storage protein in soybean (*Glycine max* (L.) Merri.). Journal of Integrative Agriculture. 2014;**13**:1854-1864. DOI: 10.1016/S2095-3119(13)60579-1

[137] Yamada T, Mori Y, Yasue K, Maruyama N, Kitamura K, Abe J. Knockdown of the 7S globulin subunits shifts distribution of nitrogen sources to the residual protein fraction in transgenic soybean seeds. The Plant Cell reports. 2014;**33**:1963-1976. DOI: 10.1007/s00299-014-1671-y

[138] Utsumi S, Kinsella JE. Forces involved in soy protein gelation: Effects of various reagents on the formation, hardness and solubility of heat-induced gels made from 7S, 11S and soy isolate. Journal of Food Science. 1985;**50**:1278-1282. DOI: 10.1111/j.1365-2621.1985.tb10461.x

[139] Wright DJ. The seed globulins. Developments in food. Proteins. 1987;**5**:81-157

[140] Ogawa T, Tayama E, Kitamura K, Kaizuma N. Genetic improvement of seed storage proteins using three variant alleles of 7S globulin subunits in soybean (*Glycine max* L.). Japanese Journal of Breeding. 1989;**39**:137-147. DOI: 10.1270/jsbbs1951.39.137

[141] Yang A, Yu X, Zheng A, James AT. Rebalance between 7S and 11S globulins in soybean seeds of differing protein content and 11SA4. Food Chemistry. 2016;**210**:148-155. DOI: 10.1016/j.foodchem.2016.04.095

[142] Takahashi M, Uematsu Y, Kashiwaba K, Yagasaki K, Hajika M, Matsunaga R, Komatsu K, Ishimoto M. Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. Planta. 2003;**217**:577-586. DOI: 10.1007/s00425-003-1026-3

[143] Tezuka M, Taira H, Igarashi Y, Yagasaki K, Ono T. Properties of tofus and soy milks prepared from soybeans having different subunits of glycinin. Journal of Agricultural and Food Chemistry. 2000;**48**:1111-1117. DOI: 10.1021/jf990560l

[144] Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L. Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. The Plant Journal. 2008;**54**:608-620. DOI: 10.1111/j.1365-313X.2008.03461.x

[145] Vicente-Carbajosa J, Carbonero P. Seed maturation: Developing an intrusive phase to accomplish a quiescent state. The International Journal of Developmental Biology. 2005;**49**:645-651. DOI: 10.1387/ijdb.052046jc

[146] Che NY, Yang Y, Li YD, Wang LL, Huang P, Gao Y, An CC. Efficient LEC2 activation of OLEOSIN expression requires two neighboring RY elements on its promoter. Science in China Series C-Life Sciences. 2009;**52**:854-863. DOI: 10.1007/s11427-009-0119-z

[147] Romanel EA, Schrago CG, Couñago RM, Russo CA, Alves-Ferreira M. Evolution of the B3 DNA binding superfamily: New insights into REM family gene diversification. PLoS One. 2009;**4**:e5791. DOI: 10.1371/journal.pone.0005791

[148] Sakata Y, Nakamura I, Taji T, Tanaka S, Quatrano RS. Regulation of the ABA-responsive Em promoter by ABI3 in the moss *Physcomitrella patens*: Role of the ABA response element and the RY element. Plant Signal. 2010;**5**:1061-1066. DOI: 10.4161/psb.5.9.11774

[149] Delahaie J, Hundertmark M, Bove J, Leprince O, Rogniaux H, Buitink J. LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. Journal of Experimental Botany. 2013;**64**:4559-4573. DOI: 10.1093/jxb/ert274

[150] Du R, Qiao Y, Wang X, Lv X, Wang J. Establishment and analysis of the mRNA expression patterns of ABI3-like and storage protein genes during soybean seed development. Emirates Journal of Food and Agriculture. 2016;**28**:177

[151] Harada JJ. Role of Arabidopsis leafy cotyledon genes in seed development. Journal of Plant Physiology. 2001;**158**:405-409. DOI: 10.1078/0176-1617-00351

[152] Kirik V, Kölle K, Balzer HJ, Bäumlein H. Two new oleosin isoforms with altered expression patterns in seeds of the Arabidopsis mutant FUS3. Plant Molecular Biology. 1996;**31**:413-417. DOI: 10.1007/BF00021803

[153] To A, Valon C, Savino G, Guilleminotb J, Devicb M, Giraudata J, Parcy F. A network of local and redundant gene regulation governs Arabidopsis seed maturation. The Plant Cell. 2006;**18**:1642-1651. DOI: doi.org/10.1105/tpc.105.039925

[154] Crowe AJ, Abenes M, Plant A, Moloney MM. The seed-specific transactivator, ABI3, induces oleosin expression. Plant Science. 2000;**151**:171-181. DOI: 10.1016/S0168-9452(99)00214-9

[155] Soderman EM, Brocard IM, Lynch TJ, Finkelstein RR. Regulation and function of the Arabidopsis ABA-insensitive4 gene in seed and abscisic acid response signaling networks. Plant Physiology. 2000;**124**:1752-1765

[156] Ezcurra I, Wycliffe P, Nehlin L, Ellerstrom M, Rask L. Transactivation of the Brassica napus napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. The Plant Journal. 2000;**24**:57-66. DOI: 10.1046/j.1365-313x.2000.00857.x

[157] Reidt W, Wohlfarth T, Ellerstrom M, Czihal A, Tewes A, Ezcurra I, Rask L, Baumlein H. Gene regulation during late embryogenesis: The RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. The Plant Journal. 2000;**21**:401-440. DOI: 10.1046/j.1365-313x.2000.00686.x

[158] Kroj T, Savino G, Valon C, Giraudat J, Parcy F. Regulation of storage protein gene expression in Arabidopsis. Development. 2003;**130**:6065-6073. DOI: 10.1242/dev.00814

[159] Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ. Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proceeding of the National Academy of Science of the United States of America. 2006;**103**:3468-3473. DOI: 10.1073/pnas.0511331103

[160] Baud S, Boutin JP, Miquel M, Lepiniec L, Rochat C. An integrated overview of seed development in Arabidopsis thaliana ecotype WS. Plant Physiology Biochemistry. 2002;**40**:151-160. DOI: 10.1016/S0981-9428(01)01350-X

[161] Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell Physiology. 2005;**46**:399-406. DOI: 10.1093/pcp/pci048

[162] Fatihi A, Boulard C, Bouyer D, Bauda S, Dubreucqa B, Lepinieca L. Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. Plant Science. 2016;**250**:198-204. DOI: 10.1016/j.plantsci.2016.06.013

[163] Abraham Z, Fernández RI, Martinez M, Diaz I, Carbonero P, Vicente-Carbajosa J. A developmental switch of gene expression in the barley seed mediated by HvVP1 (Viviparous1) and HvGAMYB interactions. Plant Physiology. 2016;**170**:2146-2158. DOI: 10.1104/pp.16.00092

[164] Baud S, Kelemen Z, Thévenin J, Boulard C, Blanchet S, To A, Payre M, Berger M, Effroy-Cuzzi D, Franco-Zorrilla JM, Godoy M, Solano R, Thevenon E, Parcy F, Lepiniec L, Dubreucq B. Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by L-AFL proteins in Arabidopsis seed. Plant Physiology. 2016;**171**:1099-1112. DOI: 10.1104/pp.16.00034

[165] Roesler K, Bermudez E, Damude HG, Li C, Meyer K, Shen B, Tarczynski MC. U.S. Patent Application, No. 15/239,069. 2016

[166] Carrington JC, Ambros V. Role of MicroRNAs in plant and animal development. Science. 2003;**301**:336-338. DOI: 10.1126/science.1085242

[167] Zhang B, Pan X, Cobb GP, Anderson TA. Plant microRNA: A small regulatory molecule with big impact. Developmental Biology. 2006;**289**:3-16. DOI: 10.1016/j.ydbio.2005.10.036

[168] Yu J, Zhang Z, Wei J, Ling Y, Xu Y, Su Z. SFGD: A comprehensive platform for mining functional information from soybean transcriptome data and its use in identifying acyl-lipid metabolism pathways. BMC Genomics. 2014;**15**:271. DOI: 10.1186/1471-2164-15-271

[169] Joshi T, Patil K, Fitzpatrick MR, Franklin LD, Yao Q, Cook JR, Wang Z, Libault M, Brechenmacher L, Valliyodan B, Wu X, Cheng J, Stacey G, Nguyen HT, Xu D. Soybean Knowledge Base (SoyKB): A web resource for soybean translational genomics. BMC Genomics. 2012;**13**:S15-S15. DOI: 10.1186/1471-2164-13-S1-S15

[170] Goettel W, Liu Z, Xia J, Zhang W, Zhao PX, An YQ. Systems and evolutionary characterization of microRNAs and their underlying regulatory networks in soybean cotyledons. PLoS One. 2014;**9**:e86153. DOI: 10.1371/journal.pone.0086153

[171] Ye CY, Xu H, Shen E, Liu Y, Wang Y, Shen Y, Qiu J, Zhu QH, Fan L. Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean. Frontiers in Plant Science. 2014;**5**:743. DOI: 10.3389/fpls.2014.00743

[172] Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. Genome sequence of the palaeopolyploid soybean. Nature. 2010;**463**:178-183. DOI: 10.1038/nature08670

[173] Li MW, Xin D, Gao Y, Li KP, Fan K, Muñoz NB, Yung WS, Lam HM. Using genomic information to improve soybean adaptability to climate change. Journal of Experimental Botany. 2016:erw348. DOI: 10.1093/jxb/erw348

Chapter 3

Breeding Cultivars for Heat Stress Tolerance in Staple Food Crops

Mahalingam Govindaraj, Santosh K. Pattanashetti,
Nagesh Patne and Anand A. Kanatti

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76480

Abstract

Food and nutritional security will be worsened by climate change-induced high temperatures, droughts and reduced water availability in most agricultural food crops environments, particularly in developing countries. Recent evidences indicate that countries in the southern hemisphere are more vulnerable to food production due to greater frequency of extreme weather events. These challenges can be addressed by: (i) adoption of climate mitigation tools in agricultural and urban activities; (ii) development of heat and drought tolerant cultivars in major food crops; (iii) bringing back forgotten native minor food crops such as millets and root crops; and (iv) continued investment in agricultural research and development with the strong government policy support on native crops grown by small holder farmers. The native crops have inherent potential and traits to cope with adverse climate during the course of its evolution process. Therefore, diversifying the crops should be a prime framework of the climate-smart agriculture to meet the global food and nutritional security for which policy-driven production changes are highly required in developing countries. The adverse effects of climate change on agricultural production need to be addressed by multidisciplinary team and approaches through strong network of research consortium including private sectors and multinational governments for global impact.

Keywords: breeding, cultivars, heat stress, staple food crops, tolerance

1. Introduction

The major challenge of this century is to produce sufficient food to meet the ever-growing population (10 billion by 2050) despite reductions in quantity and quality of arable land,

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

water and increasingly variable weather patterns that are associated with climate change [1]. Abiotic stresses such as drought, salt, cold, and high temperature continue to affect the crops individually or in combination. Climate change has increased the intensity of heat stress that adversely affects both agricultural and horticultural crops resulting in serious economic losses, particularly in agricultural dependent countries. Global climate change risks are expected to be as high as global mean temperature increase of ≥4°C would pose large risks to global and regional food security [2]. The combination of high temperature and humidity would be compromising the current production of the major food crops such as wheat, rice, and maize in tropical and temperate regions. The climate change without adaptation is projected to negatively impact production for local temperature increases of ≥2°C above the late twentieth-century levels [2].

Extreme climates including very high temperatures are predicted to have a general negative effect on crop growth and development leading to catastrophic loss of crop productivity and also widespread famine in future [3]. The increase of temperature by 3–4°C is expected to reduce crop yields by 15–35% in Africa and Asia, whereas by 25–35% in the Middle East [4]. Hence, adopting the mitigation strategies such as reforestation, water harvesting in field and households, optimal use of CO_2 emitting devices and reducing wetland crops to avoid methane emission, etc., are essential, but as they are dependent on government policies, it is difficult to achieve them in short term. For long term, the adoption of extinct native crops and its diversity in individual farm is highly required to meet not only the food and nutritional security but also the feed security for farm livestock. The development of heat stress tolerant cultivars would be an ideal solution for sustainable food production for which research is still in preliminary stage and needs donor investment to progress competitively to deliver climate-smart cultivars to farmers. A recent study has shown the climatic shift in >25% of its geographical area in India [5] and also significant increase of aridity in several parts of the country. Therefore, government needs to re-standardize the climate zones with respect to aridity and temperature while planning for any developmental and agricultural intervention.

2. Heat stress and associated effects on food productions

A threshold temperature (TT) refers to a value of daily mean temperature at which a measurable reduction in growth begins. This is the range wherein changes in the photosynthetic capacity are irreversible, but other characters such as growth, flowering, etc., are reversible. The upper and lower developmental threshold temperatures are the ones at which growth and development ceases and they differ based on the plant species and genotypes. Cool season and temperate crops often have lower threshold temperature values compared to tropical crops. Every crop plants have threshold temperatures for different developmental stages (**Table 1**); upon exceeding this, crop experiences the stress.

High-temperature sensitivity is particularly important in tropical and subtropical climates as heat stress may become a major limiting factor for field crop production [6]. Heat stress (HS) is often defined as the rise in temperature beyond a threshold level for a period of time sufficient to cause irreversible damage to plant growth and development. A transient increase in

Crops	Threshold temp. (°C)	Developmental stage
Rice	15–35	Germination
	33	Biomass
	25	Grain formation and yield
	34	Grain yield and quality
Wheat	10–35	Germination
	20–30	Vegetative
	15	Reproductive
	35	Postanthesis
	35	Protein accumulation
Maize	15–40	Germination
	33–38	Photosynthesis
	38	Vegetative
	36–40	Pollen viability and fertilization
Sorghum	20–40	Germination
	26–34	Vegetative
	25–28	Reproductive
Pearl millet	10–34	Germination
Chickpea	10–35	Germination
	15–30	Growth
	25	Reproductive growth
Common bean	23	Reproductive development
Pea	15–20	Vegetative growth
Soybean	26	Reproductive development
	23	Post-anthesis
	30.2	Pollen germination
	36.1	Pollen tube growth
Groundnut	10–41	Germination
	29–33	Vegetative development
	25–28	Vegetative growth
	22–24	Reproductive growth
Lentil	32/20	Reproductive stage
Cotton	31.8–43.3	Pollen germination
	28.6–42.9	Pollen tube growth

Table 1. Critical growth stages and threshold temperatures of important food crops.

temperature of 10–15°C above ambient is generally considered as heat shock or heat stress (HS). However, HS is a complex function of intensity (temperature in degrees), duration, and rate of increase in temperature [6]. Some researchers believe night temperatures are major limiting factor, while the others argue that day and night temperatures do not affect the plant independently. Hence, diurnal temperature is a better predictor of plants response to high temperature with day temperature having a secondary role [7]. At high temperatures, severe cellular injury and even cell death may occur within minutes, which could be attributed to catastrophic collapse of cellular organization [8]. At moderately high temperatures, injuries or death may occur only after long exposure. Direct injuries due to high temperatures include protein denaturation and aggregation, and increased fluidity of membrane lipids. Indirect or slower heat injuries include inactivation of enzymes in chloroplast and mitochondria, inhibition of protein synthesis, protein degradation and loss of membrane integrity [9]. These injuries eventually lead to starvation, inhibition of growth, reduced ion flux, production of toxic compounds and reactive oxygen species (ROS) [10, 11].

Plants can experience wide range of HS on daily or seasonal basis. Temperature plays an important role in all stages of crops such as seedling emergence, vegetative stage, flowering/reproductive, and grain filling stages. Optimal temperature for growth and development differs for different plant species and genotypes within species. Exposure to temperature outside optimal range though not necessary be lethal, but can be stressful. The observed effects depend on species and genotype, with abundant inter- and intra-specific variations [12, 13]. Under elevated temperatures, various physiological injuries have been observed such as scorching of leaves and stems, leaf abscission and senescence, shoot and root growth inhibition, or fruit damage that leads to decreased plant productivity [14]. Heat stress induces changes in respiration and photosynthesis and thus leads to a shortened life cycle and diminished plant productivity [12]. In many crop species, the effects of high-temperature stress are more prominent on reproductive development than on vegetative growth, and the sudden decline in yield with temperature is mainly associated with pollen infertility [15, 16]. The effect of high temperature among different crop plants during germination, vegetative growth, reproductive growth, and different physiological processes such as photosynthesis, membrane fluidity, respiration, water balance, oxidative stress, and antioxidant defense have been discussed in detail elsewhere [17].

A wide range of plant developmental and physiological processes are negatively affected by HS. Sexual reproduction and flowering in particular have been recognized as extremely sensitive to HS that often results in reduced crop productivity [18]. High temperature is found most deleterious at flower bud initiation stage, with sensitivity being maintained for 10–15 days [18, 19]. Many legumes and cereals show a high sensitivity to HS, during flowering, and cause severe reductions in seed set probably due to reduced water and nutrient transport during reproductive development [15]. Generally, the male gametophyte is found sensitive to high temperatures at all stages of development, while the pistil and female gametophyte are more tolerant [20]. But in pearl millet, female reproductive parts was found more heat sensitive than male [21] as the stigma protrudes out of the florets. The HS often accelerates rather than delays the onset of anthesis that means the reproductive phase will be initiated prior to the accumulation of sufficient resources [16]. Shorter developmental phases for field crops have relatively negative effects on final grain weight and yield [22, 23]. Male sterility due to HS is widely observed among many sensitive crop plants, wherein the impairment of pollen

development has been the prime factor for reduced yield under HS [13, 24]. Continuing HS beyond a successful fertilization can also halt further development of the embryo [12]. HS during seed development may result in reduced germination and loss of vigor, leading to reduced emergence and seedling establishment as noted among several crop plants [25, 26]. Both grain weight and grain number appears to be affected by HS in many temperate cereal crops, wherein decline in grain number was found directly proportional to increasing temperatures during flowering and grain filling [27, 28]. HS during seed development in several crop species has been found to cause reductions in quality parameters such as starch, protein, and total oil yield [29]. High temperature affects different stage and part of the crop growth in terms of morphology. The anatomical and phenological changes also affect plant growth and development.

3. Heat tolerance mechanisms in food crops

Heat tolerance (HT) is generally defined as the ability of the plant to grow and produce economic yield under high temperatures [6]. For surviving under HS, crop plants could manifest short-term (avoidance) and long-term (evolutionary changes) strategies. Short-term avoidance or acclimation mechanisms include changing leaf orientation, transcriptional cooling, altering membrane lipid composition, reflecting solar radiation, leaf shading of tissues that are sensitive to sunburn, and extensive rooting [30, 31]. Early maturation is found closely related to smaller yield losses in many crop plants [32], which is mainly due to escape mechanism. For example, tolerant wheat genotypes are defined by maintenance of photosynthesis, chlorophyll content, and stomatal conductance under heat stress, while the yield of these genotypes is maintained through higher seed set, grain weight, and extended grain filling duration (GFD) even at elevated temperatures [33]. At supraoptimal temperatures, heat tolerance grass species and cultivars exhibit higher activity in the photosynthetic apparatus [34, 35] and higher carbon allocation and nitrogen uptake rates [36]. Plants also utilize various mechanisms against HS such as ion transport, osmoprotectants, free-radical scavengers, and late embryogenesis abundant (LEA) proteins, wherein factors ubiquitin and dehydrin involved in signaling cascades and transcriptional control are essentially significant to counteract stress effects [37].

Transpiration is a mechanism of heat avoidance and serves as the primary mediator of energy dissipation. Generally, the rate of transpiration increases with increasing of canopy temperatures due to its effects on both vaporization and vapor pressure deficit (VPD). Crop transpiration is the most active and common method of cooling crop tissues (transpiration cooling effect), with plant cooling requirements increasing with temperature [10]. The ability to maintain high stomatal conductance at high temperatures promotes transpirational heat dissipation, as observed in heat tolerance bread and durum wheat genotypes [11] and various heat tolerance and sensitive chickpea genotypes [38]. Heat stress has been known to cause malfunction of photosystem (PS) II, reduced efficiency in electron transport, and increase in ROS production. Heat tolerance has been linked to increased tolerance of the photosynthetic apparatus [39]. ROS detoxification mechanisms are known to play important roles in protecting plants against HS [40, 41]. HT is closely correlated with increased capacity of scavenging and detoxifying of ROS. Induction of thermotolerance may be ascribed to maintenance of a better membrane thermostability and low ROS accumulation [36, 42] due to an improved antioxidant capacity [43].

HS responses among plants are mainly due to their inherent ability to survive and also to acquire thermotolerance to lethal temperatures. Genetic variability among crops for HT is mainly due to expression of different stress-responsive genes [44], acquisition of thermotolerance, and synthesis and accumulation of HSPs that are well correlated with the antioxidant defense system [45]. The maintenance of high membrane thermostability (MTS) is related to thermotolerance [46] and an important selection criterion which is determined by measuring the electrical conductivity. MTS has been successfully employed to assess thermotolerance in many food crops worldwide. The role of thermoprotectants such as HSPs, proline, glycine betaine, trehalose, brassinosteroids, salicylic acid, abscisic acid, polyamines, and nitric oxide in offering heat tolerance through endogenous synthesis or by exogenous application in different crops has been discussed in detail by Kaushal et al. [17]. Future pioneering studies in model plants can pave the way to identify key regulators as target for gene manipulation of stress tolerance in crop plants. It has also been envisaged that metabolic fingerprinting can be used as breeding tool for development of plants with the best potential to tolerate abiotic stresses.

4. Screening methodologies for heat tolerance breeding

Efficient screening procedures and identification of key traits in diverse donor or tolerant lines are very much essential toward breeding for heat tolerance. Screening for heat tolerance in the field is very challenging due to interactions with other environmental factors, but a wide variety of relevant traits are available that allows successful selection in the field conditions [47]. Tolerant genotypes may also be selected in controlled environments provided validated screening tools are in place. However, very often the more expensive controlled environments do not allow natural selection for other factors that interact with the heat stress tolerance mechanisms under field conditions, thus limiting its potential of wider applications in any trait screening [48]. Heat tolerance can be evaluated by a variety of viability assays, measurements, visual assessment, and testing under hotspot locations as described below.

- **i.** *Cell membrane thermo-stability test*: Cellular membrane dysfunction due to stress leads to increased permeability and leakage of ions, which can be readily measured by the efflux of electrolytes from affected leaf tissue into an aqueous medium. This method was initially developed by the C.Y. Sullivan (University of Nebraska) in the late 1960s for assessing sorghum and maize heat tolerance. This has been used to study cellular thermostability for heat in wheat [49, 50], soybean [51], maize [52], and chickpea [53]. A positive correlation between membrane injury and grain weight was observed in wheat suggesting that membrane thermostability (MTS) may be better indicator of heat tolerance [54]. The membrane thermostability (MTS) can be measured as follows: MTS = $(1 - T1/T2) \times 100$, where $T1$ is conductivity reading after heat treatment and $T2$ is conductivity reading after autoclaving [55]. This has been tested in pearl millet and found effective under field condition and thus can be used for screening large number of genotypes.
- **ii.** *Chlorophyll fluorescence measurement*: Heat damage in photosynthetic tissue can be measured by chlorophyll fluorescence [56]. Chlorophyll fluorescence has been linked to a thermal kinetic window established by enzymatic assays [57]. In this approach, leaf discs are exposed to a brief illumination period and the time of dark recovery of the fluorescence parameter F_v/F_o (ratio of variable to minimum chlorophyll fluorescence) is determined

as a function of temperature. It is simple, quick, and inexpensive and holds promise for the rapid screening in a large number of crops, e.g., wheat [58] and legumes (pigeon pea, chickpea, groundnut, and soybean) [59].

iii. *Estimation of membrane lipid saturation*: A higher share of saturated fatty acids in membrane lipids increases the lipid melting temperature and prevents a heat-induced increase in the membrane fluidity. To maintain the membrane fluidity, plants increase the content of saturated and monounsaturated fatty acids, modulating their metabolism in response to increasing temperatures [60]. Thus, increasing the saturation level of fatty acids appears to be critical for maintaining the membrane stability and enhancing heat tolerance in creeping bentgrass (*Agrostis stolonifera*) [61].

iv. *Canopy temperature depression (CTD)*: The surface temperature of the canopy is related to the amount of transpiration resulting in evaporative cooling. A hand-held infrared thermometer (IRT) allows canopy temperature (CT) to be measured directly and easily during afternoon (13:00 and 14:30 h) remotely and without interfering with the crop. The viewing angle should be around 40° to the horizontal line above the canopy so as to avoid the confounding effect of soil temperature. Studies have shown that CT is correlated with many physiological factors: stomatal conductance, transpiration rate, plant water status, water use, leaf area index, and crop yield. Genotypes with cooler canopy temperatures can be used to indicate a better hydration status. Under heat stress conditions, CTD is related to vascular capacity, cooling mechanism, and heat adaptation. CTD has been proved to be a rapid and stable test that can be used for selection, e.g., wheat [62].

v. *Visual assessment methods/morphological methods*: Male sterility in cowpea [63]; pollen viability, stigma receptivity in maize [64]; grain sterility in rice [65]; asynchrony of male and female floral organ development in chickpea [66], leaf firing, tassel blasting, tassel sterility, pollen viability, silk receptivity and some agronomic traits in maize [67].

vi. *Selection in hot production environments*: It has been effective in wheat [68] and maize [67]. Heat stress screening, one criteria for selection of site is high VPD area where low yield was found associated with high VPD during all the growing season, high maximum temperature during most of the growing season, and low photothermal quotient corrected by VPD in the critical period of grain set before flowering. The relationships found are agronomically robust and provide a guide for experimental research but cannot be taken as proof of cause-and-effect because weather variables are confounded [69].

vii. *Pollen-based screening of genotypes*: Using this method, various heat tolerance accessions have been identified in different crop species, e.g., DG 5630RR in soybean [70], AZ100 in maize [71], and ICC1205 and ICC15614 in chickpea [28, 72].

All these techniques need to be validated for a large number of crops for their applicability in future. Regardless of the screening method, a key objective for plant breeders is to develop an effective set of thermotolerance markers which can be used for further implementation of breeding for heat tolerance in various crop species.

Identification of the superior germplasm for heat tolerance is essential for effective genetic manipulation through breeding process. However, identification of reliable and effective heat screening methods is a major challenge in conventional breeding to facilitate detection of heat tolerance lines [6]. Although a number of screening methods and selection criteria that

have been developed/proposed by different researchers are briefly discussed above, however, the primary field screening methods also include seedling thermo-tolerance index (STI) [73], seed to seedling thermo-tolerance index (SSTI) in pearl millet [74], and heat tolerance index (HTI) as growth recovery after heat exposure in sorghum [75]. Thermo-tolerance screening at germination and early vegetative stage is found effective for pearl millet and maize [76]. These field techniques would help in preliminary identification of heat tolerance lines and thus proceed with minimum number of lines for further screening and validation. At the same time, breeder should ensure the quality of individual line data by comparing with tolerant check at all the times. This will facilitate the more reliable way of advancing the heat tolerance genotypes in any afore-discussed screening tools.

5. Breeding for heat tolerance: a next-generation breeding approach

The emphasis of crop improvement has been primarily on improving the economic yield in majority of the crop plants. This targeted breeding for economically desirable traits in crops has resulted in reduced genetic variability in the commercial varieties/hybrids to reach homogeneity in appearance. Hence, other essential genes that enable growth and reproduction in adverse environments can be absent from modern cultivars owing to their exclusion or loss during domestication or subsequent germplasm improvement, and linkage to non-beneficial loci or drag on productivity in optimal environments [1]. Useful loci and allelic variants often correspond to the downregulation or disruption of genes in susceptible genotypes. Hence, continued and broader mining of germplasm could be advantageous. Toward breeding for heat tolerance in crops, the initial search for tolerant sources should begin among the modern cultivars/advanced breeding lines and landraces of the species. The further search should be shifted to primary and secondary gene pool in sequence. For efficient utilization of the identified sources, we need to understand the underlying component traits, their inheritance including genes/QTLs and also association among important traits. All this information generated would facilitate targeted breeding for heat tolerance in crops.

5.1. Germplasm as sources for heat tolerance breeding

Wide variation for heat tolerance has been noted in both cultivated and related wild species among different crop plants (**Table 2**). Landraces are the varieties preferably handled by local farmers which are adapted to their native environment and could be the potential sources of HT. Significant variability for HT has been noted among wheat landraces, wherein tolerant ones tend to have higher leaf chlorophyll contents [64] and higher stomatal conductance, which can be utilized in breeding programs. Early maturity under high-temperature conditions is closely correlated with lesser yield losses in many crop plants. In wheat, early heading varieties performed better than later-heading varieties because they (i) produced fewer leaves per tiller and retained more green leaves, (ii) had longer grain-filling periods, and (iii) completed grain filling earlier in the season when air temperatures were lower [107].

The early maturity-led escape mechanism enabled addressing heat stress in wheat in Eastern Gangetic Plains and various South Asian locations [108, 109]. Selection for early flowering and maturity has also enabled to escape heat stress in spring-sown chickpea in Mediterranean

Crop	Heat tolerance sources*	HT associated trait/index	References
Wheat	CWI # 59788, 60155, 60391	Leaf chlorophyll content (LCC) and canopy temperature depression (CTD)	[77]
	Raj 4014 × WH730 (HT) RIL population (113)	1000-grain weight (TGW)	[78]
	Aegilops tauschii Coss.	Cell membrane stability and TTC-based cell viability	[79]
	A. speltoides Tausch; *A. geniculata* Roth	Spikelet fertility	[80]
	ALTAR 84/AO'S'; ALTAR 84/*A. tauschii*	Leaf chlorophyll, grain weight, and grain yield	[81]
	Moomal-2000, Mehran-89	Germination-related traits	[82]
	Jimai-22	Photosynthesis, PS II, carboxylation, and grain yield	[83]
	CB # 367, 333, 335	Grain development and survival	[84]
	WH # 1021, 730	Grain yield	[85]
	SYN # 11, 36, 44	1000-grain weight	[86]
Rice	Dular, Todorokiwase, Milyang23, IR2006-P12-12-2-2, Giza178	Spikelet fertility and seed set	[87]
	N22, Bala, Co 39; CG14 (*O. glaberrima*)	Spikelet fertility and seed set	[88]
	N22, NH219	Spikelet fertility and pollen viability	[89]
	Bala (HT) × Azucena RIL population	Spikelet fertility	[90]
	—	Spikelet fertility	[91]
	Oryza meridionalis	Growth rate and photosynthesis	[92]
	N-22	Spikelet fertility	[93]
	Nipponbare, Akitakomachi	Spikelet fertility	[94]
Maize	ZPBL 1304 (HT); (ZPBL 1304 × ZPL 389) F2 population (160)	Heat shock protein (HSP)	[95]
	B76, Tx205, C273A, BR1, B105C, C32B, S1W, C2A554-4	Leaf firing and tassel blast	[96]
	—	Grain filling duration, kernel dry weight, starch, protein, and oil contents	[97]
	Hybrids: YH-1898, KJ.Surabhi, FH-793, ND-6339, NK-64017	Yield	[98]
Pearl millet	9444, Nandi 32, ICMB 05666, ICMB 92777; ICMB 02333	Seed set	[21]
	F1's: H77/29-2 × CVJ-2-5-3-1-3; H77/833-2 × 96 AC-93	Seedling thermotolerance index (STI), seed to seedling thermotolerance index (SSTI), and membrane thermostability (MTS)	[99]
	CVJ-2-5-3-1-3; 77/371 × BSECT CP 1	STI and SSTI	[100]

Crop	Heat tolerance sources*	HT associated trait/index	References
Sorghum	DeKalb 28E	Pollen viability, seed set, seed yield, and harvest index	[101]
Cowpea	California Blackeye 27 (CB27)	Flower production and pod set	[102]
	B89-200, TN88-63	Seed yield	[103]
Common bean	SRC-1-12-1-182; SRC-1-12-1-48; 98020-3-1-7-2; 98012-3-1-2-1	Heat tolerance index (HTI), and heat susceptibility index (HSI)	[104]
Chickpea	ICCV 92944; several genotypes	Seed set	[105, 106]

*Heat tolerance sources include germplasm, breeding lines, populations, varieties/hybrids, wild species, etc.

Table 2. Heat tolerance sources and associated traits/indices of important food crops.

region and south India [32, 110]. The genetic variability for HT in rice could be exploited to screen germplasm and select cultivars that open flowers earlier in the morning or that maintain a high number of spikelets per panicle in warm environments [111]. A positive correlation between canopy temperature depression (CTD) and membrane stability with grain yield have been noticed and recommended as useful traits in selecting high-temperature tolerant genotypes in wheat [80, 112].

Generally, there is a strong correlation between pollen production and viability, anther dehiscence, and seed set. The anthers of heat tolerance rice cultivars dehisce more easily than those of susceptible cultivars under high-temperature conditions [90, 93]. Higher pollen grain fertility under HS may serve as an important criterion for measuring HT [113]. Similarly, gametic selection has been proposed as a viable option for addressing HS in maize [71]. In wheat, maintaining grain weight under heat stress during grain filling is a measure of HT [114, 115]. Hence, it has been proposed that high grain-filling rate and high potential grain weight can be useful selection criteria for improving HT [116]. Stay-green character has also been suggested for mass screening of wheat genotypes for HT [117]. However, this trait may be disadvantageous as it is associated with the tendency to retain the stem reserves [118].

The sources identified for heat tolerance using suitable screening method have to be confirmed for their level of tolerance across different temperature regimes, and breeding approaches have to be outlined to incorporate the tolerance into desirable agronomic background. Transfer of tolerance from cultivated germplasm could be easy but the chances of finding the sources of tolerance are quite less. The crop wild relatives may have higher level of tolerance, but their incorporation into cultivated background needs a perfect prebreeding/backcross program. The availability of molecular markers in crop plants of economic importance could rather be put to use by forward and background selection for introgression of desirable genomic regions associated with heat tolerance.

5.2. Conventional breeding: a traditional approach

The conventional breeding efforts toward development of heat tolerance cultivars are comparatively less among different crops. However, the emphasis has been quite recent and some efforts are being made in few important food crops such as wheat, rice, maize, tomato, potato, etc. Heat escape is an alternative mechanism through which plant completes its life cycle

before the onset of heat stress. In durum wheat, this property has been utilized in development of early maturing genotypes such as Waha-1, Omrabi-5, and Massara-1 [119]. In rice, introgression breeding has facilitated the transfer of HT from "N22" to "Xieqingzao B" line by developing BC_1F_8 lines [120]. Additionally, the advanced line derived from Gayabyeo/N22 cross has offered HT as well as high yield [121]. In wheat, *Aegilops tauschii* was successfully used as a donor for incorporating HT-relevant component traits such as cell membrane stability and chlorophyll retention into cv. PBW550 through backcrossing [122]. More recently, attempts were made in wheat aiming at introgression of wheat-*Leymus racemosus* chromosome to cv. Chinese spring to enhance HT and better adaptation under heat stress [123]. Impressive accomplishments were achieved in harnessing the natural genetic variation for HT, and additional efforts are underway to introduce the heat tolerance QTLs/genes into different genetic backgrounds [88, 120]. In cotton, through pollen selection under HS, heat tolerance genes were transferred from a donor line "7456" (*G. barbadense* L.) to a heat-sensitive genotype "Paymaster 404" through backcrossing [113]. In sunflower, by using temperature induction response (TIR) technique, adequate genetic variability was observed for thermotolerance among the parental lines of the hybrid KBSH-1, viz. CMS234A, CMS234B, and 6D-1 [124]. The availability of potential donors for HT would encourage plant breeders not only to deploy these novel sources directly in breeding schemes but also to excavate the resilient alleles that underlie tolerance.

5.3. Physiological traits-based breeding

The efficiency of direct selection for yield improvement under stressed conditions is hindered by low heritability and a complex network of major and minor QTLs governing them [125, 126]. Breeding for high-yielding and heat tolerance lines is limited by the influence of environmental factors, poor understanding of genetic inheritance of HT, and less availability of validated QTLs/cloned gene(s) for HT in plants [127]. Physiological trait-based breeding would be an ideal strategy for incorporating gene(s)/QTLs that determine heat tolerance. Such an approach has been adopted in wheat at CIMMYT to develop heat tolerance varieties [128]. Toward breeding for heat tolerance, physiological traits that need to be considered include those related to canopy structure, delayed senescence, photosynthesis efficiency, less respiration rates, reproductive traits, and harvest index [127, 129]. Genetic variability has been assessed in several crop plants under HS for several physiological traits and suitable tolerant sources and associated traits have been identified (**Table 2**).

Genetic variability existing for the plant phenologies conferring HT need to be exploited. Alternatively, selection for morphophysiological traits involved in heat stress adaptation, and also indirectly associated with yield, can be utilized for enhancing HT in crop plants that has been explored in wheat [127]. Substantial genetic variability for photosynthetic rate under HS has been noted in wheat and rice, which would serve as a potential indicator of HS tolerance [130, 131]. While screening over 1000 wheat genotypes, the chlorophyll fluorescence was established as an important physiological parameter [132]. Canopy temperature depression (CTD) has been found to act as heat escape mechanism in cotton [133], while for HT in wheat [81, 134]. The cooler canopy temperature (CT) under HS caused higher yield in wheat [109, 135]. Under HS, the CTD, flag leaf stomatal conductance and photosynthetic rate together are found to be positively correlated with yield in wheat [62].

Membrane thermostability MTS is considered as a useful component for measuring HT, while assessing genetic variability in different crops [49]. Selection for MTS during anthesis stage under HS led to significant yield improvements in wheat [50]. Various physiological traits and their relative contributions to HT in wheat have been discussed in detail earlier [129]. Screening against HS based on parameters such as electrolyte leakage from cell membrane and chlorophyll fluorescence revealed negative association of membrane injury with specific leaf weight in some legume species including groundnut and soybean [59]. The combination of the two selection parameters, viz. high chlorophyll content and MTS, was implicated to carry out selections in Brassica and wheat [34, 136]. The relative cell injury level (RCIL) under HS could also be taken as a reliable index in determining HS tolerance in cotton [137].

The adaptation of root respiratory carbon metabolism can offer tolerance to soil temperature by managing the ion uptake load as reported in *Agrostis* species [138]. The efficient carbon and protein metabolism is known to confer higher thermotolerance to roots at 45°C in *Agrostis scabra* (a C_3 perennial grass species) [139]. In wheat, stay-green trait associated with CTD has been indicated as a strong indicator of HT [140]. However, stay-green trait is less important in the context of yield on account of disability in translocation of stem reserves to grain under HS [118]. Whereas, the conditions encompassing heat stress alone, as well as heat stress and drought, stay-green trait is measured as normalized difference vegetation index (NDVI) at physiological maturity exhibiting a positive correlation with the yield [141]. Therefore, physiological trait-based breeding remains a promising improvement strategy to develop heat tolerance genotypes without causing yield penalty.

5.4. Molecular markers in breeding for HT

The efforts by conventional breeding schemes have led to identification of several HT-relevant gene(s) and their inheritance patterns [6, 142]. Recent advances in marker discovery and genotyping assays have led to the precise determination of chromosomal position of the QTLs responsible for HT in different crops [90, 143–145]. The identification of markers linked to QTLs enables breeding of stress-tolerant crops by combining or pyramiding of QTLs governing tolerance to various stresses. An elaborative list of QTLs associated with HT in various crops along with details of mapping populations used, number of QTLs identified, associated markers, chromosomal positions, and phenotypic variation explained (PVE) has been summarized by Jha et al. [146]. Several major or minor QTLs and linked markers for HT have been identified in major food crops such as rice [90, 147–149], wheat [143, 144, 150–152], and maize [153, 154]. QTLs for several HT-related traits have been identified such as cellular membrane stability, pollen germination, and pollen tube growth in maize [153, 154]; stay-green trait, photosynthetic genes, and HSPs in sorghum [155]; and pollen viability in adzuki bean [156].

In wheat, one candidate SNP marker that clearly distinguished heat tolerance and heat-sensitive cultivars was identified [157]. For grain-filling rate (GFR) that governs grain yield in wheat under HS, 12 closely linked SSR markers were identified [158]. Using a SNP marker for a RIL population, five important genomic regions that offered HT in cowpea were identified [159]. Genome-wide as well as candidate-gene-based association mapping using SNP and DArT markers in chickpea could establish marker-trait associations for HT [160]. Recently, a major dominant

locus *OsHTAS* (*Oryza sativa* heat tolerance at seedling stage) was identified from the genotype HT54, which contributed high-temperature tolerance at 48°C especially during seedling and grain-filling stages [161]. The QTLs identified using molecular markers in different crops provide a way to transfer the causative heat tolerance gene(s)/QTLs to elite cultivars. In parallel, the fine mapping accompanied by cloning of candidate QTL will help the breeders to commence marker-assisted breeding for incorporating HT in various important crops in near future.

For transfer of quantitative traits such as HT, molecular markers would enable the recovery of desirable genotypes in a precise and time-saving fashion [162]. Molecular markers have been useful in identifying heat sensitive advanced generation introgression lines in rice [120, 149]. The near isogenic lines (NILs) created by introducing desirable allele into the heat sensitive cultivar showed considerable reduction in the incidence of heat-induced injuries such as white-back kernels [163]. Recently in rice, a 1.5-Mb chromosomal region harboring a robust QTL controlling better grain quality under HS has been transferred from "Kokoromachi" to "Tohoku 168" using marker-assisted backcrossing [162]. The resultant NILs had improved grain quality over the susceptible parent.

DNA markers related to various HT/component traits have been identified in different crops such as rice [148, 161, 163–165], wheat [143–145, 150, 152, 166, 167], and cowpea [159, 168]. Once the markers associated with QTLs have been identified, the candidate QTLs can further be introgressed into elite lines through marker-assisted selection (MAS) strategies. One of the difficulties of developing superior genotypes for heat stress is that these traits are generally controlled by small effect QTLs or several epistatic QTLs [3]. To overcome this problem, approaches that can be employed are pyramiding several QTLs in the same genetic background using large populations through marker-assisted recurrent selection (MARS) or genomic selection (GS) [169].

MAS programs for complex traits such as heat tolerance are not effective mainly due to the genotype × environment and gene-gene (i.e., epistasis) interactions, which frequently result in a low breeding efficiency [170]. In contrast to MAS strategies which use markers for which a significant association with a trait has been identified, the GS method predicts breeding values using data derived from a vast number of molecular markers with a high coverage of the genome. Its novelty is that it uses all marker data as predictors of performance and subsequently delivers more accurate predictions [3]. Simulation studies indicated that GS may increase the correlations between predicted and true breeding value over several generations, without the need to re-phenotype. Thus, GS may result in lower analysis costs and increased rates of genetic gain [171, 172].

QTLs often do not translate well across genetic backgrounds and often produce smaller than expected adaptation effects. Thus, improving crop abiotic stress tolerance by exploiting the segregation of natural alleles rather seems challenging for such an adaptive QTL strategy [170]. When quantitative hereditary characteristics such as heat stress tolerance are involved, recurrent selection seems to be one of the most efficient methods in plant breeding. In multiple crosses, the probability is very small of obtaining superior genotypes that reunite all the favorable alleles. However, in this circumstance, a large segregating population is required, aspect that becomes unfeasible in practice [3]. The alternative is to adopt recurrent selection to gradually accumulate, by recombination cycles, the desirable and available alleles in different

parents [173]. The main aim of a recurrent selection program is to increase the frequency of favorable alleles for traits of interest, while conserving the genetic variability. The major advantages are: (i) greater genetic variability obtained by intercrossing of multiple parents; (ii) greater opportunity for recombination through successive crossings; (iii) greater efficiency in increasing the favorable gene frequency since it is a repetitive and accumulative process; and (iv) greater facility to incorporate exotic germplasm in the population [174]. In potato breeding program at the Federal University of Lavras, heat stress tolerance genotypes were successfully developed using recurrent selection, which led to gains in tuber production with improved quality [175].

A current approach to the challenge of high-temperature tolerance is "physiological" or "developmental" trait breeding through recurrent selection using crop germplasm from regions with hot growing seasons [127]. In both cases, the targets are loci with high heritability that sustain yield at normal and elevated temperatures. The selection can be on combined heat and drought tolerance, as required by many crops. Recurrent selection has been successfully used to improve heat tolerance in wheat [176, 177] and potato [175]. In wheat, the stable introduction of chromosomes from its wild relative *Leymus racemosus* provided heat tolerance in hot and arid fields [123]. The heat- and drought-tolerant rice variety N22 (*aus* ssp.) has provided QTLs associated with high levels of HSPs in anthers, spikelets, and flag leaves associated with maintenance of yield under high night temperatures [148, 178]. Additional targeted developmental breeding in rice takes advantage of QTLs from *Oryza officinalis* that avoid heat-induced spikelet sterility by promoting dehiscence and fertilization in the cooler early mornings [179]. In time, these approaches may yield loci and knowledge that can accelerate improvement in heat tolerance in combination with drought tolerance.

Field trials under real stress conditions allow for conclusive remarks on stress tolerance and yield performance of a genotype. Development of more precise phenotyping tools that can be applied to field conditions is a prerequisite for enabling the assessment of the complex genetic networks associated with QTLs. The small, yet significant phenotypic changes delivered by introducing single genes into breeding material require precision phenotyping protocols and the resource capacity to carry these out on very large populations. The integrated approach for HT has been illustrated in **Figure 1** with details of whole breeding process.

5.5. Transgenic approach

Tolerant sources if not found in a given species or not giving enough protection, transgenic approach is an alternative option. However, it requires identification of the gene responsible for the desired trait, but poses no barrier for transferring useful genes across different species within the plant kingdom or even from animal systems. Genes of nonplant species could potentially be introgressed as well and generally, several combinations of beneficial genes could be transferred into the same plant [3]. With increasingly refined transformation and regeneration protocols, transgenic techniques are becoming attractive tool for designing both biotic and abiotic stress-tolerant crops via manipulating native genes or introducing gene(s) that lie beyond the crop gene pools [180]. Toward development of transgenics for HT, primarily the focus has been on engineering genes that encode transcription factors (TFs), HSPs, chaperones, organic osmolytes, antioxidants and plant growth regulators. This has been earlier summarized by Grover et al. [181]. Most of the transgenics developed for HT are mostly

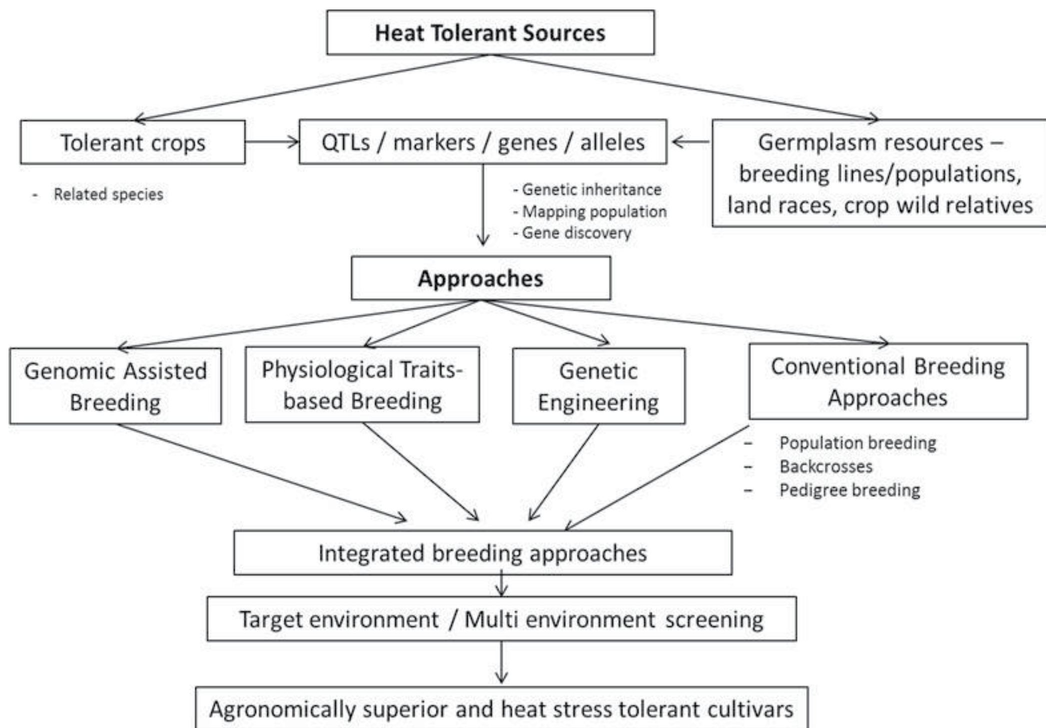


Figure 1. Schematic presentation of next-generation (integrated) breeding approach to develop heat tolerant cultivars.

in model plants such as *Arabidopsis*, tobacco, and rice. Further, it needs to be extended to other agriculturally important field crops.

To combine stress tolerance with high yield potential while avoiding the negative effects of a stress gene on plant growth under favorable conditions, strategies that spatially and temporally restrict transgene expression via tissue-specific and stress-inducible promoters are used [182]. Engineering promoters will facilitate gene pyramiding through genetic modification, addressing the issue of tolerance to multiple stresses at different stages of plant growth [183]. As an alternative, engineering with specific transcription factors and signaling components could be employed. Ultimately, this leads to the expression of their target transcriptome that consists of several genes involved in the response to stress. Transcriptome engineering emerges thus as a promising avenue for the development of abiotic stress-tolerant crops. Currently, however, plant genetic engineering is hampered by nonbiological constraints mainly related to the commercialization of transgenic crops, particularly in Europe [184]. Thus, the future commercial success of transgenic breeding will depend upon the development of clearly defined and scientifically based regulatory frameworks, and upon public acceptance of genetically modified plants and their produce [185].

5.6. Phenomics for precise and high-throughput phenotyping

Enormous sequence information is being generated through new-generation high-throughput DNA sequencing technologies [186]. But, precise, accurate, and high-throughput phenotyping of the traits on a large scale remains strenuous [187]. Field-based phenotyping (FBP)

facilities are now being initiated to have a more realistic evaluation of the plant responses to environment [188]. Infrared thermography is being utilized for large-scale phenotyping of plants responses to abiotic stress including HT in wheat and chrysanthemum [189–191]. The chlorophyll fluorescence (Fv/Fm) and canopy temperature were assessed in wheat to identify HT types [132, 143, 192, 193]. Recently, high-resolution thermal imaging system was used to precisely measure the leaf temperature [194]. The phenotyping platforms established to screen for HT include "HTpheno" for image analysis [195] and "Rootscope" used to quantify heat-shock responses in plants [196]. In the near future, the new-generation phenomics platforms would allow cost-effective and user-friendly screening for HT in crop plants.

6. Conclusion and way forward

High temperatures reduce global crop productivity by limiting either through growth limit and grain set or grain filling, and can also affect the end-use quality of the grain by reducing its compositions. Thus, temperatures during flowering time are very common in the photosensitive genotypes or in crops growing in the arid and semi-arid regions of the world, where extreme heat is more frequent with the climate change. Breeders should consider and devise tools for flowering time heat tolerance screening which direct link to productivity of a crop. Tolerance to heat is difficult to assess in the field due to variation in the timing and severity of natural heat events; so, rapid and cost-effective screening tools should be in place for applied crop breeding. All the growing food crops are selected upon short duration and better yield in small piece of land through domestication; therefore, there is a long way to go in understanding the mechanisms and its dissections before we draw a truly comprehensive picture of heat tolerance breeding for food crops. At the same time, these type of research is essential to counteracting a future where climate change may lead to moderate-to-severe reduction in crop yield in tropical and subtropical regions by the end of this century. Finally, farmers and breeders are blamed for these development-oriented research activities which are not always true as connecting lab evidences to farmers' field still is undermanaged in agricultural systems, especially in public not in private sector (champions of seed systems). A day or two scientific and political conferences will never bring long-term comprehensive solution for climate-smart agriculture but building a strong global research network consortium is highly need of the century that would be a potential seed of the future climate-smart agricultural strategies to capture crop diversity including native food and fodder crops such as millets and grain legumes.

Author details

Mahalingam Govindaraj[1], Santosh K. Pattanashetti[1]*, Nagesh Patne[2] and Anand A. Kanatti[1]

*Address all correspondence to: s.pattanashetti@cgiar.org

1 International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

2 International Maize and Wheat Improvement Center (CIMMYT), ICRISAT, Patancheru, India

References

[1] Mickelbart MV, Hasegawa PM, Bailey-Serres J. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nature Reviews Genetics. 2015;**16**:237-251

[2] IPCC. In: Field CB et al., editors. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge University Press; 2014. Available from: http://www.ipcc.ch/report/ar5/wg2/

[3] Bita CE, Gerats T. Plant tolerance to high temperature in a changing environment: Scientific fundamentals and production of heat stress-tolerant crops. Frontiers in Plant Science. 2013;**4**:273-1-18

[4] Ortiz R, Sayre KD, Govaerts B, Gupta R, Subbarao GV, Ban T, Hodson D, Dixon JM, Ortiz-Monasterio JI, Reynolds M. Climate change: Can wheat beat the heat? Agriculture, Ecosystems and Environment. 2008;**126**:46-58

[5] Raju BMK, Rao KV, Venkateswarlu B, Rao AVMS, Rama Rao CA, Rao VUM, Bapuji Rao B, Ravi Kumar N, Dhakar R, Swapna N, Latha P. Revisiting climatic classification in India: A district-level analysis. Current Science. 2013;**105**(4):492-495

[6] Wahid A, Gelani S, Ashraf M, Foolad M. Heat tolerance in plants: An overview. Environmental and Experimental Botany. 2007;**61**:199-223

[7] Peet MM, Willits DH. The effect of night temperature on greenhouse grown tomato yields in warm climate. Agricultural and Forest Meteorologyis. 1998;**92**:191-202

[8] Shinozaki K, Yamaguchi-Shinozaki K, editors. Molecular Responses to Cold, Drought, Heat and Salt Stress in Higher Plants. Austin, Texas: R.G. Landes Co.; 1999. pp. 81-98

[9] Ashraf M, Harris PJC, editors. Abiotic Stresses: Plant Resistance through Breeding and Molecular Approaches. New York: Howarth Press Inc.; 2005. pp. 277-300

[10] Seginer I. Transpirational cooling of a greenhouse crop with partial ground cover. Agricultural and Forest Meteorology. 1994;**71**:265-281

[11] Dias CV, Mendes JS, dos Santos AC, Pirovani CP, da Silva Gesteira A, Micheli F, Gramacho KP, Hammerstone J, Mazzafera P, de Mattos Cascardo JC. Hydrogen peroxide formation in cacao tissues infected by the hemibiotrophic fungus *Moniliophthora perniciosa*. Plant Physiology and Biochemistry. 2011;**49**:917-922

[12] Barnabás B, Jäger K, Fehér A. The effect of drought and heat stress on reproductive processes in cereals. Plant, Cell & Environment. 2008;**31**:11-38

[13] Sakata T, Higashitani A. Male sterility accompanied with abnormal anther development in plants–genes and environmental stresses with special reference to high temperature injury. International Journal of Plant Developmental Biology. 2008;**2**:42-51

[14] Vollenweider P, Günthardt-Goerg MS. Diagnosis of abiotic and biotic stress factors using the visible symptoms in foliage. Environmental Pollution. 2005;**137**:455-465

[15] Young LW, Wilen RW, Bonham-Smith PC. High temperature stress of *Brassica napus* during flowering reduces micro- and mega gametophyte fertility, induces fruit abortion, and disrupts seed production. Journal of Experimental Botany. 2004;**55**:485-495

[16] Zinn KE, Tunc-Ozdemir M, Harper JF. Temperature stress and plant sexual reproduction: Uncovering the weakest links. Journal of Experimental Botany. 2010;**61**:1959-1968

[17] Kaushal N, Bhandari K, Siddique KHM, Nayyar H. Food crops face rising temperatures: An overview of responses, adaptive mechanisms, and approaches to improve heat tolerance. Cogent Food & Agriculture. 2016;**2**:1134380

[18] Hedhly A, Hormaza JI, Herrero M. Global warming and sexual plant reproduction. Trends in Plant Science. 2009;**14**:30-36

[19] Nava GA, Dalmago GA, Bergamaschi H, Paniz R, dosSantos RP, Marodin GAB. Effect of high temperatures in the pre-blooming and blooming periods on ovule formation, pollen grains and yield of 'Granada' peach. Scientia Horticulturae. 2009;**122**:37-44

[20] Hedhly A. Sensitivity of flowering plant gametophytes to temperature fluctuations. Environmental and Experimental Botany. 2011;**74**:9-16

[21] Gupta SK, Rai KN, Singh P, Ameta VL, Gupta SK, Jayalekha AK, Mahala RS, Pareek S, Swami ML, Verma YS. Seed set variability under high temperatures during flowering period in pearl millet (*Pennisetum glaucum* L. (R.) Br.). Field Crops Research. 2015;**171**(1):41-53

[22] Porter J, Gawith M. Temperatures and the growth and development of wheat: A review. European Journal of Agronomy. 1999;**10**:23-36

[23] Sánchez B, Rasmussen A, Porter JR. Temperatures and the growth and development of maize and rice: A review. Global Change Biology. 2013;**20**:408-417

[24] Wassmann R, Jagadish S, Sumfleth K, Pathak H, Howell G, Ismail A, Serraj R, Redona E, Singh RK, Heuer S. Regional vulnerability of climate change impacts on Asian rice production and scope for adaptation. Advances in Agronomy. 2009;**102**:91-133

[25] Akman Z. Comparison of high temperature tolerance in maize, rice and sorghum seeds by plant growth regulators. Journal of Animal and Veterinary Advances. 2009;**8**:358-361

[26] Ren C, Bilyeu KD, Beuselinck P. Composition, vigor, and proteome of mature soybean seeds developed under high temperature. Crop Science. 2009;**49**:1010-1022

[27] Porter JR, Semenov MA. Crop responses to climatic variation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2005;**360**:2021-2035

[28] Banowetz G, Ammar K, Chen D. Temperature effects on cytokinin accumulation and kernel mass in a dwarf wheat. Annals of Botany. 1999;**83**:303-307

[29] Mahmood S, Wahid A, Javed F, Basra SMA. Heat stress effects on forage quality characteristics of maize (*Zea mays*) cultivars. International Journal of Agriculture and Biology. 2010;**12**:701-706

[30] Wahid A. Physiological implications of metabolite biosynthesis for net assimilation and heat-stress tolerance of sugarcane (*Saccharum officinarum*) sprouts. Journal of Plant Research. 2007;**120**:219-228

[31] Lehman VG, Engelke MC. Heritability of creeping bentgrass shoot water content under soil dehydration and elevated temperatures. Crop Science. 1993;**33**:1061-1066

[32] Yadav SS, Redden R, Chen W, Sharma B, editors. Chickpea Breeding and Management. Wallingford, UK: CAB International; 2007. pp. 474-496

[33] Yang J, Sears RG, Gill BS, Paulsen GM. Genotypic differences in utilization of assimilate sources during maturation of wheat under chronic heat and heat shock stresses. Euphytica. 2002;**125**:179-188

[34] Ristic Z, Bukovnik U, Prasad PVV. Correlation between heat stability of thylakoid membranes and loss of chlorophyll in winter wheat under heat stress. Crop Science. 2007;**47**:2067-2073

[35] Allakhverdiev SI, Kreslavski VD, Klimov VV, Los DA, Carpentier R, Mohanty P. Heat stress: An overview of molecular responses in photosynthesis. Photosynthesis Research. 2008;**98**:541-550

[36] Xu S, Li J, Zhang X, Wei H, Cui L. Effects of heat acclimation pretreatment on changes of membrane lipid peroxidation, antioxidant metabolites, and ultrastructure of chloroplasts in two cool-season turfgrass species under heat stress. Environmental and Experimental Botany. 2006;**56**:274-285

[37] Wang W, Vinocur B, Shoseyov O, Altman A. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in Plant Science. 2004;**9**:244-252

[38] Kaushal N, Awasthi R, Gupta K, Gaur P, Siddique KHM, Nayyar H. Heat-stress-induced reproductive failures in chickpea (*Cicer arietinum*) are associated with impaired sucrose metabolism in leaves and anthers. Functional Plant Biology. 2013;**40**:1334-1349

[39] Hemantaranjan A, Nishant Bhanu A, Singh MN, Yadav DK, Patel PK. Heat stress responses and thermotolerance. Advances in Plants & Agriculture Research 2014;**1**:1-10

[40] Larkindale J, Knight MR. Protection against heat stress induced oxidative damage in *Arabidopsis* involves calcium, abscissic acid, ethylene and salicylic acid. Plant Physiology. 2002;**128**:682-695

[41] Suzuki N, Mittler R. Reactive oxygen species and temperature stresses: A delicate balance between signaling and destruction. Physiologia Plantarum. 2006;**126**:45-51

[42] Hameed A, Goher M, Iqbal N. Heat stress-induced cell death, changes in antioxidants, lipid peroxidation and protease activity in wheat leaves. Journal of Plant Growth Regulation. 2012;**31**:283-291

[43] Chakraborty U, Pradhan D. High temperature-induced oxidative stress in *Lens culinaris*, role of antioxidants and amelioration of stress by chemical pre-treatments. Journal of Plant International. 2011;**6**:43-52

[44] Kumar MS, Kumar G, Srikanthbabu V, Udayakumar M. Assessment of variability in acquired thermotolerance: Potential option to study genotypic response and the relevance of stress genes. Journal of Plant Physiology. 2007;**164**:111-125

[45] Maestri E, Klueva N, Perrotta C, Gulli M, Nguyen HT, Marmiroli N. Molecular genetics of heat tolerance and heat shock proteins in cereals. Plant Molecular Biology. 2002;**48**:667-681

[46] Howarth CJ, Pollock CJ, Peacock JM. Development of laboratory based methods for assessing seedling thermotolerance in pearl millet. New Phytologist. 1997;**137**:129-139

[47] Yadav SS, Redden R, Hatfield JL, Lotze-Campen H, Hall AJW, editors. Crop Adaptation to Climate Change. Hoboken: JohnWiley & Sons; 2011. DOI: 10.1002/9780470960929.ch24

[48] Fritsche-Neto R, Borém A, editors. Plant Breeding for Abiotic Stress Tolerance; Berlin: Springer. 2012. pp. 137-156

[49] Blum A, Ebercon A. Cell membrane stability as a measure of drought and heat tolerance in wheat. Crop Science. 1981;**21**:43-47

[50] Shanahan JF, Edwards IB, Quick JS, Fenwick JR. Membrane thermostability and heat tolerance of spring wheat. Crop Science. 1990;**30**:247-251

[51] Martineau JR, Specht JE, Williams JH, Sullivan CY. Temperature tolerance in soybeans. I. Evaluation of a technique for assessing cellular membrane thermostability. Crop Science. 1979;**19**:75-78

[52] Ristic Z, Williams G, Yang G, Martin B, Fullerton S. Dehydration, damage to cellular membranes, and heatshock proteins in maize hybrids from different climates. Journal of Plant Physiology. 1996;**149**:424-432

[53] Tongden C, Basant M, Chakraborty U. Screening of thermotolerant cultivars of chickpea using cell membrane stability test and biochemical markers. Journal of Hill Research. 2006;**19**:52-58

[54] Xu RQ, Sun QX, Zhang SZ. Screening methods and indices of heat tolerance in spring wheat. Acta Agriculturae Boreali Sinica. 1997;**12**:22-29

[55] Ibrahim AMH, Quick JS. Genetic control of high temperature tolerance in wheat as measured by membrane thermal stability. Crop Science. 2001;**41**(5):1405-1407

[56] Schreiber U, Berry JA. Heat-induced changes of chlorophyll fluorescence in intact leaves correlated with damage of the photosynthetic apparatus. Planta. 1977;**136**:233-238

[57] Burke JJ. Variation among species in the temperature dependence of the reappearance of variable fluorescence following illumination. Plant Physiology. 1990;**93**:652-656

[58] Al-Khatib K, Paulsen GM. Mode of high temperature injury to wheat during grain development. Physiologia Plantarum. 1984;**61**:363-368

[59] Srinivasan A, Takeda H, Senboku T. Heat tolerance in food legumes as evaluated by cell membrane thermo stability and chlorophyll fluorescence techniques. Euphytica. 1996;**88**: 35-45

[60] Zhang M, Barg R, Yin M, Gueta-Dahan Y, Leikin-Frenkel A, Salts Y, Shabtai S, Ben-Hayyim G. Modulated fatty acid desaturation via overexpression of two distinct omega-3 desaturases differentially alters tolerance to various abiotic stresses in transgenic tobacco cells and plants. The Plant Journal. 2005;**44**:361-371

[61] Larkindale J, Huang B. Thermotolerance and antioxidant systems in *Agrostis stolonifera*: Involvement of salicylic acid, abscisic acid, calcium, hydrogen peroxide, and ethylene. Journal of Plant Physiology. 2004;**161**:405-413

[62] Reynolds MP, Balota M, Delgado MIB, Amani I, Fischer RA. Physiological and morphological traits associated with spring wheat yield under hot, irrigated conditions. Australian Journal of Plant Physiology. 1994;**21**:717-730

[63] Ahmed FE, Hall AE, DeManson DA. Heat injury during floral development in cowpea (*Vigna unguiculata*, fabaceae). American Journal of Botany. 1992;**79**(7):784-791

[64] Dupuis I, Dumas C. Influence of temperature stress on *in vitro* fertilization and heat shock protein synthesis in maize (*Zea mays* L.) reproductive tissues. Plant Physiology. 1990;**94**:665-670

[65] Weerakoon WMW, Maruyama A, Ohba K. Impact of humidity on temperature-induced grain sterility in rice (*Oryza sativa* L). Journal of Agronomy and Crop Science. 2008;**194**:135-140

[66] Devasirvatham V, Gaur P, Mallikarjuna N, Raju TN, Trethowan RM, Tan DKY. Effect of high temperature on the reproductive development of chickpea genotypes under controlled environments. Functional Plant Biology. 2012;**39**:1009-1018

[67] Zaidi PH, Zaman-Allah M, Trachsel S, Seetharam K, Cairns JE, Vinayan MT. Phenotyping for Abiotic Stress Tolerance in Maize – Heat Stress. A Field Manual. Hyderabad, India: CIMMYT; 2016

[68] Reynolds MP, Nagarajan S, Razzaque MA, Ageeb OAA. Using Canopy Temperature Depression to Select for Yield Potential of Wheat in Heat-Stressed Environments. Wheat Special Report No. 42. Mexico, D.F.: CIMMYT; 1997

[69] Dreccer FM, Fainges J, Sadras V. Association between wheat yield and temperature in South-Eastern Australia. In: Proceedings of the 17th ASA Conference on Building Productive, Diverse and Sustainable Landscapes; 20-24 Sep 2015. Australia: Hobart; 2015. Available from: http://www.agronomy2015.com.au/papers/agronomy2015final00223.pdf

[70] Salem MA, VG Kakani SK, Reddy KR. Pollen-based screening of soybean genotypes for high temperatures. Crop Science. 2007;**47**:219-231

[71] Ottaviano E, Mulcahy DL, Sari Gorla M, Bergamini Mulcahy G, editors. Angiosperm Pollen and Ovules. New York: Springer-Verlag; 1992. pp. 355-358

[72] Devasirvatham V, Gaur P, Mallikarjuna N, Raju TN, Trethowan RM, Tan DKY. Reproduc tive biology of chickpea response to heat stress in the field is associated with the performance in controlled environments. Field Crops Research. 2013;**142**:9-19

[73] Peacock JM, Soman P, Jayachandran R, Rani AV, Howarth CJ, Thomas A. Effect of high soil surface temperature on seedling survival in pearl millet. Experimental Agriculture. 1993;**29**:215-225

[74] Yadav AK, Narwal MS, Arya RK. Genetic dissection of temperature tolerance in pearl millet (*Pennisetum glaucum*). Indian Journal of Agricultural Sciences. 2011;**81**:203-213

[75] Setimela PS, Andrews DJ, Partridge J, Eskridge KM. Screening sorghum seedlings for heat tolerance using a laboratory method. European Journal of Agronomy. 2005;**23**(2):103-107

[76] Ashraf M, Hafeez M. Thermotolerance of pearl millet and maize at early growth stages: Growth and nutrient relations. Biologia Plantarum. 2004;**48**:81-86

[77] Hede AR, Skovmand B, Reynolds MP, Crossa J, Vilhelmsen AL, Stølen O. Evaluating genetic diversity for heat tolerance traits in Mexican wheat landraces. Genetic Resources and Crop Evolution. 1999;**46**(1):37-45

[78] Pandey GC, Sareen S, Siwach P, Tiwari R. Molecular characterization of heat tolerance in bread wheat (*Triticum aestivum* L.) using differences in thousand-grain weights (dTGW) as a potential indirect selection criterion. Cereal Research Communications. 2014;**42**(1):38-46

[79] Gupta S, Kaur S, Sehgal S, Sharma A, Chhuneja P, Bains NS. Genotypic variation for cellular thermotolerance in *Aegilops tauschii* Coss., the D genome progenitor of wheat. Euphytica. 2010;**175**:373-381

[80] Pradhan GP, Prasad PVV, Fritz AK, Kirkham MB, Gill BS. High temperature tolerance in *Aegilops* species and its potential transfer to wheat. Crop Science. 2012;**52**:292-304

[81] Pradhan GP, PVV Prasad, Fritz AK, Kirkham MB, Gill BS. Effects of drought and high temperature stress on synthetic hexaploid wheat. Functional Plant Biology. 2012;**39**:190-198

[82] Buriro M, Oad FC, Keerio MI, Tunio S, Gandahi AW, Hassan SWU, Oad SM. Wheat seed germination under the influence of temperature regimes. Sarhad Journal of Agriculture. 2011;**27**:539-543

[83] Feng B, Liu P, Li G, Dong ST, Wang FH, Kong LA, Zhan JW. Effect of heat stress on the photosynthetic characteristics in flag leaves at the grain-filling stage of different heat-resistant winter wheat varieties. Journal of Agronomy and Crop Science. 2014;**200**:143-155

[84] Rehman AIH, Ahmad N, Hussain M, Khan MA, Farooq J, Ali MA. Screening wheat germplasm for heat tolerance at terminal growth stage. Journal Plant Omics. 2009;**2**:9-19

[85] Dhanda SS, Munjal R. Heat tolerance in relation to acquired thermotolerance for membrane lipids in bread wheat. Field Crops Research. 2012;**135**:30-37

[86] Sharma DK, Andersen SB, Ottosen CO, Rosenqvist E. Wheat cultivars selected for high Fv/Fm under heat stress maintain high photosynthesis, total chlorophyll, stomatal conductance, transpiration and dry matter. Plant Physiology. 2015;**153**(2):284-298

[87] Tenorio FA, Ye C, Redoña E, Sierra S, Laza M, Argayoso MA. Screening rice genetic resources for heat tolerance. SABRAO Journal of Breeding and Genetics. 2013;**45**(3):371-381

[88] Jagadish SVK, Craufurd PQ, Wheeler TR. Phenotyping parents of mapping populations of rice for heat tolerance during anthesis. Crop Science. 2008;**48**:1140-1146

[89] Poli Y, Basava RK, Panigrahy M, Vinukonda VP, Dokula NR, Voleti SR, Desiraju S, Neelamraju S. Characterization of a Nagina22 rice mutant for heat tolerance and mapping of yield traits. Rice. 2013;**6**:36

[90] Jagadish SVK, Cairns J, Lafitte R, Wheeler TR, Price AH, Craufurd PQ. Genetic analysis of heat tolerance at anthesis in rice. Crop Science. 2010;**50**:1633-1641

[91] Yan C, Ding Y, Liu Z, Wang Q, Li G, He Y, Wang S. Temperature difference between the air and organs of rice plant and its relation to spikelet fertility. Agricultural Sciences in China. 2008;**7**(6):678-685

[92] Scafaro AP, Haynes PA, Atwell BJ. Physiological and molecular changes in *Oryza meridionalis* Ng, a heat-tolerant species of wild rice. Journal of Experimental Botany. 2010;**61**:191-202

[93] Prasad PVV, Boote KJ, Allen LH, Sheehy JE, Thomas JMG. Species, ecotype and cultivar differences in spikelet fertility and harvest index of rice in response to high temperature stress. Field Crops Research. 2006;**95**(2):398-411

[94] Matsui T, Omasa K. Rice (*Oryza sativa* L.) cultivars tolerant to high temperature at flowering: Anther characteristic. Annals of Botany. 2002;**89**:683-687

[95] Ristic Z, Yang G, Martin B, Fullerton S. Evidence of association between specific heat-shock protein(s) and the drought and heat tolerance phenotype in maize. Journal of Plant Physiology. 1998;**153**(3):497-505

[96] Chen J, Xu W, Velten J, Xin Z, Stout J. Characterization of maize inbred lines for drought and heat tolerance. Journal of Soil and Water Conservation. 2012;**67**(5):354-364

[97] Wilhelm EP, Mullen RE, Keeling PL, Singletary GW. Heat stress during grain filling in maize effects on kernel growth and metabolism. Crop Science. 1999;**39**(6):1733-1741

[98] Rahman SU, Arif M, Hussain K, Hussain S, Mukhtar T, Razaq A, Iqbal RA. Evaluation of maize hybrids for tolerance to high temperature stress in Central Punjab. American journal of Bioenergy and Biotechnology. 2013;**1**:30-36

[99] Yadav AK, Arya RK, Narwal MS. Screening of pearl millet F1 hybrids for heat tolerance at early seedling stage. Advances in Agriculture. 2014; Article ID 231301, 1-17. DOI: 10.1155/2014/231301

[100] Yadav SS, Redden R, Hatfield JL, Lotze-Campen H, Hall AJW, editors. Crop Adaptation to Climate Change. Hoboken: JohnWiley & Sons. 2011. pp. 340-355

[101] Prasad PVV, Boote Jr KJ, Allen LH. Adverse high temperature effects on pollen viability, seed-set, seed yield and harvest index of grain-sorghum [*Sorghum bicolor* (L.) Moench] are more severe at elevated carbon dioxide due to higher tissue temperatures. Agricultural and Forest Meteorology. 2006;**139**:237-251

[102] Ehlers JD, Hall AE, Patel PN, Roberts PA, Matthews WC. Registration of California Blackeye 27. Crop Science. 2000;**40**:854-855

[103] Ehlers JD, Hall AE. Heat tolerance of contrasting cowpea lines in short and long days. Field Crops Research. 1998;**55**:11-21

[104] Porch TG. Application of stress indices for heat tolerance screening of common bean. Journal of Agronomy and Crop Science. 2006;**192**:390-394

[105] Gaur PM, Jukanti AK, Varshney RK. Impact of genomic technologies on chickpea breeding strategies. Agronomy. 2012;**2**:199-221

[106] Jha UC, Chaturvedi SK, Bohra A, Basu PS, Khan MS, Barh D. Abiotic stresses, constraints and improvement strategies in chickpea. Plant Breeding. 2014;**133**:163-178

[107] Tewolde H, Fernandez CJ, Erickson CA. Wheat cultivars adapted to post-heading high temperature stress. Journal of Agronomy and Crop Science. 2006;**192**:111-120

[108] Joshi AK, Mishra B, Chatrath R, Ferrara GO, Singh RP. Wheat improvement in India: Present status, emerging challenges and future prospects. Euphytica. 2007;**157**:431-446

[109] Mondal S, Singh RP, Crossa J, Huerta-Espino J, Sharma I, Chatrath R, Singh GP, Sohu VS, Mavi GS, Sukaru VPS, Kalappanavar IK, Mishra VK, Hussain M, Gautam NR, Uddin J, Barma NCD, Hakim A, Joshi AK. Earliness in wheat: A key to adaptation under terminal and continual high temperature stress in South Asia. Field Crops Research. 2013;**151**:19-26

[110] Berger JD, Milroy SP, Turner NC, Siddique KHM, Imtiaz M, Malhotra R. Chickpea evolution has selected for contrasting phenological mechanisms among different habitats. Euphytica. 2011;**180**:1-15

[111] Gonzalez Fontes A, editor. Abiotic Stress Adaptation in Plants; Houston: Studium Press Llc. 2010. pp. 387-415

[112] Blum A, Klueva N, Nguyen HT. Wheat cellular thermotolerance is related to yield under heat stress. Euphytica. 2001;**117**:117-123

[113] Rodriguez-Garay B, Barrow JR. Pollen selection for heat tolerance in cotton. Crop Science. 1988;**28**:857-859

[114] Tyagi PK, Pannu RK, Sharma KD, Chaudhary BD, Singh DP. Response of different wheat (*Triticum aestivu*m L.) cultivars to terminal heat stress. Tests of Agrochemicals and Cultivars. 2003;**24**:20-21

[115] Singha P, Bhowmick J, Chaudhury BK. Effect of temperature on yield and yield components of fourteen wheat (*Triticum aestivum* L.) genotypes. Environment and Ecology. 2006;**24**:550-554

[116] Dias AS, Lidon FC. Evaluation of grain filling rate and duration in bread and durum wheat, under heat stress after anthesis. Journal of Agronomy and Crop Science. 2009;**195**:137-147

[117] Reynolds MP, Ortiz-Monasterio JI, McNab A, editors. Application of Physiology in Wheat Breeding. Mexico, DF: CIMMYT; 2001. pp. 124-135

[118] Blum A. Improving wheat grain filling under stress by stem reserve mobilization. Euphytica. 1998;**100**:77-83

[119] Al-Karaki GN. Phenological development-yield relationships in durum wheat cultivars under late-season high-temperature stress in a semiarid environment. ISRN Agronomy. 2012;Article ID 456856, 1-7

[120] Jiang-lin L, Hong-yu Z, Xue-lian S, Ping-an Z, Ying-jin H. Identification for heat tolerance in backcross recombinant lines and screening of backcross introgression lines with heat tolerance at milky stage in rice. Rice Science. 2011;**18**:279-286

[121] Manigbas NL, Lambio LAF, Madrid LB, Cardenas CC. Germplasm innovation of heat tolerance in rice for irrigated lowland conditions in the Philippines. Rice Science. 2014;**21**:162-169

[122] Sehgal SK, Kaur S, Gupta S, Sharma A, Kaur A, Bains NS. A direct hybridization approach to gene transfer from *Aegilops tauschii* Coss. to *Triticum aestivum* L. Plant Breeding. 2011;**130**:98-100

[123] Mohammed YSA, Tahir ISA, Kamal NM, Eltayeb AE, Ali AM, Kamal NM. Impact of wheat-*Leymus racemosus* added chromosomes on wheat adaptation and tolerance to heat stress. Breeding Science. 2014;**63**:450-460

[124] Senthil Kumar M, Srikanthbabu V, Raju BM, Ganeshkumar SN, Udayakumar M. Screening of inbred lines to develop a thermo tolerant sunflower hybrid using the temperature induction response (TIR) technique: A novel approach by exploiting residual variability. Journal of Experimental Botany. 2003;**54**:2569-2578

[125] Leung H. Stressed genomics - bringing relief to rice fields. Current Opinion in Plant Biology. 2008;**11**:201-208

[126] Manavalan LP, Guttikonda SK, Tran LS, Nguyen HT. Physiological and molecular approaches to improve drought resistance in soybean. Plant & Cell Physiology. 2009;**50**:1260-1276

[127] Cossani CM, Reynolds MP. Physiological traits for improving heat tolerance in wheat. Plant Physiology. 2012;**160**:1710-1718

[128] Reynolds M, Manes Y, Izanloo A, Langridge P. Phenotyping approaches for physiological breeding and gene discovery in wheat. The Annals of Applied Biology. 2009;**155**:309-320

[129] Gupta PK, Balyan HS, Gahlaut V, Kulwal PL. Phenotyping, genetic dissection, and breeding for drought and heat tolerance in common wheat: Status and prospects. Plant Breeding Reviews. 2012;**36**:85-168. DOI: 10.1002/9781118358566.ch2

[130] Reynolds MP, Delgado MI, Gutierrez-Rodriguez M, Larqué-Saavedra A. Photosynthesis of wheat in warm, irrigated environment. I: Genetic diversity and crop productivity. Field Crops Research. 2000;**66**:35-50

[131] Restrepo-Diaz H, Garces-Varon G. Response of rice plants to heat stress during initiation of panicle primordia or grain-filling phases. Journal of Stress Physiology & Biochemistry. 2013;**9**:318-325

[132] Sharma DK, Andersen SB, Ottosen CO, Rosenqvist E. Phenotyping of wheat cultivars for heat tolerance using chlorophyll a fluorescence. Functional Plant Biology. 2012;**39**:936-947

[133] Cornish K, Radin JW, Turcotte EL, Lu ZM, Zeiger E. Enhanced photosynthesis and gs of pima cotton (*Gossypium barbadense* L.) bred for increased yield. Plant Physiology. 1991;**97**:484-489

[134] Balota M, Payne WA, Evett SR, Peters TR. Morphological and physiological traits associated with canopy temperature depression in three closely related wheat lines. Crop Science. 2008;**48**:1897-1910

[135] Kumari M, Pudake RN, Singh VP, Joshi A. Association of stay green trait with canopy temperature depression and yield traits under terminal heat stress in wheat (*Triticum aestivum* L.). Euphytica. 2012;**190**:87-97

[136] Kumar S, Sairam RK, Prabhu KV. Physiological traits for high temperature stress tolerance in *Brassica juncea*. Indian Journal of Plant Physiology. 2013;**18**:89-93

[137] Khan AI, Khan IA, Sadaqat HA. Heat tolerance is visible in cotton (*Gossypium hirsutum* L.) and can be exploited for breeding of better yielding cultivars under high temperature. Pakistan Journal of Botany. 2008;**40**:2053-2058

[138] Rachmilevitch S, Lambers H, Huang B. Root respiratory characteristics associated with plant adaptation to high soil temperature for geothermal and turf-type *Agrostis* species. Journal of Experimental Botany. 2006;**57**:623-631

[139] Huang B, Rachmilevitch S, Xu J. Root carbon and protein metabolism associated with heat tolerance. Journal of Experimental Botany. 2012;**63**:3455-3465

[140] Buck HT, Nisi JE, Salomón N, editors. Wheat Production in Stressed Environments. Developments in Plant Breeding. 2007;**12**:357-363

[141] Lopes MS, Reynolds MP. Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. Journal of Experimental Botany. 2012;**63**:3789-3798

[142] Farooq M, Bramley H, Palta JA, Siddique KHM. Heat stress in wheat during reproductive and grain filling phases. Critical Reviews in Plant Sciences. 2011;**30**:1-17

[143] Pinto RS, Reynolds MP, Mathews KL, McIntyre CL, Olivares-Villegas JJ, Chapman SC. Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. Theoretical and Applied Genetics. 2010;**121**:1001-1021

[144] Paliwal R, Röder MS, Kumar U, Srivastava JP, Joshi AK. QTL mapping of terminal heat tolerance in hexaploid wheat (*T. aestivum* L). Theoretical and Applied Genetics. 2012;**125**:561-575

[145] Bonneau J, Taylor J, Parent B, Bennett D, Reynolds M, Feuillet C, Langridge P, Mather D. Multi-environment analysis and improved mapping of a yield-related QTL on chromosome 3B of wheat. Theoretical and Applied Genetics. 2013;**126**:747-761

[146] Jha UC, Bohra A, Singh NP. Heat stress in crop plants: Its nature, impacts and integrated breeding strategies to improve heat tolerance. Plant Breeding. 2014;**133**:679-701

[147] Zhu CL, YH Xiao, Wang CM, Jiang L, Zhai HQ, Wan JM. Mapping QTL for heat-tolerance at grain filling stage in rice. Rice Science. 2005;**12**:33-38

[148] Ye C, Argayoso MA, Redona ED, SN Sierra, MA Laza, Dilla CJ, Mo Y, Thomson MJ, Chin J, Delavina CB, Diaz GQ, Hernandez JE. Mapping QTL for heat tolerance at flowering stage in rice using SNP markers. Plant Breeding. 2012;**131**:33-41

[149] Lei D, Tan L, Liu F, Chen L, Sun C. Identification of heat-sensitive QTL derived from common wild rice (*Oryza rufipogon* Griff). Plant Science. 2013;**201-202**:121-127

[150] Vijayalakshmi K, Fritz AK, Paulsen GM, Bai G, Pandravada S, Gill BS. Modeling and mapping QTL for senescence related traits in winter wheat under high temperature. Molecular Breeding. 2010;**26**:163-175

[151] Mason RE, Mondal S, Beecher F, Hays D. Genetic loci linking improved heat tolerance in wheat (*Triticum aestivum* L) to lower leaf and spike temperatures under controlled conditions. Euphytica. 2011;**180**:181-194

[152] Beecher FW, Mason E, Mondal S, Awika J, Hays D, Ibrahim A. Identification of quantitative trait loci (QTLs) associated with maintenance of wheat (*Triticum aestivum* Desf) quality characteristics under heat stress conditions. Euphytica. 2012;**188**:361-368

[153] Ottaviano E, Sari-Gorla M, Pé E, Frova C. Molecular markers (RFLPs and HSPs) for the genetic dissection of thermo tolerance in maize. Theoretical and Applied Genetics. 1991;**81**:713-719

[154] Frova C, Sari-Gorla M. Quantitative trait loci (QTLs) for pollen thermo tolerance detected in maize. Molecular & General Genetics. 1994;**245**:424-430

[155] Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT. Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome. 2000;**43**:461-469

[156] Singh RJ, Jauhar PP, editor. Genetic Resources, Chromosome Engineering and Crop Improvement. Vol. 1. Grain Legumes. Boca Roca, FL: CRC Press; 2005. pp. 347-359

[157] Garg D, Sareen S, Dalal S, Tiwari R, Singh R. Heat shock protein based SNP marker for terminal heat stress in wheat (*Triticum aestivum* L.). Australian Journal of Crop Science. 2012;**6**:1516-1521

[158] Barakat MN, Al-Doss AA, Elshafei AA, Moustafa KA. Bulked segregant analysis to detect quantitative trait loci (QTL) related to heat tolerance at grain filling rate in wheat using simple sequence repeat (SSR) markers. African Journal of Biotechnology. 2012;**11**:12436-12442

[159] Lucas MR, Ehlers JD, Huynh BL, Diop NN, Roberts PA, Close TJ. Markers for breeding heat-tolerant cowpea. Molecular Breeding. 2013;**31**:529-536

[160] Thudi M, Upadhyaya HD, Rathore A, Gaur PM, Krishnamurthy L, Roorkiwal M, Nayak SN, Chaturvedi SK, Basu PS, Gangarao NV, Fikre A, Kimurto P, Sharma PC, Sheshashayee MS, Tobita S, Kashiwagi J, Ito O, Killian A, Varshney RK. Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. PLoS One. 2014;**9**:e96758

[161] Wei H, Liu J, Wang Y, Huang N, Zhang X, Wang L, Zhang J, Tu J, Zhong X. A dominant major locus in chromosome 9 of rice (*Oryza sativa* L) confers tolerance to 48°C high temperature at seedling stage. The Journal of Heredity. 2013;**104**:287-294

[162] Shirasawa K, Sekii T, Ogihara Y, Yamada T, Shirasawa S, Kishitani S, Sasaki K, Nishimura M, Nagano K, Nishio T. Identification of the chromosomal region responsible for

high-temperature stress tolerance during the grain-filling period in rice. Molecular Breeding. 2013;**32**:223-232

[163] Kobayashi A, Sonoda J, Sugimoto K, Kondo M, Iwasawa N, Hayashi T, Tomita K, Yano M, Shimizu T. Detection and verification of QTLs associated with heat-induced quality decline of rice (*Oryza sativa* L.) using recombinant inbred lines and near-isogenic lines. Breeding Science. 2013;**63**:339-346

[164] Xiao Y, Yi P, Luo L, Zhang G, Deng H, Dia L, Liu X, Tang W, Chen L, Wang GL. Quantitative trait loci associated with seed set under high temperature stress at the flowering stage in rice (*Oryza sativa* L.). Euphytica. 2011;**178**:331-338

[165] Buu BC, Ha PTT, Tam BP, Nhien TTT, Hieu NV, Phuoc NT. Quantitative trait loci associated with heat tolerance in rice (*Oryza sativa* L.). Plant Breeding and Biotechnology. 2014;**2**:14-24

[166] Bennett D, Reynolds M, Mullan D, Izanloo A, Langridge P, Schnurbusch T. Detection of two major grain yield QTL in bread wheat (*Triticum aestivum* L.) under heat, drought and high yield potential environments. Theoretical and Applied Genetics. 2012;**125**:1473-1485

[167] Sadat S, Said KA, Bihamta MR, Torabi S, Salekdeh SGH, Ayeneh GAL. Marker assisted selection for heat tolerance in bread wheat. World Applied Sciences Journal. 2013;**21**:1181-1189

[168] Pottorff M, Roberts PA, Close TJ, Lonardi S, Wanamaker S, Ehlers JD. Identification of candidate genes and molecular markers for heat-induced brown discoloration of seed coats in cowpea [*Vigna unguiculata* (L.) Walp]. BMC Genomics. 2014;**15**:328

[169] Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. Science. 2010;**327**:818-822

[170] Collins NC, Tardieu F, Tuberosa R. Quantitative trait loci and crop performance under abiotic stress: Where do we stand? Plant Physiology. 2008;**147**:469-486

[171] Habier D, Fernando RL, Dekkers JCM. Genomic selection using low-density marker panels. Genetics. 2009;**182**:343-353

[172] Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. Crop Science. 2009;**49**:1-12

[173] Doná AA, Miranda GV, DeLima RO, Chaves LG, eGama EEG. Genetic parameters and predictive genetic gain in maize with modified recurrent selection method. Chilean Journal of Agricultural Research. 2012;**72**(1):33-39

[174] Ramalho MAP, Abreu ÂFB, Santos JB. Genetic progress after four cycles of recurrent selection for yield and grain traits in common bean. Euphytica. 2005;**144**:23-29

[175] Benites FRG, Pinto CABP. Genetic gains for heat tolerance in potato in three cycles of recurrent selection. Crop Breeding and Applied Biotechnology. 2011;**11**:133-140

[176] Gororo NN, Eagles HA, Eastwood RF, Nicolas ME, Flood RG. Use of *Triticum tauschii* to improve yield of wheat in low-yielding environments. Euphytica. 2002;**123**:241-254

[177] Machado JC, de Souza MA, de Oliveira DM, Cargnin A, Pimentel AJB, de Assis JC. Recurrent selection as breeding strategy for heat tolerance in wheat. Crop Breeding and Applied Biotechnology. 2010;**10**:9-15

[178] Shi W, Muthurajan R, Rahman H, Selvam J, Peng S, Zou Y, Jagadish KSV. Source-sink dynamics and proteomic reprogramming under elevated night temperature and their impact on rice yield and grain quality. The New Phytologist. 2013;**197**:825-837

[179] Ishimaru T, Hirabayashi H, Ida M, Takai T, San-Oh YA, Yoshinaga S, Ando I, Ogawa T, Kondo M. A genetic resource for early-morning flowering trait of wild rice *Oryza officinalis* to mitigate high temperature-induced spikelet sterility at anthesis. Annals of Botany. 2010;**106**:515-520

[180] Ashraf M, Athar HR, Harris PJC, Kwon TR. Some prospective strategies for improving crop salt tolerance. Advances in Agronomy. 2008;**97**:45-110

[181] Grover A, Mittal D, Negi M, Lavania D. Generating high temperature tolerant transgenic plants: Achievements and challenges. Plant Science. 2013;**205-206**:38-47

[182] Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiology. 2009;**149**:188-195

[183] Datta K, Baisakh N, Maung Thet K, Tu J, Datta S. Pyramiding transgenes for multiple resistance in rice against bacterial blight, yellow stem borer and sheath blight. Theoretical and Applied Genetics. 2002;**106**:1-8

[184] Cattivelli L, Rizza F, Badeck FW, Mazzucotelli E, Mastrangelo AM, Francia E, Marè C, Tondelli A, Stanca AM. Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. Field Crops Research. 2008;**05**:1-14. DOI: 10.1016/j.fcr.2007.07.004

[185] Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. Food security: The challenge of feeding 9 billion people. Science. 2010;**327**:812-818

[186] Bevan MW, Uauy C. Genomics reveals new landscapes for crop improvement. Genome Biology. 2013;**14**:206

[187] Cobb JA, DeClerck G, Greenberg A, Clark R, McCouch S. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics. 2013;**126**:867-887

[188] White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffeli TA, Conley MM, Feldmann KA, French AN, Heun JT, Hunsaker JD, Jenks MA, Kimball BA, Roth RL,

Strand RJ, Thorp KR, Wall GW, Wang G. Field-based phenomics for plant genetics research. Field Crops Research. 2012;**133**:101-112

[189] Amani I, Fischer RA, Reynolds MP. Canopy temperature depression association with yield of irrigated spring wheat cultivars in hot climate. Journal of Agronomy and Crop Science. 1996;**176**:119-129

[190] Janke E, Körner O, Rosenqvist E, Ottosen CO. High temperature stress monitoring and detection using chlorophyll a fluorescence and infrared thermography in chrysanthemum (*Dendranthema grandiflora*). Plant Physiology and Biochemistry. 2013;**67**:87-94

[191] Prashar A, Jones HG. Infra-red thermography as a highthroughput tool for field phenotyping. Agronomy. 2014;**4**:397-417

[192] Sharma DK, Andersen SB, Ottosen CO, Rosenqvist E. 2015. Wheat cultivars selected for high Fv/Fm under heat stress maintain high photosynthesis, total chlorophyll, stomatal conductance, transpiration and dry matter. Plant Physiology. 2015;**153**(2):284-298. DOI: 10.1111/ppl.12245

[193] Mason RE, Singh RP. Considerations when deploying canopy temperature to select high yielding wheat breeding lines under drought and heat stress. Agronomy. 2014;**4**:191-201

[194] Jones HG, Sirault XRR. Scaling of thermal images at different spatial resolution: The mixed pixel problem. Agronomy. 2014;**4**:380-396

[195] Hartmann A, Czauderna T, Hoffmann R, Stein N, Schreiber F. HTPheno: An image analysis pipeline for high-throughput plant phenotyping. BMC Bioinformatics. 2011;**12**:148

[196] Kast EJ, Nguyen MDT, Lawrence RE, Rabeler C, Kaplinsky NJ. The rootscope: A simple high-throughput screening system for quantitating gene expression dynamics in plant roots. BMC Plant Biology. 2013;**13**:158

Chapter 4

Functional Mapping of Plant Growth in *Arabidopsis thaliana*

Kaiyue Liu, Wenhao Bo, Lina Wang,
Rongling Wu and Libo Jiang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74424

Abstract

Most traits important to agriculture, biology, and biomedicine are complex traits, determined by both genetic and environmental factors. The complex traits that change their phenotypes over different stages of development are called dynamic traits. Traditional quantitative trait loci (QTLs) mapping approaches ignore the dynamic changes of complex traits. Functional mapping, as a powerful statistical tool, can not only map QTLs that control the developmental pattern and process of complex traits, but also describe the dynamic changes of complex traits. In this study, we used functional mapping to identify those QTLs that affect height growth in 10th generation recombinant inbred lines derived from two different *Arabidopsis thaliana* accessions. Functional mapping identified 48 QTLs that are related to height traits. The growth curves of different genotypes can be drawn for each significant locus. By GO gene function annotations, we found that these QTLs detected are associated with the synthesis of biological macromolecules and the regulation of biological functions. Our findings provide unique insights into the genetic control of height growth of *A. thaliana* and will provide a theoretical basis for the study of complex traits.

Keywords: complex traits, QTL, functional mapping, growth, *Arabidopsis thaliana*

1. Introduction

Complex traits are genetic traits controlled by multiple genes. They are sensitive to environmental changes and easily affected by the environment [1]. The phenotypic expression of complex traits in individuals within a population displays a continuous variation and generally a normal distribution. Most important biochemical, medical, and agronomic traits

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (cc) BY

and the majority of human diseases are complex traits that are controlled by interconnected genetic networks and environmental factors. In these gene networks, the effect of each gene is small [2]. The characteristics also change at different developmental stages and demonstrate the dynamic features of complex traits. The study of these complex traits is an important topic in modern biology.

Genetic mapping aimed at mapping underlying genes to genomic locations is a powerful tool for dissecting the genetic architecture of complex traits. Lander and Botstein have proposed an approach for mapping quantitative trait loci (QTLs) based on a sparse-density linkage map of molecular markers. This so-called interval mapping method can overcome the confounding problem of marker-QTL recombination [3]. Composite interval mapping includes other markers as covariates to control the overall genetic background, which displays increased power in QTL detection [4]. Considering QTL-QTL epistatic interactions in a linkage map, Kao et al. proposed using multiple marker intervals to map QTLs [5]. Currently, statistical methodologies for QTL mapping include regression analysis, maximum likelihood, and the Bayesian approach. With the development of high-throughput single nucleotide polymorphism (SNP) genotyping techniques, genome-wide association studies (GWAS) have provided a powerful means of mapping a complete set of genes underlying complex traits [6–8]. The genetic structure of a trait is explained by GWAS, which identify the numbers and chromosomal locations of each gene, the size of each gene's unique and pleiotropic effects, and the relative contributions of additive, dominant, and epistatic genetic effects. This provides an unprecedented tool for preparing a genotype-phenotype map [9]. So far, GWAS have detected many genetic variants for a wide range of complex traits, including those pertaining to agriculture, forestry, and human disease [9, 10].

It should be noted, however, that traits such as height and weight vary with time or other independent environmental stimuli. The traditional QTL mapping method directly associates a single marker with a single phenotype at a time point, which ignores the dynamic characteristics of organisms at different developmental stages, and cannot exactly reflect the whole genetic architecture of complex traits. Although thousands of QTLs were detected in many individuals, only a small number of QTLs were cloned and separated [11]. The reason for this problem is that the QTLs that have undergone rigorous statistical testing are divorced from biological relevance, which limited the projections of the genetic structure of traits. Genetic analysis of dynamic traits presents a serious statistical challenge. To solve these problems, Ma et al. [13] proposed a QTL mapping method based on a logistic-mixture model [12]. The QTL effect on developmental traits during ontogeny is considered as a function of time, and a series of growth formulas can be derived from the logistic curve describing plant height, size, and weight [13], arriving at a model which is expected to be improved in parameter estimation and statistical inference over previous models. Ma et al. [13] developed a maximum likelihood statistical framework based on a logistic-mixture model for the characteristics of function-valued traits, which change as a function of a specific variable. This QTL mapping strategy is called functional mapping.

Functional mapping combines mathematical functions that describe biological processes and assembles mathematical formulas into the statistical framework of QTL mapping to study the interactions between genes and phenotypic traits of organisms during growth and development. We estimate the parameters of a specific genotype that determines the development of

http://dx.doi.org/10.5772/intechopen.74424

a trait by statistical hypothesis testing and parameter estimation, rather than directly estimating the genetic effects of the gene at all-time points. Because of the combination of biological laws (defined by mathematical models) and the reduced number of genetic parameters that need to be estimated, functional mapping increased the potency of detecting significant QTLs [14]. Statistical merits of functional mapping will be exemplified in which case data are recorded irregularly, bringing on data sparsity. Such sparse longitudinal data cannot be well analyzed by traditional mapping for two reasons. First, because the problem of missing data existing at a given time point, traditional method is unable to use all of individuals, thus leading to a biased parameter estimation and reduced power of QTL detection. Second, individuals are measured at a few number of time points, limiting the fit of growth equation. Functional mapping is robust for handling longitudinal sparse data in which no single time point has the phenotypic data for all individuals, facilitating the QTL mapping to study the genetic architecture of hard-to-measure traits.

A set of tree data is used to assess the statistical validity of functional mapping. Several QTLs affecting the developmental trajectories of poplar were detected with the QTL mapping method based on the logistic-mixture model [15], and these QTLs were located on a genetic linkage map constructed by polymorphic markers. Studies have shown that functional mapping is useful in establishing gene-phenotype relationships and predicting QTL phenotypes of individual organisms based on the control of a trait. Functional mapping combines the principles of Mendelian genetics with statistical and developmental mechanisms, and is superior to traditional QTL mapping methods that combine the principles of statistics and genetics. To date, functional mapping has been used for mapping dynamic QTLs in poplars [16], jujubes [17], soybeans [18], rice [19], maize [20], yeast [21], oysters [22], mice [23], humans [24], and drug responses [25, 26].

Arabidopsis thaliana is a small, annual or winter annual, rosette plant. *A. thaliana* is a central genetic model and universal reference organism in plant and crop science. The successful integration of different fields of research in the study of *A. thaliana* has made a large contribution to our molecular understanding of key concepts in biology. The *Arabidopsis* reference genome sequence was the first published nuclear genome of a flowering plant in 2000 (http://www.arabidopsis.org) [27, 28]. The ease and speed with which experiments can be conducted on *A. thaliana* has allowed enormous fundamental progress in our knowledge of the molecular principles of plant development, cell biology, metabolism, physiology, genetics, and epigenetics [29]. The uses of *Arabidopsis* as the universal reference plant continue to expand, particularly in the field of systems biology [30, 31]. Important work has been done to investigate the molecular networks that mediate environmentally controlled developmental switches in *A. thaliana*. Examples include the transition from vegetative to reproductive development, also termed flowering time control [32–34], seed dormancy and germination control [35, 36], and the light-regulated development of seedlings [37, 38]. *A. thaliana* has also served as a model research organism for exploring many areas of fundamental biology, including photobiology, the circadian clock, DNA methylation, DNA repair, RNA silencing, protein degradation, and G-protein signaling, many of which have direct application in human health [39–42]. Insights into the functions of a multitude of individual genes, as well as the elucidation of biosynthetic pathways and regulatory networks, in *A. thaliana* have proven invaluable for identifying the genetic basis of agronomically important traits in crops such as plant height and flowering time [43, 44]. Classical molecular genetics approaches have been used in

A. *thaliana* to dissect the patterning and development of flowers [45], embryos [46], leaves, and roots [47]. In this study, we describe the implementation of functional mapping to identify and map QTLs for height trajectories in a population of *A. thaliana*. The mapping population is composed of 144 recombinant inbred lines (RILs) derived from a Landsberg erecta (LER) cultivar and a Shadara (SHA) cultivar. Functional mapping identified 48 QTLs that determine the height growth of *A. thaliana*. The identification of these QTLs will help us address fundamental questions about the genetic mechanisms of height growth.

2. Materials and methods

2.1. Materials

In this study, the mapping population included 114 RILs derived from two cultivars of *Arabidopsis*. The main phenotype of the LER cultivar includes round leaves, short petiole, short pedicel, flowers clustered at inflorescence tips, short pod width, sharp tip, inflorescence compact, and a plant height of 10–25 cm. The general traits of the SHA cultivar are described in The *Arabidopsis* Information Resource database as a slightly narrow leaf with a height of about 30 cm. To ensure adequate seeds for the sustainability of the material, each RIL carried out a generation of expansion. For the sterilization of *Arabidopsis* seeds, the seeds were shaken in a centrifuge tube with 10% sodium hypochlorite for 6 min; 1 ml of 95% alcohol was then added, and the tubes were shaken for 5 min. The seeds were rinsed 5–6 times to thoroughly wash away the remaining sodium hypochlorite solution on the seed surface. The alcohol was poured out, and the tubes were dried at a super clean bench to obtain sterile *Arabidopsis* seeds.

For vernalization of *A. thaliana* seeds, distilled water was added to the seeds and the seeds were placed in a 4°C refrigerator for 3 days of dark treatment before sowing. To sow the seeds after vernalization, vermiculite, peat, and limestone were mixed evenly in a 1:1:1 ratio and used as the soil. An 18 × 18 cm plastic grid with 90 squares was placed over the plant pot. The seeds were placed into each mesh with a toothpick. The planting diagram of the experiment is shown in **Figure 1**. Circles and triangles represent the two different lines of *A. thaliana*, planted in the same growing space. After planting, the plant pot was covered with plastic wrap and then transferred to the long day (16 h) and short day (8 h) artificial climate room (22°C). When the *Arabidopsis* sprouts displayed green shoots, the plastic wrap was removed.

2.2. Phenotypic data acquisition

Plant height was measured after 1 week of planting. Each line of *A. thaliana* was randomly selected for 10 strains. In total, 1160 strains of *Arabidopsis* were selected. The height of *Arabidopsis* plants were measured once per week until the end of the plant's life cycle by manual measurement with a 1 mm ruler (accurate to 1 mm).

The plant height phenotype was measured for 1160 *A. thaliana* strains. The height of 1–60 lines of *Arabidopsis* was measured seven times, and the height of 60–116 lines of *Arabidopsis* was measured seven times. Because of environmental or genotypic reasons, nine lineages of *Arabidopsis* did not complete a life cycle (numbers 1–4, 22, 28–30, and 60).

http://dx.doi.org/10.5772/intechopen.74424

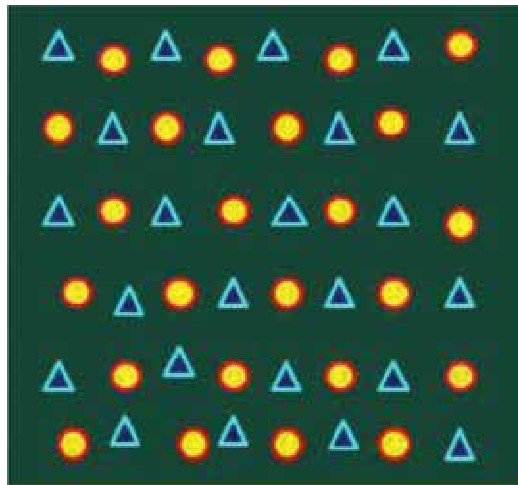


Figure 1. Schematic diagram of *Arabidopsis* planting.

The growth of *Arabidopsis* strains can be seen in **Figure 2a–c**, which indicates the growth of the strains for 1, 3, and 6 weeks, respectively.

2.3. Acquisition of SNP data

A. thaliana DNA was extracted with a DNA extraction kit (TIANGEN DP305) according to the manufacturer's instructions. The plant material used was 100 mg of the young leaf. After DNA was detected in the samples, the DNA sequence was determined by an outside company. The resulting sequence data was subjected to SNP calling.

2.3.1. DNA quality inspection

The degree of degradation of the DNA was analyzed by agarose gel electrophoresis and imaged with an ultraviolet gel imager. The DNA sample was determined to be without degradation if

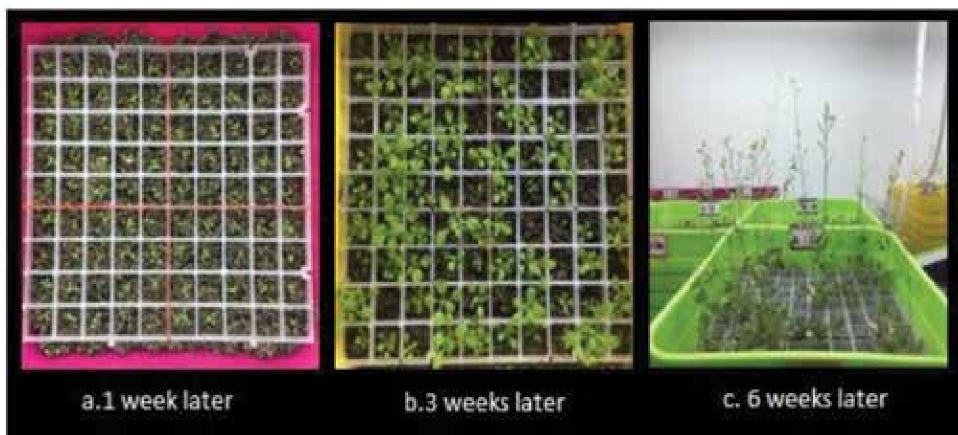


Figure 2. *Arabidopsis* growth conditions at 1 week (a), 3 weeks (b), and 6 weeks (c).

the sample did exhibit dispersion phenomena. All of the samples were determined not to be degraded and were sequenced.

We measured the absorbance of the DNA samples at wavelengths of 260 and 280 nm and calculated the value of the optical density (OD) 260/OD280. Samples were determined to be of sufficient quality if this value was greater than 1.8 and less than 2.2. After testing, all samples were found to meet the requirements for sequencing.

2.3.2. Sequencing and mapping

DNA from the sample was sequenced on the Illumina HiSeq 2000 platform to generate 125 bp paired-end reads at greater depth. All these RILs sequencing were performed at Total Genomics Solution (TGS) Institute. After sequencing quantity control, the data of 107 samples reached 117.68 G, with an average of 1.10 G per sample. The percentage of Q30 bases was more than 90%, the percentage of Q20 bases was more than 95%, and the distribution of GC was normal. Thus, the quantity and quality of the sequencing met the requirements for subsequent analysis.

The per-base coverage depth across all contigs was calculated by mapping raw reads from each RIL against reference genomes. The results showed that the average mapping ratio and sequencing coverage of the samples were higher than 95 and 92%, respectively. The sequence depth of the samples was 9.91×. In addition to the line 110, the other lines are better for subsequent variation detection and analysis.

2.4. SNP calling

SNP detection was performed with the widely accepted mutation detection software GATK. SNP. The screening criteria were as follows: The depth of sequencing for each sample was greater than or equal to 4, otherwise the sample was marked as missing. Additionally, the quality value of the comparison must be greater than or equal to 20, and the variation detection quality value must be greater than or equal to 50. If the allele frequency was less than 5% or the absence rate of the sample was greater than 50%, the site was filtered out.

The RIL population contains 105 progeny and two parents; the parent number is "A518_LER" and "A518_SHA." The reads of each sample were compared with the reference genome. The average ratio of the samples was above 95%, and the coverage was greater than 92%. The average sequencing depth was 9.19×. By processing, we obtained 107 samples and 1,023,325 whole genome SNP markers. According to the situation of this population, individuals with heterozygous genotype ratios over 25% and markers indicating that parents possessed heterozygous genotypes were eliminated. Finally, 609,427 SNPs and 80 individuals met the model requirements and could be used to detect the QTLs that affect the growth height of *A. thaliana*.

3. Functional mapping

3.1. Statistical model

Let $y_i = c(y_i(1), \cdots, y_i(T))$ denote the vector of trait values for RIL, *i* measured at *T* time-points. Consider a SNP with two alleles *Q* and *q*, generating two genotypes: *QQ* with n_1 RILs and *qq* with n_2 RILs. In this study, the likelihood for height growth data of *A. thaliana* is expressed as

$$L(\Phi \mid y) = \prod_{i=1}^{n_1} f_1(y_i) \prod_{i=1}^{n_2} f_2(y_i) \tag{1}$$

where Φ indicates the unknown parameters including the time-dependent effects of different QTL genotypes and the time-dependent residual variance and correlations. $f_1(y_i)$ and $f_2(y_i)$ are a multivariate normal distribution with a time-dependent mean vector for genotype *QQ* and *qq*,

$$\begin{cases} \mu_1 = (\mu_1(1), \ldots \mu_1(T)) \text{ for } QQ, \\ \mu_2 = (\mu_2(1), \ldots \mu_2(T)) \text{ for } qq. \end{cases} \tag{2}$$

(T × T)-dimensional longitudinal covariance matrix is expressed as Σ, which can be modeled by using a statistical approach such as the first-order autoregressive [AR(1)] model or an autoregressive moving-average process (ARMA). The maximum-likelihood estimates (MLEs) of the unknown parameters are implemented with the simplex algorithm in R software [48].

3.2. Modeling the mean vector

One of the most important equations for capturing time-specific change in growth is the logistic curve [49, 50], which we used to describe height growth of the QTL genotype according to the following expression:

$$g(t) = \frac{a}{1 + be^{-rt}}, \tag{3}$$

where $g(t)$ represents the trait value at time point t, *a* indicates the asymptotic value of g when $t \to \infty$, *b* is a parameter to position the curve on the time axis, and *r* indicates the relative growth rate. Consequently, any specific growth characteristics described by the logistic growth equation can be captured by parameter *a*, *b*, and *r*, and these can be used to determine the coordinates of biologically important benchmarks along the growth trajectory. The mean vector for the QTL genotypes *QQ* and *qq* from time *1* to *T* in the multivariate normal density function is expressed as:

$$\begin{cases} \mu_1 = \left(\frac{a_1}{1+b_1 e^{-r_1}}, \ldots, \frac{a_1}{1+b_1 e^{-r_1 T}} \right) for\ QQ, \\ \mu_2 = \left(\frac{a_2}{1+b_2 e^{-r_1}}, \ldots, \frac{a_2}{1+b_2 e^{-r_2 T}} \right) for\ qq. \end{cases} \tag{4}$$

3.3. Modeling the covariance structure

The first-order autoregressive [AR(1)] model has been successfully applied to model the structure of the within-subject covariance matrix for functional mapping. The AR(1) model includes two basic assumptions: (1) variance stationarity (the residual variance, σ^2) is unchanged over time points and (2) covariance stationarity (the correlation between different time points) decreases proportionally (in ρ) with increased time interval. The AR(1) is described as:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \mathrm{L} & \rho^{T-1} \\ \rho & 1 & L & \rho^{T-2} \\ \mathrm{L} & \mathrm{L} & \mathrm{L} & \mathrm{L} \\ \rho^{T-1} & \rho^{T-2} & \mathrm{L} & 1 \end{bmatrix}, \tag{5}$$

where $0 < \rho < 1$ is the proportion parameter with which the correlation decays with a time point.

3.4. Hypothesis tests

After the MLEs of the parameters are obtained, the hypothesis concerning the existence of a QTL affecting overall growth can be written as:

$$\begin{cases} H0: a_1 = a_2, b_1 = b_2, r_1 = r_2 \\ H1: at\ least\ one\ of\ the\ equalities\ above\ does\ not\ hold' \end{cases} \tag{6}$$

where the null hypothesis H_0 corresponds to the reduced model and the alternative hypothesis H_1 corresponds to the full model. The test statistics for testing the hypotheses is the log-likelihood ratio (LR) of the full over reduced model. The critical threshold is determined from permutation tests.

3.5. Candidate gene function annotation

Gene ontology (GO) annotation analysis was performed using Blast2GO [51]. Finally, R script language programming was used to translate the GO annotation results into charts.

4. Results

4.1. QTL detection

By plotting the total growth against growth week, it was observed that each of the 116 mapped genotypes followed the logistic growth curve. **Figure 3** illustrates the growth trajectories of

http://dx.doi.org/10.5772/intechopen.74424

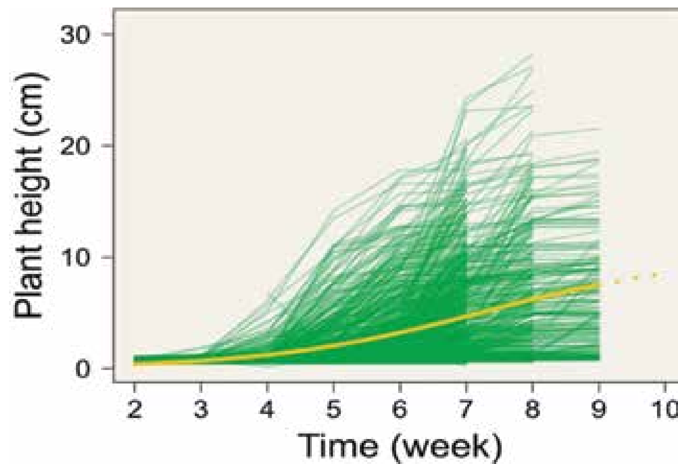


Figure 3. The growth trajectories of height for individual recombinant inbred lines (RILs) are shown in green, whereas yellow lines are the mean growth trajectories of all RILs fitted to a curve. The x- and y-axes of the plot denote time (in weeks) and height (in cm).

height for each RIL over 9 weeks. A least-squares approach was used to fit height growth with the logistic curve (Eq. (3)) for each RIL. Based on statistical tests, all genotypes can be fit by a logistic curve containing parameters a, b, and r ($P < 0.01$, **Figure 3**).

According to the fit results, the relative growth rate of *A. thaliana* was $r = 0.61$, and the asymptotic value of plant height was calculated to be $a = 10.08$. We can see from **Figure 3** that the plant height of *A. thaliana* increases exponentially over a period of time, and as growth time increases, the trend of the curve begins to flatten, until the plant height does not change anymore. In addition, different genotypes showed different growth curves, suggesting the possibility of genetic control over the growth trajectories. The statistical model built upon the logistic growth curve model is used to map QTLs responsible for growth trajectories in plant height.

Functional mapping was implemented to analyze the mapping population. A Manhattan plot of LR values against the genome locations of SNPs distributed throughout the genome is shown in **Figure 4**. We found QTLs that affect the growth trajectory of plant height on the fourth chromosome of the RIL population (**Figure 4**). The genome-wide empirical estimate of the critical value is obtained from permutation tests. The profile of the LR value of the full and reduced model across the length of chromosome 4 has a clear peak from 7.9 to 22.5 kb. A total of 48 significant QTLs responsible for growth trajectories of plant heights were identified. The LR value at this peak is 437.78, which is beyond the empirical critical threshold at the significance level $p = 0.05$.

Functional mapping can also be used to observe the dynamic expression of a QTL over time such as when a QTL starts to affect a growth process. The parameter combination of the *QQ* genotype was $a_1 = 9.07$, $b_1 = 370.08$, and $r_1 = 0.92$. The parameter combination of the *qq* genotype was $a_2 = 39.62$, $b_2 = 114.61$, and $r_2 = 0.28$. Different QTL genotypes corresponded to different parameter combinations, which indicate that the QTL controls the developmental

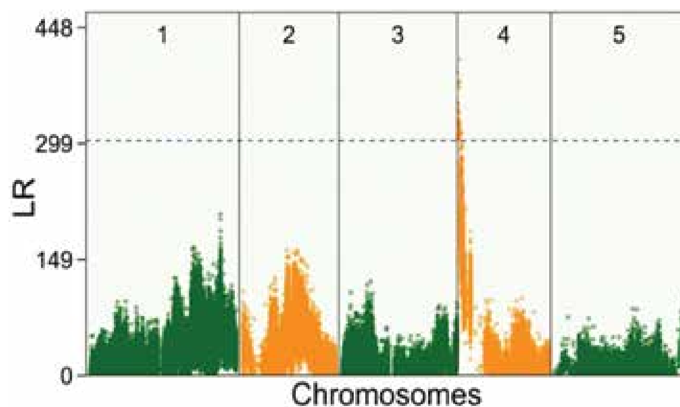


Figure 4. The log-likelihood ratio (LR) of height change of *Arabidopsis thaliana*. The profiles of the LR between the full model (there is a quantitative trait locus [QTL]) and reduced model (there is no QTL) for height growth trajectories throughout the *A. thaliana* genome. The critical thresholds for claiming the genome-wide existence of a QTL are obtained from permutation tests.

trajectory of the plant height. Using the estimates from the growth curves, we drew two different curves, each corresponding to a genotype at each of the detected QTLs on the fourth chromosome (**Figure 5**).

On the basis of the hypothesis test (6), this QTL is detected to be inactive until *A. thaliana* grew to 4 weeks, and its effect on height growth increased with time. At 9 weeks, the genotype *qq* exhibited height growth more than its alternative *QQ*. The effect of *qq* on plant height was more significant than that of *QQ*. If different genotypes at a given QTL correspond to different trajectories, the QTL must affect the differentiation of this trait. Apparently, this mapped QTL interacts significantly with time to affect the height growth of *A. thaliana*.

4.2. Candidate gene function annotation

GO classification is widely used for gene classification and functional annotation, and GO provides three types of semantic terms, including cell component, molecular function, and biological

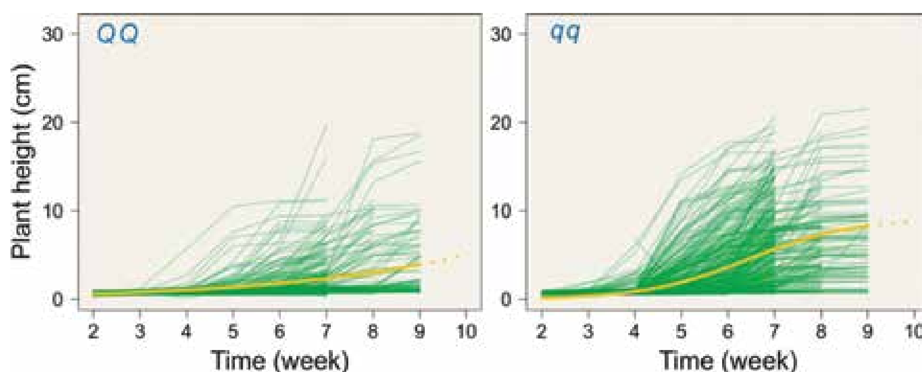


Figure 5. Growth trajectory for different genotypes at the QTLs detected on groups.

process, to describe the characteristics of genes and gene products [52, 53]. To further summarize the data of GO classification, the cytological components include cell structure, tissue, protein complex, extracellular structure, and cell process. Biological processes include developmental processes, physiological processes, regulatory processes, and the processes of responding to stress. Molecular functions include binding, catalysis, activation, structural molecules, and transcriptional regulatory functions (**Figure 6**).

Functional prediction and classification analysis of 48 loci were screened with the National Center for Biotechnology and the Joint Genome Institute (JGI) databases. There are 20 gene families in these loci, which include the F-box and calcium-binding EF-hand protein families. Pathway cluster analysis was used to compare the 48 genes with the known protein sequences in the JGI database.

It was found that 7 of the 48 genes corresponded to functions in the JGI database. They were divided into nine groups, including carbon fixation pathways in prokaryotes, biosynthesis of

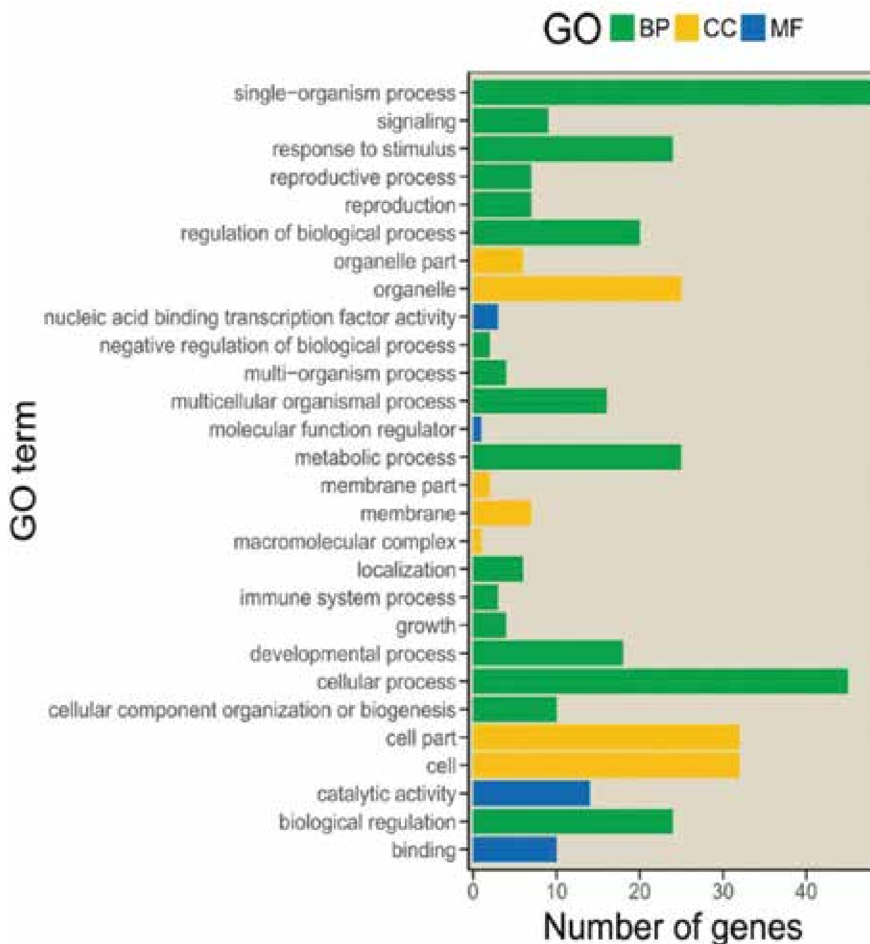


Figure 6. Gene ontology annotations of significant sites.

antibiotics, one-carbon pool by folate, starch, and sucrose metabolism, glycerolipid metabolism (synthase), glycerolipid metabolism (lipase), polyketide sugar unit biosynthesis, streptomycin biosynthesis, and pentose and glucuronate interconversions. We found that these nine pathways are mostly associated with carbohydrate, energy, and amino acid metabolism. These biological processes play an important role in the development of height in *Arabidopsis*.

We also found that AT4G00160.1 encodes an F-box protein in the signal transduction pathway. It has been shown that F-box is an auxin receptor [54]. Plant height development is regulated by gibberellin (GA) and auxin (indole-3-acetic acid [IAA]) [55], and GA20ox and GA30ox are encoded by multiple genes, and mutations in these loci can result in dwarfing of the plant in the later stage of GA biosynthesis [56]. The most significant site, AT4G01150.1, is related to protein curvature thylakoid chloroplastic. Protein curvature thylakoid chloroplastic tends to be located in leaves and stems, and plays an important role in plant photosynthesis, which affects plant growth [57].

5. Discussion

Gene mapping has been shown to be a powerful approach for the study of the genetic architecture of complex traits. It has been instrumental for the characterization of QTLs that control quantitative traits of interest to agriculture, biology, and human disease [58, 59]. However, traditional mapping strategies do not provide much insight into the genetic control mechanisms for phenotypic variations if some statistical and biological issues related to the approach are not resolved. Ma et al. (2002) integrated some fundamental biological principles into the mapping framework, aimed at generating more biologically meaningful discoveries related to trait formation and development, further proposing so-called functional mapping [13]. Functional mapping attempts to combine strong statistical and molecular genetics with the developmental mechanisms of biological features, and to elucidate the genetic mechanisms of complex traits. Since functional mapping combines different mathematical functions with biological significance, it possesses three advantages over traditional mapping methods in QTL mapping: (1) because the underlying biological mechanism is considered, the results of functional mapping are closer to biological reality; (2) a smaller sample size can be used to achieve sufficient accuracy for QTL detection because multiple measurements of the same individual improve mapping accuracy; and (3) by treating the growth process as a smooth curve, a large number of variables can be analyzed simultaneously, and the estimation of a small number of parameters can improve the accuracy of the parameter estimation and flexibility of the model.

With the development of high-throughput sequencing technology and the reduction of sequencing cost, GWAS have become an important tool for studying complex traits and have been widely used in genetic studies of complex traits in humans, animals, and plants [60]. Most GWAS only use single phenotypic data to perform regression analysis with each SNP such as with Plink software [61]. In addition, some GWAS have been developed to solve the false positive loci of population structure and genetic relationship [62–64]. The successes and potential of GWAS have not been explored when complex phenotypes arise as a curve. In any regard, a curve is more informative than a point in describing the biological features of a trait. To apply functional mapping to GWAS by integrating GWAS and functional aspects of

dynamic traits, a new analytical model for genome-wide association analysis of dynamic data, called functional GWAS (fGWAS), has been derived [65]. There are two advantages to fGWAS over GWAS: (1) fGWAS is able to identify genes that determine the final form of the trait and (2) it provides the ability to study the temporal pattern of genetic control over a time course.

The regulation network of plant height traits has been studied intensively in molecular biology. We already know that the development of plant height traits is regulated by growth hormones such as GA and IAA. Using functional mapping, we found 48 growth QTLs in *A. thaliana*. Through the GO annotation of QTLs, we found that there are many genes among the significant loci identified in this study that are related to the pathways for synthesis and conduction of growth hormones, such as AT4G00160.1, which encodes an F-box protein in the signal transduction pathway. It has been shown that the F-box is an auxin receptor. Thus, we can see that the QTLs identified in this study may not only be applicable to *A. thaliana*, but also to other plants. These results show that functional mapping can reveal more intricate details of dynamic traits such as height growth and other phenotypes.

Functional mapping is far from enough to fully study complex traits, and there are still many limitations in describing the developmental pathways leading to the final phenotype and revealing the underlying genetic mechanisms for the formation and development of these traits. It is too simple to draw a complete dynamic diagram of complex traits. Wu [66] extends functional mapping to system mapping. By identifying the dynamic formation process of complex traits as a system and decomposing it into several parts, the QTL that controls the interaction of each component during the development of complex characters is identified. From the point of view of ecology, the process of character formation is extremely complex. To draw a complete quantitative genetic structure, we need to study the characteristics of an organism affected by its own genes as well as the influence from the community partner genome. In nature, most organisms live in groups, and individuals compete with each other. Wu combined game theory with QTL mapping, which opens up new opportunities for improving the accuracy and resolution of complex phenotype QTL recognition [67].

Acknowledgements

This work is supported by National Natural Science Foundation of China (grant 31700576) and grant 201404102 from the State Administration of Forestry of China.

Author details

Kaiyue Liu[1], Wenhao Bo[1], Lina Wang[1], Rongling Wu[1,2] and Libo Jiang[1*]

*Address all correspondence to: libojiang@bjfu.edu.cn

1 Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

2 Center for Statistical Genetics, Pennsylvania State University, Hershey, PA, USA

References

[1] Zhong WH, Xin JX. Genetic analysis method of developmental characters QTL. Journal of Henan Institute of Science and Technology (Nature Science Edition). 2008;**36**(04):7-10

[2] Cheverud JM. Genetics and analysis of quantitative traits. American Journal of Physical Anthropology. 1999;**108**(03):548-549

[3] Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989;**121**(01):185-199

[4] Zeng ZB. Precision mapping of quantitative trait loci. Genetics. 1994;**136**(04):1457-1468

[5] Kao CH, Zeng ZB, Teasdale RD. Multiple interval mapping for quantitative trait loci. Genetics. 1999;**136**(04):1457-1468

[6] Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genetics. 2010;**6**(09):e1001139

[7] Hindorff LA, MacArthur J, Morales J, Junkins HA, Hall PN, Klemm AK, Manolio TA. A Catalog of Published Genome-Wide Association Studies. 2011. www.genome.gov/gwastudies/ [Accessed May 27, 2011]

[8] Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C. Identification of ten loci associated with height highlights new biological pathways in human growth. Nature Genetics. 2008;**40**:584-591

[9] Sun L, Wu R. Mapping complex traits as a dynamic system. Physics of Life Reviews. 2015;**13**:155

[10] Shilin Y, Yi W, Zu ZZ, Jun XW, Nan C, Guang YX, Jian C. Genome-wide association study on susceptibility genes associated with Yang-deficiency constitution: A small sample case-control study. Chinese Journal of Integrative Medicine. 2015;**21**(08):601-609

[11] Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Lai J, et al. Genome-wide genetic changes during modern breeding of maize. Natue Genetics. 2012;**44**(7):813

[12] Leal SM. Genetics and analysis of quantitative traits. American Journal of Human Genetics. 2001;**68**(2):548

[13] Ma CX, Casella G, Wu R. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. Genetics. 2002;**161**(04):1751-1762

[14] Rice SH. The analysis of ontogenetic trajectories: When a change in size or shape is not Heterochrony. Proceedings of the National Academy of Sciences of the United States of America. 1997;**94**(03):907-912

[15] Zhong WH. Genetic analysis and functional mapping of complex quantitative traits. Henan Agricultural Science. 2006;**35**(11):19-22

[16] Wu R, Ma CX, Lou XY, Casella G. Molecular dissection of allometry, ontogeny, and plasticity: A genomic view of developmental biology. AIBS Bulletin. 2003;**53**(11):1041-1047

[17] Zhang B, Tong C, Yin T. Detection of quantitative trait loci influencing growth trajectories of adventitious roots in Populus using functional mapping. Tree Genetics and Genomes. 2009;**5**:539-552

[18] Huang Z, Tong C, Bo W, Pang X, Wang Z, Xu J, Wu R, et al. An allometric model for mapping seed development in plants. Briefings in Bioinformatics. 2013;**15**(4):562-570

[19] Li H, Huang W, Wu S. A conceptual framework to mapping quantitative trait loci governing ontogenetic allometry. PLoS One. 2007;**2**(08):1245

[20] Cui Y, Zhu J, Wu RL. Functional mapping for genetic control of programmed cell death. Physiological Genomics. 2006;**25**:458-469

[21] Coelho WS, Li Y. Identification of quantitative trait loci that affect endo reduplication in Zea mays endosperm. Theoretical and Applied Genetics. 2007;**115**:1147-1162

[22] Li N, McMurry T, Berg A. Functional clustering of periodic transcriptional profiles through ARMA. PLos One. 2010;**5**(04):e9894

[23] Liu T, Wu R. A Bayesian algorithm for: Functional mapping of dynamic traits. Algorithms. 2009;**2**:667-691

[24] Li N, Das K, Wu R. Functional mapping of human growth trajectories. Journal of Theoretical Biology. 2009;**261**:33-42

[25] Lin M, Aqvilonte C, Johnson J. Sequencing drug response with HapMap. Pharmaco Genomics. 2005;**5**:149-156

[26] Lin M, Hou W, Li H. Modeling: Sequence-sequence interactions for drug response. Bioinformatics. 2007;**23**:1251-1257

[27] Somerville C, Koornneef M. A fortunate choice: The history of Arabidopsis as a model plant. Nature Reviews Genetics. 2002;**3**:883-889

[28] Koornneef M, Meinke D. The development of Arabidopsis as a model plant. The Plant Journal. 2010;**61**:909-921

[29] Krämer U. Planting molecular functions in an ecological context with *Arabidopsis thaliana*. Elife. 2015;**4**:1-13

[30] Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. A highresolution root spatiotemporal map reveals dominant expression patterns. Science. 2007;**318**:801-806

[31] Endo M, Shimizu H, Nohales MA, Araki T, Kay SA. Tissue-specific clocks in Arabidopsis show asymmetric coupling. Nature. 2014;**515**:419-422

[32] Simpson GG, Dean C. Arabidopsis, the Rosetta stone of flowering time. Science. 2002; **296**:285-289

[33] Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. 2004

[34] Andres F, Coupland G. The genetic basis of flowering responses to seasonal cues. Nature Reviews Genetics. 2012;**13**:627-639

[35] Bentsink L, Jowett J, Hanhart CJ, Koornneef M. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. Proceedings of the National Academy of Sciences of USA. 2006;**103**:17042-17047

[36] Graeber K, Nakabayashi K, Miatton E, Leubner Metzger G, Soppe WJ. Molecular mechanisms of seed dormancy. Plant, Cell & Environment. 2012;**35**(10):1769-1786

[37] Deng XW, Matsui M, Wei N, Wagner D, Chu AM, Feldmann KA, Quail PH. COP1, an Arabidopsis regulatory gene, encodes a protein with both a zincbinding motif and a G beta homologous domain. Cell. 1992;**71**:791-801

[38] Nemhauser JL. Dawning of a new era: Photomorphogenesis as an integrated molecular. Current Opinion in Plant Biology. 2008;**11**:4-8

[39] Hays JB. *Arabidopsis thaliana*, a versatile model system for study of eukaryotic genome-maintenance functions. DNA Repair. 2002;**1**:579-600

[40] Thomma BP. Disease induction by human microbial pathogens in plant-model systems. Drug Discovery Today. 2007;**12**:167-173

[41] Jones AM, Chory J, Dangl JL, Estelle M, Jacobsen SE, Meyerowitz EM, Nordborg M, Weigel D. The impact of Arabidopsis on human health: Diversifying our portfolio. Cell. 2008;**133**:939-943

[42] Schlaich NL. *Arabidopsis thaliana* – The model plant to study host-pathogen interactions. Current Cancer Drug Targets. 2011;**12**:955-966

[43] Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. Cell. 2006;**127**:1309-1321

[44] Olsen KM, Wendel JF. A bountiful harvest: Genomic insights into crop domestication phenotypes. Annual Review of Plant Biology. 2013;**64**:47-70

[45] Meyerowitz EM, Bowman JL, Brockman LL, Drews GN, Jack T, Sieburth LE, Weigel D. A genetic and molecular model for flower development in *Arabidopsis thaliana*. Development. 1991;**1**:157-167

[46] Mayer U, Torres Ruiz RA, Berleth T, Misera S, Jurgens G. Mutations affecting body organization in the Arabidopsis embryo. Nature. 1991;**353**:402-407

[47] Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. Cell. 1996;**86**:423-433

[48] Ortutay C, Ortutay Z. Molecular Data Analysis Using R. Hoboken, New Jersey: Wiley Blackwell

[49] Niklas KJ. Plant Allometry: The Scaling of Form and Process. Chicago: University of Chicago Press. 1994

[50] West GB, Brown JH, Enquist BJ. A general model for ontogenetic growth. Nature. 2001;**413**:628-631

[51] Huang ZX, Ke CH, Chen J. Bioinformatics process of large scale GO annotation. Journal of Xiamen University (Natural Science Edition). 2012;**51**(1):139-143

[52] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Epping JT, Harrise MA, Hill DP, Issel-Tarver L, Kasarsk A, Lew S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: Tool for the unification of biology. Gene Ontology Constortium Nature Genetics. 2000;**25**:25-29

[53] Berardini TZ, Mundodi S, Reiser L, Huala E, Garacia-Hernandze M, Zhang P, Mueller Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckley B, Montoya M, Miller N, Weems D, Rhee SY. Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiology. 2004;**135**:745-755

[54] Sakamoto TK, Miura H, Itoh T, Tatsumi M. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiology. 2014;**134**:1642-1653

[55] Hedden P, Phillips AL. Gibberellin metabolism: New insights revealed by the genes. Trends in Plant Science. 2000;**5**(12):523-530

[56] Itoh H, Tatsumi T, Sakamoto T, Otomo K. A rice semi-dwarf gene, Tan-Ginbozu (D35), encodes the gibberellin biosynthesis enzyme, ent-kaurene oxidase. Plant Molecular Biology. 2004;**54**:533-547

[57] Li P, Liu Y, Wang H, He Y, Wang X, Lv F, Chen H, Pang X, Liu M, Shi T, et al. PubAngioGen: A database and knowledge for angiogenesis and related diseases. Nucleic Acids Research. 2015;**43**:D963-D967

[58] Niedziela A, Bednarek PT, Labudda M, Mańkowski DR, Anioł A. Genetic mapping of a 7R al tolerance QTL in triticale (x Triticosecale Wittmack). Journal of Applied Genetics. 2014;**55**(01):1-14

[59] Recla JM, Robledo RF, Gatti DM, Bult CJ, Churchill GA, Chesler EJ. Precise genetic mapping and integrative bioinformatics in diversity outbred mice reveals Hydin as a novel pain gene. Mammalian Genome. 2014;**25**(5-6):211-222

[60] Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and-complex traits. Nature Reviews Genetics. 2005;**6**(2):95-108

[61] Christopher CC, Carson CC, Laurent CT, Shashaank V, Shaun MP, James JL. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience. **4**(1):7

[62] Yu J, Pressoir G, Briggs WH. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics. 2006;**38**(02):203-208

[63] Klasen JR, Barbez E, Meier L. A multi-marker association method for genome-wide association studies without the need for population structure correction. Nature Communications. 2016;**7**:13299

[64] Stephan J, Stegle O, Beyer A. A random forest approach to capture genetic effects in the presence of population structure. Nature Communications. 2015;**6**:7432

[65] Stefan E, Lorenzo BJ. Handbook of Adaptive Designs in Pharmaceutical and Clinical Development. A. Pong and S. C. Chow (2010). Boca Raton: Chapman & Hall/CRC Press. Biometrical Journal. 2011;**53**(4):708-709

[66] Wu R, Cao JG, Huang ZW, Wang Z, Gai JY, Vallejos E. Systems mapping: How to improve the genetic mapping of complex traits through design principles of biological systems. BMC Systems Biology. 2011;**5**:84

[67] Zhu X, Jiang L, Ye M. Integrating evolutionary game theory into mechanistic genotype-phenotype mapping. Trends in Genetics. 2016;**32**(05):256-268

Chapter 5

The Usage of Genomic Selection Strategy in Plant Breeding

Mohd Shamshad and Achla Sharma

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76247

Abstract

Major paradigm shift in plant breeding since the availability of molecular marker technology is that mapping and characterizing the genetic loci that control a trait will lead to improved breeding. Often, one of the rationales for cloning of QTL is to develop the "perfect marker" for MAS, perhaps based on a functional polymorphism. In contrast, an advantage of genomic selection is precisely its black box approach to exploiting genotyping technology to expedite genetic progress. This is an advantage in our view because it does not rely on a "breeding by design" engineering approach to cultivar development requiring knowledge of biological function before the creation of phenotypes. Breeders can therefore use genomic selection without the large upfront cost of obtaining that knowledge. In addition, genomic selection can maintain the creative nature of phenotypic selection which couple's random mutation and recombination to sometimes arrive at solutions outside the engineer's scope. Currently, the lion's share of research on genomic selection has been performed in livestock breeding, where effective population size, extent of LD, breeding objectives, experimental design, and other characteristics of populations and breeding programs are quite different from those of crop species. Nevertheless, a great number of findings within this literature are very illuminating for genomic selection in crops and should be studied and built upon by crop geneticists and breeders. The application of powerful, relatively new statistical methods to the problem of high dimensional marker data for genomic selection has been nearly as important to the development of genomic selection as the creation of high-density marker platforms and greater computing power. The methods can be classified by what type of genetic architecture they try to capture.

Keywords: genomic selection, training population, breeding population, linkage disequilibrium, genomic prediction model, breeding value

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (cc) BY

1. Introduction

Marker-assisted selection (MAS) is an important scheme in plant breeding since the 1990s, after promising analysis results for tagging genes or mapping QTL [1]. Marker assisted selection and molecular breeding have been used in the identification of underlying major genes in gene pools and their transfer to desirables traits of major plant breeding programs. Using of MAS have shown some shortcomings due to long selection schemes and also the look for vital marker-QTL associations being unable to capture "minor" gene effects. Thus marker-assisted selection (MAS) is difficult to improve traits having complex inheritance such as grain yield and abiotic stresses.

Using whole-genome prediction models, the genomic selection (GS) strategy has paved the way to over-come these limitations. High-density molecular markers using is one of the main features of genomic selection. Therefore, each of the trait loci has the likelihood of being in linkage disequilibrium (LD) with a minimum of one marker locus within the entire breeding population. Genome selection strategy removes the need to mapping of genes and search for linked QTL–marker loci associated individually. Rather, Genomic selection accounts for bunches of predictors simultaneously and is characterized by constraining random estimates towards zero. Moreover, Genomic selection helpful for accelerate breeding cycles in such a way that the rate of annual genetic gain per unit of time and cost can be decreased. Genomic selection has been well established in the field of animal breeding, but is in its beginning in crops plants and forest tree breeding.

Genome-wide selection or genomic selection estimates marker effects across the full ordering of the breeding population (BP) supported the prediction model developed within the training population (TP). Training population could be a group of related individuals (such as half-sibs or lines) that are each phenotypes and genotypes. Breeding population typically is just genotyped not phenotypes. Hence, Genomic selection depends on the degree of genetic similarity between training population and breeding population within the Linkage disequilibrium between marker and trait loci. Breeding values have not been a preferred index in plant breeding, however it is in animal breeding. Once plan of genomic estimated breeding value (GEBV) was planned, it had been considered an unrealistic approach due to lack of enormous scale genotyping technologies. However, currently, it has been a possible approach with recent advances in high throughput genotyping platforms (3rd generation platforms). Generally processes of genomic selection and marker assisted selection used for Quantitative Traits are shown in **Figure 1**.

The main schemes of the two approaches are similar, wherever each marker assisted selection and genomic selection consist of breeding and training phases. In the training phase, phenotypes and genome-wide (GW) genotypes are investigated in an exceedingly set of a population, i.e., the mapping population in marker assisted selection and also the training population in genomic selection. Among populations, important relationships between phenotypes and genotypes are expected utilizing statistical models. Within the breeding phase, genotype data are obtained in an exceedingly breeding population, on the basis of genotypic information favorable individuals are selected. There are three prominent variations

http://dx.doi.org/10.5772/intechopen.76247

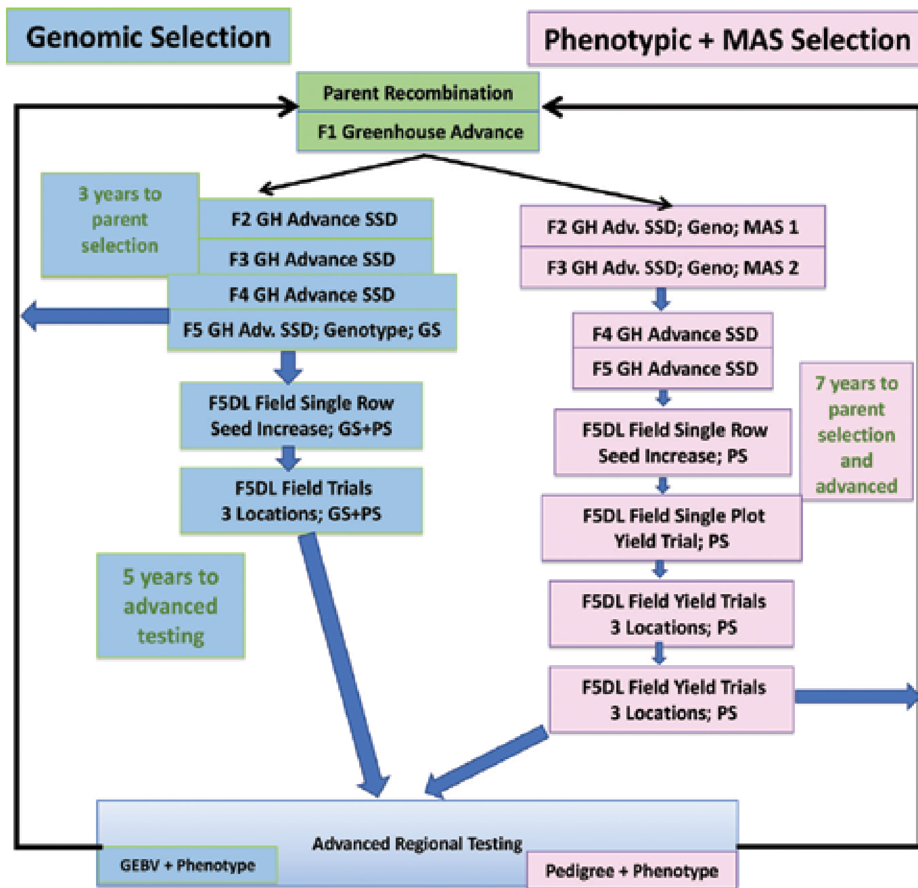


Figure 1. Self-pollinated crop genomic selection vs. phenotypic/MAS selection timeline.

between the two approaches: (1) within the training section, quantitative trait loci (QTLs) are known in marker assisted selection whereas formulae for genetic estimation of breeding value prediction are generated in genomic selection, called genomic selection models; (2) within the breeding section, genotype data are solely needed for targeted regions in marker assisted selection, whereas genomic selection genotype data are considered to be mandatory in genomic selection (3) within the breeding phase, favorable individuals are selected on the bases of the linked markers in marker assisted selection, whereas GEBVs are used for selection in GS. Thus, GS collectively analyses all the genetic variance of every individual by summing the marker impacts of GEBV and it is expected to deal with little effect genes that cannot be captured by traditional MAS.

The statistical ways employed by GS are comparatively new the plant-breeding community. The ways of marker-assisted selection (MAS) or marker-assisted recurrent selection (MARS) assume that the user is aware of that alleles are favorable, and what their average effects on the phenotype are. This assumption is viable for major-gene traits however not for quantitative traits that are influenced by several loci of little impact and the environment. To deal with

quantitative traits, new statistical approaches that might account for this uncertainty were required to get the most effective predictions potential. Finding problem with locus identification, entailed that the consequences for all marker loci be at the same time estimated. Once a prediction based on allele effects, the allele becomes the unit of analysis. Alleles are so the units that need to be replicated inside and across environments. However that replication will occur in spite of the particular lines carrying the alleles such lines themselves no longer need to be replicated. Within the breeding context, removing the requirement for line replication opens the likelihood of dramatically increasing the amount of lines pushed through the pipeline of a breeding program, and successively of accelerating selection intensity.

2. Genomic selection scheme

Genomic selection is to assemble a training population for individuals for which both genotypes and phenotypes are available and use those data to create a statistical model that relates variation in observed genotypes marker loci to variation in the observed phenotypes of the individuals. Multiple generations of parents and progenies provided powerful training population than a single generation individuals and larger number of individual's generations and markers provide more powerful training population (TP). The statistical model obtained from genotype and phenotype is then applied to a prediction population comprised of individuals for which genotypes are available, but phenotypes are not. GS is based on similarity between the training population (TP) and breeding population (BP) in the LD between marker loci and trait loci. This similarity may exist because breeding population is selected from training population or descended from training population or because density of markers is so high that every trait locus is in disequilibrium with at least one marker locus across the entire population of the target species. The training population is genotyped and phenotyped to train the genomic selection (GS) prediction model. In Genomic selection main role of phenotyping is to calculate effect of markers & cross validation. Genotypic information from the breeding material is then fed into the model to calculate genome estimated breeding values (GEBV) for these lines (**Figure 2**).

2.1. Need of genomic selection

Traditional marker assisted selection, whereas helpful for merely transmitted traits controlled by few loci, loses effectiveness because the number of loci will increase. This is often true for individual quantitative traits or once multiple traits are below selection. Quantitative traits like grain yield, abiotic stress have verified hard to enhance with marker-assisted selection. The main limitations are (i) tiny population sizes and traditional statistical strategies that have inadequate power to find and accurately estimate effects of small-effect quantitative trait loci (QTL) and (ii) gene x gene interactions (epistasis) and (iii) genotype x environment interactions (G.E) that have restricted the exchangeability of quantitative trait loci result estimates across populations and environments. The Beavis effect is a statistical phenomenon in biology that refers to the overestimation of the effect size of quantitative trait loci (QTL) as a result of small sample sizes in QTL studies.

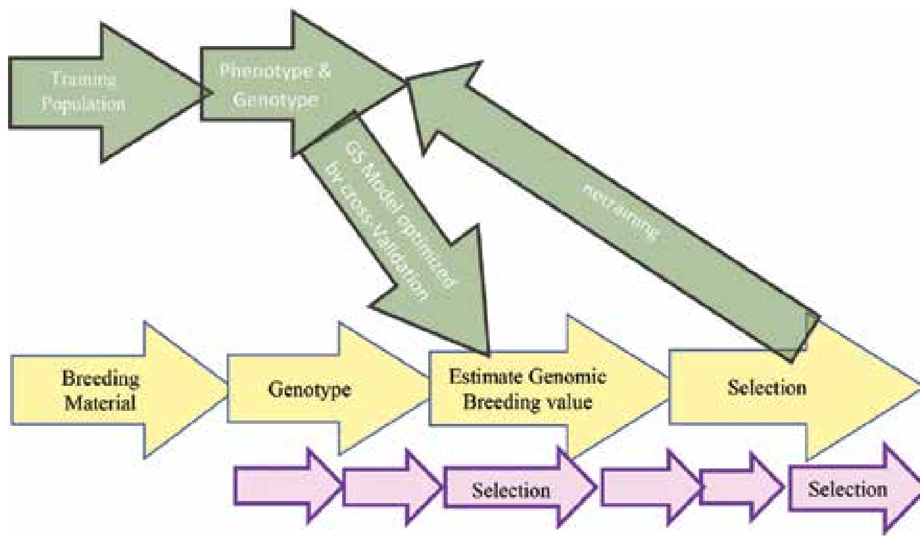


Figure 2. Genomic selection scheme. Information on phenotype and genotype for a training population allows estimating parameters for the model. (Modified: Castro et al. 2012) [12].

The availability of low cost and extensive molecular markers in plants has allowed breeders to raise however molecular markers might best be used to win breeding progress. Additionally advances in high-throughput genotyping have markedly reduced the value per data point of molecular markers and increasing genome coverage. This reduction was in the main the results of three parallel developments [2] (i) the invention of huge numbers of single nucleotide polymorphism (SNP) markers in several species; (ii) development of high-throughput technologies, like multiplexing and gel-free deoxyribonucleic acid arrays, for screening SNP polymorphisms; and (iii) automation of the marker-genotyping method, together with efficient procedures for deoxyribonucleic acid extraction [2]. Phenotyping prices are increased Genotyping prices are being reduced and marker densities are being increased speedily.

Statistical strategies are inadequate for improving polygenic traits controlled by several loci of small impact. There will be more markers (explanatory variables) than lines (observations) that introduce statistical issues. Drawback of small p (number of traits) and enormous m (number of markers) ends up in a lack of degrees of freedom. The foremost acceptable statistical model is required to at the same time estimate several marker effects from a limited range of phenotypes. In so-called “large p, small m” problems, standard multiple linear regression cannot be used without variable selection, that conflicts with the initial goal of avoiding marker selection. To overcome these issues, a range of ways, e.g., best linear unbiased prediction, ridge regression, Bayesian regression, kernel regression and machine learning methods are projected to develop prediction models for genomic selection.

The most economical use of GS is to exchange expensive and long phenotyping by a prediction of the genetic worth of the character below selection (or any multi trait index). Thus, the foremost expected advantage is to shorten selection cycles. However, to learn from shorter

cycles, the genetic gain per selection cycle ought to be near that predicted from phenotypic or combined MAS + phenotypic selection. Progeny testing schemes have a high accuracy of selection, however the time interval is also additional, takes long term to perform a cycle of selection that decreases the genetic gain. The univariate breeder's equation was used for the GS-BPs as a result of they include just one stage of selection [3]. Selection accuracy is adequate to the correlation between selection criteria and breeding value (i.e., correlation between phenotypes or GEBVs and true breeding values [TBVs]). In oxen, Schaeffer [4] determined that the time and value savings exploitation GS with GEBV accuracy of 0.75 would increase genetic gain twofold and supply a price savings of ninety two in comparison to the present ways.

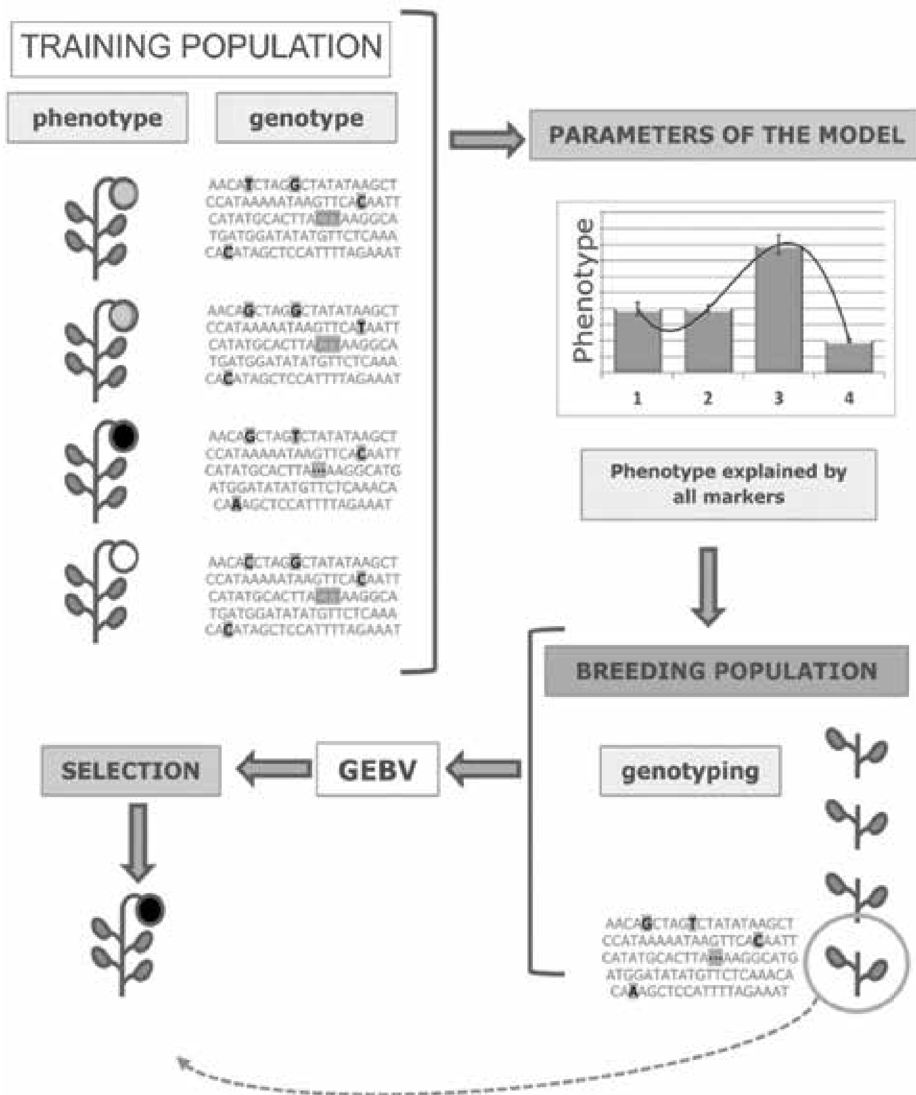


Figure 3. Genomic selection scheme.

The power to calculate extremely correct GEBVs and also the potential to drastically cut back makeup analysis frequency and selection cycle time expedited a speedy adoption of genomic selection and is revolutionizing the oxen breeding trade (**Figure 3**).

2.2. Model for genomic selection

The basic model may be denoted as

$$Y_i = g(x_i) + e_i \quad (1)$$

where Y_i is an observed phenotype of individual i (i = 1 … n) and x_i is a *1 x p* vector of SNP genotypes on individual *i*, *g(xi)* is a function relating genotypes to phenotypes, and e_i a residual term. The GEBV is generally equal to g (x_i). Further similarities among GS models can be seen by recognizing that they all seek to minimize a certain cost function. In least squares analysis, the well-known cost function is simply the sum of squared residuals.

2.3. Cross-validation

Evaluating GEBV accuracy through cross validation (CV). CV entails splitting the data into training and validation set. The ratio of observations in each set varies, but often a fivefold CV is used, that is, the data set is randomly divided into five sets, with four sets being combined to form the training set and the remaining set designated as the validation set. Each subset of the data is used as the validation set once, before applying of the prediction model to the breeding population, the accuracy of the model should be tested. For this, most of the training population is used to create a prediction model, which is then used to estimate the genomic estimation breeding values of the remaining individuals in the training population, using genotypic data only. This permits researchers to "test" and refine the prediction model to make sure the prediction accuracy is high enough that future predictions are often relied upon. Once valid, the model is often applied to a breeding population to calculate GEBVs of lines that genotypical, however no phenotypical, information is available.

2.4. Genomic selection prediction accuracies

The prediction accuracy of the GEBVs is evaluated by the correlation between the GEBVs and empirically estimated breeding values, r(GEBV: EBV), where the EBV can be obtained in a number of ways, most simply, as a phenotypic mean. This correlation provides an estimate of selection accuracy and thus directly relates GEBV prediction accuracy to selection response [2]. Other statistics such as mean-square error (MSE) are used occasionally [3]. Genomic selection accuracy is defined as the correlation between GEBV and the true breeding value (TBV), that is, r(GEBV:TBV). Since we can only measure r(GEBV:EBV), this measure needs to be converted to an estimate of r(GEBV:TBV). To do so, it is assumed that r(GEBV:EBV) = r(GEBV:TBV) X r(EBV:TBV). This assumption is correct if the only component common between the GEBV and the EBV is the TBV itself. In other words, the assumption holds if GEBV = TBV + e1 and EBV = TBV + e2, where e1 and e2 are uncorrelated residuals. The assumption could be violated if the training and validation data were collected in the same

environment. In that case, genotype by environment (G.E) interaction would generate a common component of error in both GEBV and EBV, biasing their correlation upward. Thus, training and validation data should be collected in different environments to ensure sound estimates of GEBV prediction accuracy. The correction, r(EBV:TBV), accounts for the fact that the EBV in the validation set is not free of error. When the EBVs are phenotypes, r(EBV:TBV) is equal to the square root of heritability (h) within the validation set [2].

Accurate GEBV predictions offer the possibility that future elite and parental lines will be selected on GEBV rather than phenoypic data from extensive field testing. Immediate impact would be a great increase in speed of breeding cycle increasing selection gains per unit time. Thus, GS could radically change the practice of field evaluation for breeders. Of course, regardless of the breeding method used, final field evaluations of varieties across the target environments will be needed before they are distributed to farmers.

Breeding cycle time is shortened by removing phenotypic evaluation of lines before selection as parents for the next cycle. Model training and line development cycle length will be crop and breeding program specific. In a GS breeding schema, genome-wide DNA markers are used to predict which individuals in a breeding population are most valuable as parents of the next generation of offspring. The prediction model is additionally continuously rejuvenated as genotypical and phenotypic data from elite lines derived from the collaborating breeding programs is incorporated into the prediction models. In this manner, new germplasm may be infused into the system at any point. As lines derived from the recently infused germplasm advance within the breeding process, their genotypical and phenotypic data may be incorporated into the prediction models.

The purpose of phenotyping now is to pick the best lines from a segregating population and to judge fewer lines with larger replication in every cycle of selection. However during a GS driven breeding cycle, the aim of phenotyping is to estimate or re-estimate marker effects. It is far from clear at this point whether or not it will be advantageous to evaluate solely the best lines or to evaluate few lines with high replication. So separates the germplasm improvement cycle from the prediction model improvement cycle. Indeed, if we tend to use the rules for optimum QTL linkage mapping, analysis should include not just the best however the best and the the} worst lines **Figure 3** also emphasizes the requirement for model updating and re-evaluation. Marker effects might amendment as a results of allomorph frequency changes [6] or of epistatic gene action. Model updating with every breeding cycle should mitigate reduced gains from GS caused by these mechanisms. Thus, GS may radically change the practice of field evaluation for breeders. Of course, despite the breeding technique used, final field evaluations of varieties across the target environments are going to be needed before they are distributed to farmers.

Accuracy declines as generation number between the last model update and selection candidates increases [4–6], because selection causes changes in variances, allele frequencies, and LD relationships between markers and QTL [4]. Under random mating, simulations have shown model accuracy to decrease by about 5% per generation [5, 6], but accuracy decrease was much more rapid under selection.

http://dx.doi.org/10.5772/intechopen.76247

3. Factors responsible for the estimation accuracy of GS models

The response of genomic selection is that the output of varied factors responsible for estimation accuracy of GEBVs. These factors are interconnected in an exceedingly advanced and comprehensive manner. They include model performances, sample size and relatedness, marker density, gene effects, heritability and genetic design, and therefore the extent and distribution of linkage disequilibrium between markers and QTL.

3.1. Population size

The most important characteristic of the population is its effective size. An obvious measure of population size is its census: how many individuals it contains. But populations with the same census size can behave quite differently. For a population of a given rate of inbreeding, the effective size is equal to the census size of a randomly mating ("ideal") population that would have that same rate.

Accuracy due to genetic relationships can represent from a small minority to a large majority proportion of the overall accuracy. The combination of long-distance LD due to pedigree relatedness (e.g., full sibs and half sibs) and short-distance ancestral LD due to small effective population size are among the key features of our training population. With improved marker technology, large TPs that use a representative sample of germplasm in a given breeding program may be a good strategy for long-term accuracy over a broad range of families. It has been observed to be monotonic increase in the prediction accuracy for grain yield with increasing population size without any substantial decrease in the slope (**Figure 4**). Studies in this the size of the training population is of crucial importance in genomic selection. The impact of the population size on the accuracy of genomic selection is less pronounced for fewer characters like grain moisture, which might be due to presence of larger variance among populations

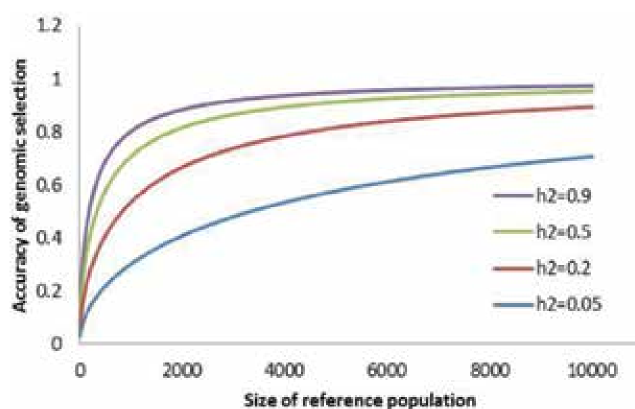


Figure 4. Relation between number of plants in the training population and accuracy of genomic selection for traits with different heritabilities.

that can be efficiently utilized by few individuals per population. Parameters such as effective population size and QTL number strongly influence marker densities and TP sizes required for acceptable accuracy. Indeed, simulations similar to those of Meuwissen et al. [6] have shown that marker density needs to scale with effective population size [7]. Until very low marker densities were reached, marker number had very little, if any, effect on prediction accuracies within families from various plant species [8]. Likewise, GEBV accuracy of several traits in cattle, including net merit, was hardly affected when as many as 75% of the original markers were masked. Adequate marker density and TP size depend on QTL number and trait heritability. Calus and Veerkamp [9] used the average r^2 between adjacent markers as a measure of marker density relative to decay of linkage disequilibrium. They found that for a highly heritability trait, average adjacent marker r^2 of 0.15 was sufficient, but for a low heritability trait, increasing the r^2 to 0.20 improved prediction accuracy. Heritability dramatically affects TP sizes required for successful GS, especially at h^2 less than 0.40 [3].

3.2. Linkage disequilibrium

Genetic drift is an important cause in generating LD, the non-independence of alleles at different loci. This non-independence allows marker alleles to predict the allelic state of nearby QTL, enabling marker genotypes to predict the phenotype. The LD intensity decays with greater distance between two markers. Decay rates which vary widely across species, populations, and genomes due to forces of mutation, recombination, population size, population mating marker density must increase with increases in Ne*c, where Ne is the effective population size and c is the recombination rate between loci. LD patterns. Marker density can be inferred by the rate of LD decay across the genome as inferred by the relationship b/w inter marker coefficient of determination r^2 and genetic distance. LD estimates can be used to determine target marker densities for GS at equilibrium, drift generating LD is balanced by recombination, causing it to decay, such that nearby loci are expected to be in higher LD than faraway loci. LD has a major effect on the operability of GS, so it has to be well understood before performing GS. LD is defined as the non-random association of alleles at different loci. It has been found that for high heritability trait average adjacent marker r^2 of 0.15 is sufficient but for low heritability trait increasing r^2 value to 0.2 improve accuracy of GEBV predictions.

4. Types of marker platforms

Since, then marker of choice is very important to accurate estimate GEBVBs, different platforms are available

4.1. SNP chips

Single nucleotide polymorphisms (SNPs) differentiate individuals based on variations detected at the level of a single nucleotide base in the genome. SNPs have become the marker of choice for crop genetics and breeding applications because of their high abundance in genomes, and the availability of a wide array of genotyping platforms with various multiplex capabilities for

SNP analysis [10]. Recent breakthroughs in next generation sequencing (NGS) technologies enabled millions of sequences reads to be generated from a single run at a more affordable cost. The ability to perform GS requires routine genotyping at a high number of loci. Single nucleotide polymorphisms (SNPs) differentiate individuals based on variations detected at the level of a single nucleotide base in the genome. SNPs have become the marker of choice for crop genetics and breeding applications because of their high abundance in genomes, and the availability of a wide array of genotyping platforms with various multiplex capabilities for SNP analysis [10]. Recent breakthroughs in next generation sequencing (NGS) technologies enabled millions of sequences reads to be generated from a single run at a more affordable cost. The resulting large amount of data provided sequence depth adequate for de novo sequence assembly, which has made the SNP discovery on a large scale a feasible task, particularly for species without completed genome sequences. Successful results on large-scale discovery of SNPs based on NGS methods have been reported in several plant species, including both and polyploid species, and more are on the way. The development of highly parallel SNP assays, such as Illumina's Golden Gate assay with 1536-plex platform enabled the genome-wide studies previously not feasible for economically important crops. Using these techniques, SNP-based high-density genetic maps are now available in several crop plants such as soybean, maize, barley and wheat. Thus, genotyping lines for use in GS using SNP and direct resequencing with next-generation.

4.2. Genotyping by sequencing (GBS)

Advances in next generation technologies have driven the costs of DNA sequencing down to the point that genotyping-by sequencing (GBS) is now feasible for high diversity, large genome species. GBS is a highly multiplexed approach is based on high-throughput, next-generation sequencing of genomic subsets targeted by restriction enzymes (REs). Genotyping by sequencing (GBS) in any large genome species requires reduction of genome complexity. Genotyping-by-sequencing can be applied to different populations or even different species without any prior genomic knowledge as marker discovery is simultaneous with the genotyping of the population. GBS sequence allows access to any sequence within low copy genomic regions or regulatory regions controlling the expression of plant genes responsible for agronomically important phenotypes are often located in non-coding DNA. The use of GBS for GS, therefore, should be applicable to a range of model and non-model crop species to implement genomics-assisted breeding. Genotyping-by sequencing combines marker discovery and genotyping of large populations, making it a superb marker platform for breeding applications even within the absence of a reference genome sequence or previous polymorphism discovery. Additionally, the pliability and low price of GBS create this an ideal approach for genomics-assisted breeding.

5. Advantages of genomic selection

1. The marker effects are calculable from the training population and used directly for GS within the involved breeding population, and QTL discovery, mapping, etc., are not needed.

2. Each simulation and empirical studies reveal that GS produces larger gains per unit time than constitution selection. For instance, a simulation study in maize showed GS to be superior to MARS, notably for traits having low heritability. Further, GS will predict the performance of breeding lines additional accurately than that supported pedigree data, and GS appears to be an efficient tool for rising the potency of rice breeding.

3. The selection index approach integrates appropriately weighted data from multiple traits into an index that is the premise for concurrent selection for the concerned traits. The genome-wide marker data is integrated into a range index either alone or in conjunction with phenotype data on one or additional traits. Simulation studies show that the on top of combined selection index approach of GS can increase the effectiveness of selection, considerably for low heritability traits.

4. GS would tend to cut back the speed of inbreeding and also the loss of genetic variability as compared to selection based on breeding values calculable from phenotype data; this may be achieved while not sacrificing selection gains. This might be notably vital in species that show severe inbreeding depression.

5. Genomic selection scheme consist of phenotyping for each selection cycle within the breeding population is not needed. This greatly reduces the length of breeding cycle, notably in perennial species. For instance, GS was calculable to reduce the selection cycle time from 19 years to simply 6 years just in case of oil palm (*Elaeis guineensis*). Further, GS was calculable to outperform MARS and phenotypic selection even with a population size of fifty once selection gain was considered on per unit time and price, however not on per selection cycle, basis. (The selection the choice) cycle is reduced as a result of GS does not need analysis of interbreeding performance of the plants being subjected to selection that is critical within the case of phenotypic selection. In perennial species, GS is anticipated to facilitate commercialization of improved genotypes at abundant shorter intervals of time than phenotypic selection.

6. Genomic selection would possibly enable breeders to pick out parents for crossbreeding programs from among those lines that have not been evaluated within the target environment. This selection would be supported GEBVs of these lines estimated for their adaptation to the target environment. This could facilitate germplasm exchange and their utilization in breeding programs.

7. Genotype X environment interaction could be a vital a part of phenotype and its estimation is sort of demanding. GS can utilize information on marker genotype and trait phenotype accumulated over time in varied analysis programs covering a variety of environments and integrate an identical in GEBV estimates of the various individuals/lines. This could enable GEBV estimation even for traits that they have never been tested.

8. Theoretically, GEBV estimates is employed for the selection of parents for crossing programs and, possibly, for the development of hybrid varieties. These applications, however, ought to expect validation of the concept in apply.

http://dx.doi.org/10.5772/intechopen.76247

6. Limitations of genomic selection

1. GS has still not become popular plant breeding community primarily due to low evidence for its sensible utility. In fact, most discussions on its utility are for the most part statistical treatments and simulations that are not simply appreciated by plant breeders.

2. The potential value of GS should be assessed with caution because GS has been mostly evaluated in simulation studies. There is an imperative have to be compelled to judge genomic selection in crop breeding situations to demonstrate its practical utility.

3. The marker effects and, as a result, GEBV estimates would possibly modification attributable to changes in gene frequencies and epistatic interactions. This is often ready to necessitate amendment of the GS model with every breeding cycle therefore the gains from GS are not reduced.

4. The accuracy of GEBV estimates has been evaluated exploitation simulation models based on additive genetic variance. These models ignore epistatic effects that does not seem to be realistic. It has been argued that since biological process makes alone a small contribution to the breeding value, the employment of solely additive genetic models for GS is additionally expected to maximize selection gains. However, this argument are planning to be entirely valid only for self-fertilizing species, where homozygous lines are used as parents as well as varieties.

5. Our information concerning the genetic design of quantitative traits is severely restricted, that limits our ability to develop applicable models of GS to realize the most prediction accuracy.

6. The selection response declines at a faster rate under GS than with pedigree selection. This may be reduced by continually together with new markers for the prediction of GEBVs. The long response under GS can also be raised by putting higher weights on the low-frequency favorable alleles, considerably within the start of GS program.GS is simpler than phenotypic selection on per unit time basis only if off-season/greenhouse facilities are accustomed grow up to three generations per annum. The utility and also the cost-effectiveness of GS would be uncertain wherever such facilities are not offered.

7. The necessity for genotyping of an oversized variety of marker loci in every generation of selection adds considerably to the price of breeding programs. It has been projected that, inside the future, a bigger stress are going to be placed on the use of marker information than on composition information. Such a shift, however, would need the value of one marker information to be merely 1/5000 the price of phenotyping one entry.

8. Implementation of GS would need intensive infrastructure and completely different resources, which might get on the so much aspect the reach of most moderate size public sector breeding programs, considerably within the developing countries. To boot, planning and execution of GS is kind of exigent, and additionally the breeders would be required to reorient their approach to the breeding programs.

7. Conclusions

Currently, the lion's share of research on GS has been performed in livestock breeding, where effective population size, extent of LD, breeding objectives, experimental design, and other characteristics of populations and breeding programs are quite different from those of crop species. Nevertheless, a great number of findings within this literature are very illuminating for GS in crops and should be studied and built upon by crop geneticists and breeders. The application of powerful, relatively new statistical methods to the problem of high dimensional marker data for GS has been nearly as important to the development of GS as the creation of high-density marker platforms and greater computing power. The methods can be classified by what type of genetic architecture they try to capture. Somewhat surprisingly, RR-BLUP, which makes the ostensibly unrealistic assumption that genetic effects are uniformly spread across the genome, often performs as well as more sophisticated models. Exceptions do exist, though, and there is abundant evidence that BayesB is superior for traits with strong QTL effects. Additionally, since BayesB better identifies markers in strong LD with QTL than RR-BLUP, it maintains accuracy for more generations. Finally, the question of whether or not to model epistasis remains open. If epistasis is important for a particular trait in a particular population, the kernel methods and machine-learning techniques such as SVM may be preferable. It is important for the practitioner to consider such issues or test methods on a relevant data set before a method for GEBV calculation is chosen. Although the increasing marker density, training population size, and trait heritability are obvious ways to improve GEBV accuracy; these options add cost to the program. Implementing algorithms for markers imputation and training population design holds the potential for essentially free additional accuracy, leading to greater overall GS efficiency.

The current drops in genotyping costs, while phenotyping costs remain constant or increase, suggest that efforts to understand how to choose which lines to phenotype on the basis of their genotype, that is, how to design training populations, will be rewarding. Combining training populations from different populations is another way to boost accuracy when individual populations lack sufficient size and assuming that the marker densities required are available. With respect to maximizing long-term selection, we discussed several promising approaches that strive to retain favorable, low-frequency alleles while minimizing loss of short-term gain. Both simulation and empirical results for GS have been quite impressive. Empirical results of GS accuracy in crops, however, are not yet available for the public sector, except in the form of CV within families. Further empirical studies of the effects of statistical models, marker density, TP size and composition, and different selection criteria for the effectiveness of GS in breeding populations are urgently needed. In addition, while the CV approach can be instructive, an important caveat should be mentioned. In CV, the training and validation sets belong to the same population. But in GS, the selected candidates will rarely belong to the same population as the training set and may well be several generations removed from it. Recombination during meiosis between generations erodes the association between marker and QTL, systematically reducing accuracy. The effect of selection on allele frequencies and the Bulmer effect can also have detrimental effects on accuracy. In order to realistically evaluate GS for crops, studies designed for this purpose should be performed.

http://dx.doi.org/10.5772/intechopen.76247

Clearly, exciting times are ahead of us as public breeding programs launch GS efforts. This review compiles several immediately useful results for breeders wanting to maximize gains through GS. Knowledge of breeding program parameters (effective population size, extent of LD, and trait heritability) allows marker density and training population size to be determined using analytical formulae. The greatest impact of GS on gain per unit time will come from shortening the breeding cycle [11]. Therefore, redesigning crossing and population development schemes incorporate GS as early as possible will likely be the most effective. Consequently, phenotyping resources will need to be shifted from early generation, evaluation for selection to evaluation for model training. The importance of epistasis will need to be assessed for each trait. A major paradigm in plant breeding since the availability of molecular marker technology is that mapping and characterizing the genetic loci that control a trait will lead to improved breeding. Often, one of the rationales for cloning of QTL is to develop the "perfect market" for MAS, perhaps based on a functional polymorphism. In contrast, an advantage of the GS is precisely its black box approach to exploiting genotyping technology to expedite genetic progress. This is an advantage in our view because it does not rely on a "breeding by design" engineering approach to cultivar development requiring knowledge of biological function before the creation of phenotypes. Breeders can therefore use GS without the large upfront cost of obtaining that knowledge. In addition, GS can maintain the creative nature of phenotypic selection which couples random mutation and recombination to sometimes arrive at solutions outside the engineer's scope.

Acknowledgements

Support from the teaching section Department of plant breeding and Genetics, Punjab Agricultural University, Ludhiana India.

Nomenclature

MAS	marker-assisted selection
GS	genomic selection
BV	breeding value
GEBV	genomic estimated breeding value
CV	cross-validation
TP	training population
BP	Breeding population
LD	linkage disequilibrium

Author details

Mohd Shamshad* and Achla Sharma

*Address all correspondence to: shamshad.ratan@gmail.com

Punjab Agricultural University, Ludhiana, India

References

[1] Coors JG. Who are plant breeders, what do they do, and why? Plant Breed Arnel R Hallauer International Symposium. 2008:51-60

[2] Sun CQ, Huang ZZ, Wang YL, et al. Overcoming pre-fertilization barriers in the wide cross between *Chrysanthemum grandiflorum* (Ramat.) *Kitamura* and *C. Nankingense* (Nakai) Tzvel. By using special pollination techniques. Euphytica. 2011;**178**:195-202

[3] Hayes BJ, Bowman PJ, Chamberlain AC, et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics, Selection, Evolution. 2009;**41**:1-9

[4] Muir WM. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. Journal of Animal Breeding and Genetics. 2007;**124**:342-355

[5] Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;**177**:2389-2397

[6] Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;**157**:1819-1829

[7] Solberg TR, Sonesson AK, Woolliams JA, et al. Genomic selection using different marker types and densities. Journal of Animal Science. 2008;**86**:2447-2454

[8] Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics. 2009;**120**:151-161

[9] Calus MPL, Veerkamp RF. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. Journal of Animal Breeding and Genetics. 2007;**124**:362-368

[10] Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology. 2002;**5**:94-100

[11] Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. Crop Science. 2009;**49**:1-12

[12] Perez-de-Castro AM, Vilanova S, Canizares J, et al. Application of genomic tools in plant breeding. Current Genomics. 2012;**13**:179-195

Section 2

RNA-seq and Gene Editing in Plants

Chapter 6

Role of Next-Generation RNA-Seq Data in Discovery and Characterization of Long Non-Coding RNA in Plants

Shivi Tyagi, Alok Sharma and
Santosh Kumar Upadhyay

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72773

Abstract

The next-generation sequencing (NGS) technologies embrace advance sequencing technologies that can generate high-throughput RNA-seq data to delve into all the possible aspects of the transcriptome. It involves short-read sequencing approaches like 454, illumina, SOLiD and Ion Torrent, and more advance single-molecule long-read sequencing approaches including PacBio and nano-pore sequencing. Together with the help of computational approaches, these technologies are revealing the necessity of complex non-coding part of the genome, once dubbed as "junk DNA." The ease in availability of high-throughput RNA-seq data has allowed the genome-wide identification of long non-coding RNA (lncRNA). The high-confidence lncRNAs can be filtered from the set of whole RNA-seq data using the computational pipeline. These can be categorized into intergenic, intronic, sense, antisense, and bidirectional lncRNAs with respect to their genomic localization. The transcription of lncRNAs in plants is carried out by plant-specific RNA polymerase IV and V in addition to RNA polymerase II and target the epigenetic regulation through RNA-directed DNA methylation (RdDM). lncRNAs regulate the gene expression through a variety of mechanism including target mimicry, histone modification, chromosome looping, etc. The differential expression pattern of lncRNA during developmental processes and different stress responses indicated their diverse role in plants.

Keywords: next-generation sequencing (NGS), high-throughput RNA-seq, long non-coding RNA (lncRNA), expression, development, stress

1. Introduction

Next-generation sequencing (NGS) technologies provide a new platform for the production of high-throughput sequencing data in less time at reduced cost. The tremendous

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (cc) BY

improvements in past years have allowed the sequencing of millions of DNA fragments in parallel. It has shifted the genomics to a newer edge by capturing the small details of DNA fragments. Earlier, Maxam and Gilbert's [1] and Sanger sequencing [2] techniques were leading approaches after the discovery of the DNA structure [3]. However, these techniques were time-consuming and limited to small-scale, dealing with few genes to the genome of simple organisms. But the necessity of sequencing the complex genome in short time and reduced cost have technologically advanced the sequencing approaches and evolved as NGS technologies. The NGS systems provide rapid, reproducible, and highly accurate sequencing techniques, and are based on the short-read sequencing approaches and a more advance single-molecule long-read sequencing [4]. The short-read sequencing approaches are dependent on sequencing by synthesis (SBS) and sequencing by ligation (SBL) methods. Further, these methods require pre-processing of DNA before directly proceeding to the sequencing steps, according to the requirement of different NGS platforms [4]. In SBS approach, the nucleotides are added by the polymerase into the elongating DNA strand and the signal is received in the form of fluorescence or ionic concentration change for every single nucleotide incorporated [5, 6]. Besides this, in SBL approach, probes having one- or two-base matching, bound to fluorophore, are ligated to the adjacent oligonucleotide on DNA fragments. The emitted fluorescent spectrum identifies the complementary bases of the probe at a specific position and reset primers are used to encrypt the complete DNA sequence [5]. Most of the short-read sequencing approaches require the clonal amplification of DNA on the solid surface such as bead-based, solid-state, and generation of DNA nanoball [5]. In all the methods, initially the DNA is fragmented and then ligated to a common set of adaptor for amplification and consequently ensue for DNA sequencing [5]. The short-read sequencing approaches include 454, illumina, SOLiD, and Ion Torrent platforms. Moreover, the *in-silico* approaches are used for the assembly of data generated by after these techniques [6]. The limitations in short-read sequencing approaches like *de novo* sequencing and the resolution of genomic variation leads to the development of more advance long-read sequencing approaches [6]. The long-read sequencing approaches are used for complex genomes with several long repetitive elements, structural variation, and alteration in copy number, which are significant for the occurrence of disease, and for evolution and adaptation [7–9]. It produces long reads of several kilobases and allows the higher resolution of the genome. In contrast to short reads, a single long-read can completely span the repetitive or complex region of genome, thus reducing the probability of vagueness in the size and positions of the genomic element [6]. Pacific Biosciences and Oxford Nanopore are commercially available sequencing technologies which provide the platform for sequencing the long reads with thousands of bases per read. These technologies are based on single-molecule sequencing, but have different methods of nucleotide detection. Oxford Nanopore is based on the detection through nanopores while Pacific Biosciences uses optical detection inside zero-mode waveguide [10]. Besides this, in synthetic approach, the data of short-read sequencing is combined with informatics and biochemical approaches for the construction of synthetic long reads. Long reads allow researchers for a deep transcriptomic study such as allele-specific transcription, alternative splicing, and in the identification of exact connectivity of exons and discern gene isoforms [6, 11, 12].

2. High-throughput RNA sequencing

Transcriptome consists of a whole set of transcripts present in a cell, and their expression level in particular developmental stage and cellular conditions. The detailed study of an organism at transcriptome level is necessary for revealing the molecular constituents involved in that particular stage or condition of the tissue. The high-throughput RNA-sequencing (RNA-seq) has emerged as an important technique in the field of transcriptomics for studying all the aspects of gene expression at large scale. It is one of the most commonly used techniques for quantification and mapping of transcriptomes. It involves the conversion of RNA into cDNA, followed by random sequencing of cDNAs fragments by using NGS platforms [13]. The generated millions of short reads were assembled by various bioinformatics approaches. Consequent mapping of these short reads reveals the position of gene transcribing the RNA on the reference genome or sets of a gene [13]. The high-throughput technologies also include direct RNA sequencing (DRS), in which the native RNA is directly sequenced without proceeding to the step of cDNA preparation. The technique is successful in sequencing native polyA+, where reverse transcription is undesirable. It is applicable in determining the precise sequence, identification of alternative polyadenylation sites, and deals with the small amount of nucleic acid [14]. In cap-assisted gene expression (CAGE) technique, RNAs with a 5′ cap are targeted. The short sequence tags are generated from 5′ ends of targeted RNAs with one tag per RNA molecule and allow the precise mapping of 5′ ends [15]. Series analysis of gene expression (SAGE) is another method for the sequencing of RNA molecules which target polyadenylated messages, and tags are generated near 3′ ends, typically one internal tag per RNA molecule [16]. Similarly, paired-end tags (PET) also targets polyadenylated RNA molecules, but the combined information on 5′ and 3′ ends of same RNA molecule generates the sequence tag [17]. Furthermore, rapid amplification of cDNA ends (RACE) is a PCR-based method used to identify the unknown sequences in conjunction with a known region. Together with the NGS technologies, it can be utilized for deep transcriptome sequencing of the particular locus [18]. Targeted RNA sequencing is also meant for a specific locus and by using tiling microarrays RNAs are selected and sequenced [19]. RNA profiling method by GRO-seq measures the steady-state levels of RNA and combined NGS analysis with the nuclear run-on experiments to generate information on RNA polymerase complexes competent with transcription [20]. This high-throughput RNA-seq is helpful in finding out the transcript (messenger RNAs, non-coding RNAs, and small RNAs) of species in short time and in determining the 5′ and 3′ splice sites, splicing patterns, and post-transcriptional modifications. The quantification of transcripts reveals the change in expression of genes in different conditions.

3. Long non-coding RNA (lncRNA)

3.1. Discovery and identification of lncRNA

In the era of NGS, the high-throughput RNA-seq data has lime lighted the necessity of non-coding part of the genome in the gene functioning. Non-coding RNAs (ncRNAs) are

transcribed from non-coding DNA, earlier called junk DNA. An extensive study on transcriptomes from multiple species indicated that about 90% of the genome can be transcribed, whereas only a small portion of such transcribed regions potentially codes for proteins [21]. The ncRNAs are categorized into housekeeping and regulatory ncRNAs on the basis their expression and role in different cells types. The expression of housekeeping ncRNAs (e.g., t-RNA, r-RNA, and snRNA) is prominent and has a structural role in all the cells [22]. While, the regulatory ncRNA shows temporal expression in specific cell types and includes microR NAs (miRNAs), small interfering RNAs (siRNAs), enhancer RNAs (eRNAs), promoter-associated RNAs (PARs), Piwi-interacting RNAs (piRNAs), and long non-coding (lncRNA). The criteria of >200 nt length are set for the identification of lncRNAs among all the organisms [23]. lncRNA comprises of a major group of ncRNAs and regulate various biological processes through different molecular mechanisms.

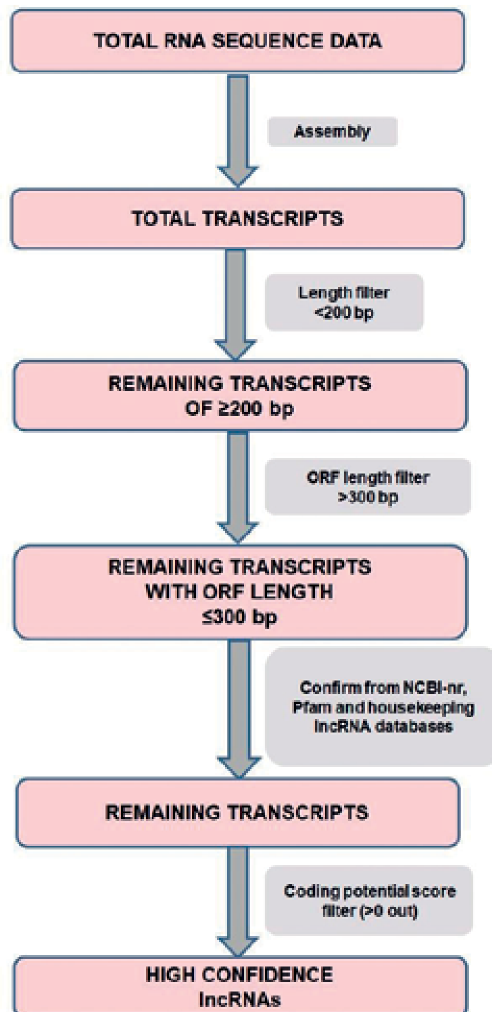


Figure 1. Pipeline for identification of long non-coding RNA.

In plants, the lncRNA was first reported in *Glycine max* [24], involved in changing the sub-cellular localization of a protein. In *Medicago truncatula* and *Oryza sativa*, *MtENOD40* and *OsENOD40* lncRNAs were discovered in nodule formation, respectively, and signify the involvement of lncRNA in biological roles [25, 26]. Likewise, in other plant species, for example, *COLD-INDUCED LONG-ANTISENSE INTRAGENIC RNA* (*COOLAIR*) and *COLD-ASSISTED INTRONIC NONCODING RNA* (*COLDAIR*), lncRNA in *Arabidopsis thaliana* [27, 28], involved in regulation of flowering, were identified and studied for their diverse function in the plant system. Furthermore, the exponential rise in high-throughput RNA-seq data have contributed to the discovery of lncRNA at genome-wide level, but the studies are limited in plants to some species. The amalgamation of experimental RNomics with the computational approaches has contributed to the identification of lncRNA and their function in wide-ranging biological processes [6]. The accurate identification and functional annotation is an ongoing challenge in

Sr. no.	Plant name	Number of lncRNAs	Tissues/organ/stress	Reference
1	*Amborella trichopoda*	2569	Tissue	[32]
2	*Arabidopsis thaliana*	~6480	Organ-specific and stress responsive	[22]
3	*Chlamydomonas reinhardtii*	2214	Cultured cells and synchronized vegetative cells	[32]
4	*Cicer arietinum*	2248	Three vegetative tissues and flower development	[30]
5	*Cucumis sativus*	3274	Fruit development and sex differentiation tissues	[33]
6	*Fragaria vesca*	1556	Floral, fruit tissue, and two vegetative tissues	[34]
7	*Medicago truncatula*	23,324	Control, osmatic, and salt stress in leaf and root tissues	[35]
8	*Oryza sativa*	2224	Development and reproductive organs	[36]
9	*Physcomitrella patens*	2711	Developmental stages	[32]
10	*Populus trichocarpa*	2542	Control and drought condition	[37]
11	*Setaria italica*	584	Drought stress	[38]
12	*Selaginella moellendorffii*	4422	Root, stem, and leaf	[32]
13	*Solanum lycopersicum*	10,774	In wild and ripening mutant	[39]
		1565	Tomato yellow-leaf curl virus stress	[40]
14	*Triticum aestivum*	44,698	Organ-specific and stress responsive	[31]
		283	Fungal-responsive lincRNAs	[41]
15	*Vitis vinifera*	4506	Organ-specific	[32]
16	*Zea mays*	1704	Different tissues	[42]
		664	Drought-stressed leaves	[43]
		7245	Leaves (under conditions of nitrogen deficiency and sufficiency)	[44]

Table 1. Occurrence of lncRNA in various plant species.

the field of bioinformatics for high-throughput RNA-seq data. The data of identified lncRNAs in plants is timely submitted to the different databases [29]. A pipeline with multiple filters has been designed for the assembly and identification of high confidence lncRNAs in **Figure 1** [30, 31]. The present status of most of the identified lncRNAs in different plant species are mentioned in **Table 1**.

3.2. Classification of lncRNAs

The biotypes of lncRNAs were identified with respect to their genomic localization, and were mainly categorized into intergenic, intronic, sense, antisense, and bidirectional lncRNAs. As the term suggest, the intergenic lncRNA are transcribed from the region amid two genes, while introns are the source of intronic lncRNA [45]. The sense and antisense lncRNAs are derived from overlapping region of exons on the sense and antisense strands, respectively [18], when the transcription of lncRNA is initiated in the juxtaposition of adjacent mRNA on complementary strand, termed as bidirectional lncRNA [45].

4. Molecular mechanisms of the functioning of lncRNAs

The dramatic change, in the past years about the knowledge of lncRNA in gene regulation mechanisms, has exponentially raised with high-throughput RNA-seq data. In plants, the studies are limited to small scale in comparison to animals, but the available reports suggested their different mechanisms as following.

4.1. lncRNA as target mimics of miRNA

Target mimicry is a mechanism of lncRNA for regulating the functions of miRNAs. They inhibit the interaction between the miRNA and their respective targets by binding to the target of miRNA via partial complementary sequence [46]. The novel mechanism of target mimicry was first discovered in *Arabidopsis*. In addition, *phosphate Starvation 1* (*IPS1*) was the first lncRNA identified as endogenous target mimic (eTM) of miR399 involved in phosphate homeostasis [46]. During phosphate starvation, the expression of miR399 is induced in companion cells and phloem [47]. Subsequently, the expression of *PHO2* gene, a target of miR399, is repressed [47–50]. This gene encodes UBC24 (E2 ubiquitin conjugase-related enzyme) and the reduction in its expression leads to the increased expression of *Pht1;8* and *Pht1;9* (phosphate transporter genes) in roots [47, 48]. Later, a similar mechanism was discovered in animals and humans suggesting target mimicry as the prevalent phenomenon [51, 52].

4.2. Histone modification

The lncRNAs are known to regulate gene expression through epigenetic changes. These epigenetic changes may result in alteration of gene expression in plants. Vernalization is the most common phenomenon of lncRNA mediated epigenetic regulation in plants. In *Arabidopsis*, FLOWERING LOCUS C (FLC) gene is the principal regulator of vernalization process and regulates the flowering time [53]. The expression of this gene is regulated by *COOLAIR* and *COLDAIR* lncRNAs through histone modifications [54].

4.3. Precursor lncRNA

lncRNAs constitute an important class of riboregulators by acting as a precursor in the synthesis of shorter ncRNAs, such as miRNAs and siRNAs. In this mechanism, some lncRNAs are processed to shorter ncRNAs or may directly act as a precursor [55]. The genes of primary miRNA transcripts (pri-miRNA) encoding miRNAs are transcribed by RNA polymerase II [56]. In plants, miRNA constitutes the modest portion in small regulatory ncRNA pool due to the presence of other complex small regulatory ncRNAs. In addition, they have plant-specific RNA polymerase IV/V involved in the transcription of siRNAs and endogenous siRNAs [57]. For example, in *Triticum aestivum*, 19 lncRNAs were predicted as a precursor of 28 miRNAs [31]. In *Arabidopsis*, the 24-nt sequence of several siRNAs were matched with five lncRNAs (npc34, npc351, npc375, npc520, and npc523), which was considered as potential precursor lncRNAs. The mapping of siRNAs on both the strands of lncRNAs also strengthened the findings [58].

4.4. RNA-dependent DNA methylation

The modification of chromatin is facilitated by recruitment of chromatin modifiers through lncRNA and small RNA (sRNA) into the specific locations in DNA. This RNA-dependent DNA methylation (RdDM) is a conserved process that recruits DNA methyltransferase and histone modifiers for DNA methylation and suppressive histone modification, respectively [59].

4.5. Chromosome looping

This mechanism is different from RdDM and histone modification, as it only involves the structural changes of chromatin. Thereby, it affects the binding potential of RNA polymerase and other transcription factors [60]. A persuasive example of chromosome looping mechanism in plants by *APOLO* lncRNA has been described in auxin transport by regulating the PID expression, an auxin transporter. When locus of *APOLO* lncRNA is transcribed by RNA Pol V and modified by RdDM, the expression of the locus is suppressed and loops to PID. This causes the inhibition of PID transcription. In contrast, when RNA Pol II carry out the transcription of APOLO lncRNA the looping of PID is restrained resulting in the expression of PID [60].

4.6. Protein re-localization

The mechanism of lncRNA in protein relocalization was first described in *G. max* and *Medicago sativa* [61, 62]. The symbiotic interactions among soil bacteria and leguminous plants are regulated by *Enod40* gene (early nodulin gene) which is induced by nitrogen-fixing bacteria in the pericycle and dividing cortical cells of roots [24, 63]. The diverse occurrence of *Enod40* lncRNA was suggested by its presence in non-leguminous plants, such as rice [26, 64]. The secondary structure of *Enod40* lncRNA is highly stable and has five highly conserved domains. The ORF of *Enod40* is very short and synthesis two short peptides. These short peptides regulate the biological activities of Enod40 and consequently help in nodulation [65, 66]. In *M. truncatula*, *Enod40* has been reported in the re-localization of MtRBP1 (*Medicago truncatula* RNA binding protein 1). *Enod40* showed direct interaction with MtRBP1 and re-localized the protein during nodulation process from nuclear speckles into cytoplasmic granules [25].

5. Expression profiling of lncRNAs

5.1. During developmental stages of different tissues

The expression of lncRNAs is regulated through different environmental and biological factors and delving into their diverse biological roles. They exhibit spatial and temporal expression during different developmental stages of various plant tissues. In contrast to the animals, a little is known about the functioning of lncRNAs in plants. The available reports reveal their role in nodule formation [26], lateral root development [67], vegetative and gametophytic development [68], cell-wall synthesis [69], flowering time [27, 54], and several others. The expression profiles developed using high-throughput RNA-seq data from various plants organs marks lncRNAs as an indispensable unit of the transcriptome. For instance, the expression profiles of lncRNAs from root, leaf, stem, spike, and grain in three developmental stages of *T. aestivum* have suggested the role in developmental processes. Furthermore, the lncRNAs show differential expression pattern comparable to the mRNA and highlight their function in related stages [31]. Besides this, the differential expression of lncRNAs in 11 different tissues of chickpea and 13 of maize also strengthens the findings [30]. These results also highlight the higher number of lncRNAs in actively dividing cells and reproductive tissues in comparison to the other [30, 33, 42, 43]. Depending on the expression values, they can be divided into different categories ranging from very low to very high expressing lncRNAs [30, 31]. Furthermore, fragments per kilobase of transcripts per million mapped reads (FPKM), reads per kilobase of transcripts per million mapped reads (RPKM) or transcripts per million (TPM) has to be determined for normalization and estimation of expression level [70]. The alteration in the expression level of various tissues within sundry plants can be correlated with the different genetic makeup and depth of transcriptome sequencing data. Tissue specificity index (TSI) is also calculated for studying the differential expression pattern of lncRNAs. The value of TSI ranges from zero to one, zero for housekeeping genes and one or near to one for sternly tissue-specific genes [31]. The criteria of TSI has revealed that lncRNAs are involved in flower and fruit development in *Fragaria vesca* [34], flower development in *Cicer arietinum* [30], development of fiber in *Gossypium arboreum* [71], and in development of root and floral tissues in *Morus notabilis* [72]. In addition to TSI, cell-type specificity can be interpreted for the expression of lncRNAs in specific cells [29]. For instance, in *Arabidopsis* cell-type specific lncRNAs have been identified in specialized cells but the expression was lower in comparison to mRNA [73]. The knowledge of lncRNAs is limited in plants, but the elevation in the survey of high-throughput RNA-seq data has allowed the prediction of their biological roles through expression profiling.

5.2. Expression under biotic and abiotic stresses

The expression of lncRNAs gets affected by biotic and/or abiotic factors in plants, but the mechanism remains poorly understood. Stress-responsive lncRNAs have emerged as an important component of plant defense machinery. The differential expression patterns in response to various stresses, including biotic and abiotic stresses, suggest the diverse function of lncRNAs at different intervals of stress exposure. For instance, the expression of 1832 lincRNAs gets remarkably affected after 2 h and/or 10 h of drought, salt, cold, and/or ABA (abscisic acid) treatments

in *Arabidopsis*. However, the expression of one of the candidate stress responsive lincRNA increased after treatment by elf18 (EF-Tu), which activates pathogen-associated molecular pat tern responses [22]. Likewise, in *T. aestivum*, 283 lncRNAs were identified as fungal-responsive lncRNAs, out of which 254 and 52 lincRNAs were specifically expressed after infestation with *Blumeria graminis* f. sp. tritici and *Puccinia striiformis* f. sp. tritici, respectively [41]. Later, a total of 44,698 lncRNAs were identified in *T. aestivum* consisting of both stress responsive and tissue-specific lncRNAs [31]. In response to tomato yellow-leaf curl virus, 1565 lncRNAs were expressed in *Solanum lycopersicum* [40]. In case of *Populus trichocarpa*, 2542 lncRNAs were expressed under drought stress condition [37]. The exploration of lncRNAs in various plant species in response to different stress conditions exhibit the dynamic role in plant defense.

6. Databases for lncRNAs

New high-throughput technologies have aided in the exponential rise of RNA-seq data from various plant species. A significant amount of lncRNAs has been identified and characterized for their diverse biological roles. Therefore, it is necessary to organize this data in web-based platforms or databases for further improvement, updates, and analysis [29]. Along with the aid of several computational tools, the data can be analyzed for phylogenetic relationships, expression patterns, molecular interactions, single nucleotide polymorphism, epigenetic variations, etc., and assist in understanding the lncRNAs in plants. The information in these databases can be managed specifically for single or numerous plant species. For instance, PLncRNAdb is specific for four plants species including *A. thaliana*, *A. lyrata*, *P. trichocarpa*, and *Z. mays* and consist of 5000 lncRNAs [74]. The information on 37 plants and 6 algae with data of >120,000 lncRNAs can be accessed on GreeNC database [75]. NONCODE v4 and PLncDB have information on 3853 and >13,000 lncRNA transcripts, respectively in Arabidopsis. Some databases cover the infor mation on both coding and non-coding transcripts like PlantNATsDB accumulating data of 70 plant species on NATs [76]. Besides this, some databases are plant-specific like TAIR10, PNRD, PlantNATsDB, etc., while certain databases (e.g., RNACentral, lncRNAdb v2.0, and NONCODE v4) consist of information from other organisms also in addition to plants [29]. These well-managed databases will allow the researchers to further study the lncRNA in more depth.

7. Biological roles of lncRNAs

The present knowledge on the function of lncRNAs is still limited in plants and a large portion of their function and mechanism is yet to be identified. In spite of this, the biological role of lncRNA has been studied in several plant species as discussed in **Table 2**. Some biological roles have been discussed here to highlight the importance of lncRNAs in plants.

7.1. lncRNA in plant fertility

The participation of lncRNAs in producing the male sterile lines in *O. sativa* is an important example of plant fertility. These male sterile lines are necessary for the hybridization and

Sr. no.	Species name	Annotated lncRNAs	Biological role	Regulatory mechanism	References
1	*Arabidopsis thaliana*	*APOLO*	Auxin-controlled development	Chromatin loop dynamics	[77]
		ASCO-lncRNA	Lateral root development	Alternative splicing regulators	[67]
		asHSFB2a	Vegetative and gametophytic development	Antisense transcription	[68]
		COLDAIR	Flowering time	Promoter interference	[27]
		COOLAIR	Flowering time	Histone modification	[28, 54, 78]
		HID1	Photomorphogenesis	Chromatin association	[79]
		IPS1	Phosphate homeostasis	Target mimicry	[46]
2.	*Glycine max*	*GmENOD40*	Nodule formation	Protein re-localization	[61]
3.	*Hordeum vulgare*	*HvCesA6 lnc-NAT*	Cell-wall synthesis	siRNA precursor	[69]
4.	*Medicago truncatula*	*MtENOD40*	Nodule formation	Protein re-localization	[66]
5.	*Oryza sativa*	*Cis-NATPHO1;2*	Phosphate homeostasis	Translational enhancer	[80]
		LDMAR (P/ TMS12-1	Fertility	Promoter interference	[81, 82]
		OsPI1	Phosphate homeostasis	Unknown	[83]
		OsENOD40	Nodule formation	Unknown	[26]
6.	*Petunia hybrid*	*SHO lnc-NAT*	Local cytokinin synthesis	dsRNA degradation	[84]
7.	*Solanum lycopersicum*	*TPS11*	Phosphate homeostasis	Unknown	[85]

Table 2. List of some annotated lncRNAs.

breeding processes. lncRNAs are known to induce photoperiod-sensitive genetic male sterility (PSMF) in *O. sativa* [82, 86], but the mechanism is not completely well understood. But according to the available reports, two different mechanisms of lncRNA can be possible [23]. In one mechanism, the high expression of the *long day* (*LD*)*-specific male-fertility-associated RNA* (*LDMAR*), a type of lncRNA, is important for the fertility of rice plant during long day (LD) conditions. During male sterility, the programme cell death (PCD) of anther cells occur due to lowered expression of *LDMAR* under LD conditions. The reduced expression of *LDMAR* is mediated by over expressing psi-LDMAR (a siRNA), transcribed in the promoter region of *LDMAR*. Enhanced expression of Psi-LDMAR caused methylation in promoter region through RdDM mechanism [81]. The other mechanism suggested the involvement of osa-sm R5864w (a 21-nt sRNA) which was formed from a unique ncRNA encoded by *LDMAR*. The

point mutation of C to G in osa-sm R5864w, resulting in the loss of function, leads to the production of light and temperature sensitive male sterile lines of rice [82].

7.2. Role in alternate splicing

Plant lncRNAs are known to increase the complexity of transcriptome and proteome by participating in alternative splicing. It was first reported in Arabidopsis, where lncRNA behaved as an alternative splicing competitor (ASCO) [67]. Together with the nuclear speckle RNA-binding protein (NSR), ASCO-lncRNA forms an alternative splicing regulatory module. The expression of AtNSR in primary and lateral root meristems regulates the development of lateral roots. The interaction of AtNSR with overexpressing ASCO-lncRNA affects the splicing pattern of mRNA targeted by NSR in transgenic plant [67, 87]. This indicates the role of lncRNA as a regulator of alternative splicing.

7.3. Plant lncRNAs in photomorphogenesis

Most of the plant growth and developmental processes are regulated by different climatic factors among them light is one of the most important factor [88]. The role of lncRNA in the regulation of photomorphogenesis is still an interesting area of research because most of the identified regulatory molecules are proteins. In *A. thaliana*, several light responsive lncRNAs have been identified associated with histone modifications [89]. Identification and functional characterization of *HIDDEN TREASURE 1* (*HID1*), a novel lncRNA, involved in photomorphogenesis have been accomplished [89]. It may control the process of photomor phogenesis by regulating the expression of PHYTOCHROME INTERACTING FACTOR 3 (PIF3), a transcription factor involved in light response [89]. It could negatively regulate the expression of *PIF3* gene by binding to its promoter directly or in association with chromatin [89]. The occurrence of *HID1* homologs has been described in other plant species exhibiting conserved functions. The findings also shed light on the involvement of other ncRNAs in light responses.

8. Limitations in computational analysis of lncRNAs

The selection of lncRNAs from the complete set of RNAs is broadly based on three criteria: (i) transcript length of ≥200 bp, (ii) small open reading frame with ≤300 bp, and (iii) transcripts without homology to known proteins. In addition to this, several other factors like the type of cDNA libraries or transcriptional sequence data, depth of sequencing, and coding potential of transcripts, also contribute in the screening of lncRNAs. The challenges during computational analysis come when some protein-coding gene which fulfill the basic selection criteria and encode a functional peptide. Besides this, the functional long non-coding transcript may have ORF >300 bp and share homology with known protein-coding genes will also produce hindrance in the identification [90]. Another challenge comes with the transcripts that not only function as an RNA molecule, but also encodes a peptide [91]. The advancement in computational approaches have been made to overcome these limitations and for more accurate

differentiation between coding and non-coding transcripts [92]. The use of support vector machines (SVMs) or other machine learning algorithms along with the computational methods have increased the confidence of disparity in between coding and non-coding transcripts [93]. However, the identity and function of computationally identified lncRNA needs to be validated individually by experimentation.

Acknowledgements

Authors are grateful to Panjab University, Chandigarh, India for research facilities. SKU is grateful to Department of Science and Technology (DST), Government of India, Science and Engineering Research Board (SERB) for Early Career Research Award (ECR/2016/001270), and DST-INSPIRE Faculty Fellowship (IFA12-LSPA-09).

Author details

Shivi Tyagi, Alok Sharma and Santosh Kumar Upadhyay*

*Address all correspondence to: skupadhyay@pu.ac.in

Department of Botany, Panjab University, Chandigarh, India

References

[1] Maxam AM, Gilbert W. A new method for sequencing DNA. PNAS. 1977;**74**:560-564

[2] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology. 1975;**94**:441-448

[3] Watson JD, Crick FH. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature. 1953;**171**:737-738

[4] Buermans HPJ, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta. 2014;**1842**:1932-1941

[5] Goodwin S, McPherson JD, Richard McCombie W. Coming of age: Ten years of next generation sequencing technologies. Nature Reviews Genetics. 2016;**17**:333-351

[6] Levy SE, Myers RM. Advancements in next-generation sequencing. Annual Review of Genomics and Human Genetics. 2016;**17**:95-115

[7] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;**431**:931-945

[8] Kaper F, Swamy S, Klotzle B, Munchel S, Cottrell J, et al. Whole-genome haplotyping by dilution, amplification, and sequencing. PNAS. 2013;**110**:5552-5557

[9] Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. Nature Methods. 2015;**12**:351-356

[10] Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science. 2003;**299**:682-686

[11] Tilgner H, Grubert F, Sharon D, Snyder MP. Defining a personal, allele-specific, and single molecule long-read transcriptome. PNAS. 2014;**111**:9869-9874

[12] Treutlein B, Gokce O, Quake SR, Sudhof TC. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. PNAS. 2014;**111**:E1291-E1299

[13] St Laurent G, Wahlestedt C, Kapranov P. The landscape of long non-coding RNA classification. Trends in Genetics. 2015;**31**(5):239-251

[14] Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, et al. Direct RNA sequencing. Nature. 2009;**461**(7265):814-818

[15] Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, et al. CAGE: Cap analysis of gene expression. Nature Methods. 2006;**3**(3):211-222

[16] Phillipe N, Samra EB, Boureux A, Mancheron A, Ruffle F, et al. Combining DGE and RNA-sequencing data to identify new polyA+ non-coding transcripts in the human genome. Nucleic Acids Research. 2014;**42**(5):2820-2832

[17] Fullwood MJ, Wei CL, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Research. 2009;**19**(4):521-532

[18] Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR. Examples of the complex architecture of the human transcriptome revealed by Race and high-density tiling arrays. Genome Research. 2005;**15**(7):987-997

[19] Mercer TR, Daniel JG, Marcel ED, Joanna C, Cole T, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nature Biotechnology. 2012;**30**:99-104

[20] Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011;**145**:622-634

[21] Kim ED, Sung S. Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. Trends in Plant Science. 2012;**17**:16-21

[22] Liu J, Jung C, Xu J, Wang H, Deng S, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. The Plant Cell. 2012;**24**:4333-4345

[23] Liu J, Wang H, Chua NH. Long noncoding RNA transcriptome of plants. Plant Biotechnology Journal. 2015;**13**:319-328

[24] Kouchi H, Hata S. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. Molecular & General Genetics. 1993;**238**:106-119

[25] Campalans A, Kondorosi A, Crespi M. Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. The Plant Cell. 2004;**16**:1047-1059

[26] Kouchi H, Takane K, So RB, Ladha JK, Reddy PM. Rice ENOD40: Isolation and expression analysis in rice and transgenic soybean root nodules. The Plant Journal. 1999;**18**:121-129

[27] Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science. 2011;**331**:76-79

[28] Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* polycomb target. Nature. 2009;**462**:799-802

[29] Bhatia G, Goyal N, Sharma S, Upadhyay SK, Singh K. Present scenario of long noncoding RNAs in plants. Non Coding RNA. 2017;**3**:16

[30] Khemka N, Singh VK, Garg R, Jain M. Genome-wide analysis of long intergenic noncoding RNAs in chick pea and their potential role in flower development. Scientific Reports. 2016;**6**:33297

[31] Shumayla, Sharma S, Taneja M, Tyagi S, Singh K, Upadhyay SK. Survey of high throughput RNA-seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. Frontiers in Plant Science. 2017;**8**:1019

[32] Szcześniak MW, Rosikiewicz W, Makalowska I. CANTATAdb: A collection of plant long non-coding RNAs. Plant and Cell Physiology. 2015;**57**:pcv201. DOI: 10.1093/pcp/pcv201

[33] Hao Z, Fan C, Cheng T, Su Y, Wei Q, et al. Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. PLoS One. 2015;**10**:e0121800

[34] Kang C, Liu Z. Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. BMC Genomics. 2015;**16**:815

[35] Wen J, Parker BJ, Weiller GF. *In silico* identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula*. In Silico Biology. 2007;**7**(4-5):485-505

[36] Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, et al. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biology. 2014;**15**:512

[37] Shuai P, Liang D, Tang S, Zhang Z, Ye C, et al. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. Journal of Experimental Botany. 2014;**65**:4975-4983

[38] Qi X, Xie S, Liu Y, et al. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. Plant Molecular Biology. 2013;**83**:459-473

[39] Scarano D, Rao R, Corrado G. *In silico* identification and annotation of non-coding RNAs by RNA-seq and De Novo assembly of the transcriptome of tomato fruits. PLoS One. 2017;**12**:e0171504

[40] Wang J, Yu W, Yang Y, Li X, Chen T, Liu T, et al. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. Scientific Reports. 2015;**5**:16946

[41] Zhang H, Hu W, Hao J, Lv S, Wang C. Genome-wide identification and functional prediction of novel and fungi-responsive lincRNAs in *Triticum aestivum*. BMC Genomics. 2016;**17**:238

[42] Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biology. 2014;**15**:R40

[43] Zhang W, Han Z, Guo Q, Liu Y, Zheng Y, et al. Identification of maize long non-coding RNAs responsive to drought stress. PLoS One. 2014;**9**(6):e98958

[44] Lv Y, Liang Z, Ge M, Qi W, Zhang T, et al. Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). BMC Genomics. 2016;**17**:350

[45] Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annual Review of Biochemistry. 2012;**81**:145-166

[46] Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, et al. Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genetics. 2007;**39**:1033-1037

[47] Aung K, Lin SI, Wu CC, Huang YT, CL S, Chiou TJ. pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. Plant Physiology. 2006;**141**:1000-1011

[48] Bari R, Datt Pant B, Stitt M, Scheible WR. PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiology. 2006;**141**:988-999

[49] Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL. Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. Plant Cell. 2006;**18**:412-421

[50] Fujii H, Chiou TJ, Lin SI, Aung K, Zhu JKA. miRNA involved in phosphate-starvation response in *Arabidopsis*. Current Biology. 2005;**15**:2038-2043

[51] Rubio-Somoza I, Weigel D. MicroRNA networks and developmental plasticity in plants. Trends in Plant Science. 2011;**16**:258-264

[52] Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, et al. Integration of human large intergenic noncoding RNAs reveals global properties and specific sub-classes. Genes & Development. 2011;**25**:1925-1927

[53] Michaels SD, Amasino RM. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell. 1999;**11**:949-956

[54] Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. Science 2013;**340**(6132):619-621

[55] Hirsch J, Lefort V, Vankersschaver M, Boualem A, Lucas A, et al. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived. Plant Physiology. 2006;**140**:1192-1204

[56] Lee Y, Kim M, Han J, Yeom KH, Lee S, et al. MicroRNA genes are transcribed by RNA polymerase II. The EMBO Journal. 2004;**23**:4051-4060

[57] Chen XM. Small RNAs and their roles in plant development. Annual Review of Cell and Developmental Biology. 2009;**25**:21-44

[58] Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, et al. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. Genome Research. 2009;**19**:57-69

[59] Castel SE, Martienssen RA. RNA interference in the nucleus: Roles for small RNAs in transcription, epigenetics and beyond. Nature Reviews. Genetics. 2013;**14**:100-112

[60] Böhmdorfer G, Wierzbicki AT. Control of chromatin structure by long noncoding RNA. Trends in Cell Biology. 2015;**25**(10):623-632

[61] Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, et al. Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. The Plant Journal. 1993;**3**:573-585

[62] Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, et al. Enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. The EMBO Journal. 1994;**13**:5099-5112

[63] Compaan B, Yang WC, Bisseling T, Franssen H. ENOD40 expression in the pericycle precedes cortical cell division in Rhizobium-legume interaction and the highly conserved internal region of the gene does not encode a peptide. Plant and Soil. 2001;**230**:1-8

[64] Gultyaev AP, Roussis A. Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40. Nucleic Acids Research. 2007;**35**: 3144-3152

[65] Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. Proceedings of the National Academy of Sciences of the United States of America. 2002;**99**:1915-1920

[66] Sousa C, Johansson C, Charon C, Manyani H, Sautter C, et al. Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. Molecular and Cellular Biology. 2001;**21**:354-366

[67] Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, et al. Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. Developmental Cell. 2014;**30**:166-176

[68] Wunderlich M, Gross-Hardt R, Schof F. Heat shock factor HSFB2a involved in gametophyte development of *Arabidopsis thaliana* and its expression is controlled by a heat-inducible long non-coding antisense RNA. Plant Molecular Biology. 2014;**85**(6):541-550

[69] Held MA, Penning B, Brandt AS, Kessans SA, Yong W, et al. Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell

wall biosynthesis in barley. Proceedings of the National Academy of Sciences of the United States of America. 2008;**105**:20534-20539

[70] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. A survey of best practices for RNA-seq data analysis. Genome Biology. 2016;**17**:13

[71] Zou C, Wang Q, Lu C, et al. Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). Science China. Life Sciences. 2016;**59**:164

[72] Song X, Sun L, Luo H, Ma Q, Zhao Y, et al. Genome-wide identification and characterization of long non-coding RNAs from mulberry (*Morus notabilis*) RNA-seq data. Genes. 2016;**7**:11

[73] Li S, Yamada M, Han X, Ohler U, Benfey PN. High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. Developmental Cell. 2016;**39**:508-522

[74] PLncRNAdb Ming Chen's Lab. Available online: http://bis.zju.edu.cn/PlncRNADB/index.php [Accessed: 25 December, 2016]

[75] Gallart AP, Pulido AH, de Lagrán IA, Sanseverino W, Cigliano RA. GREENC: A Wiki-based database of plant lncRNAs. Nucleic Acids Research. 2016;**44**:D1161-D1166

[76] Chen D, Yuan C, Zhang J, Zhang Z, Bai L, Meng Y, Chen LL, Chen M. PlantNATsDB: A comprehensive database of plant natural antisense transcripts. Nucleic Acids Research. 2012;**40**:D1187-D1193

[77] Ariel F, Romero-Barrios N, Jegu T, Benhamed M, Crespi M. Battles and hijacks: Non-coding transcription in plants. Trends in Plant Science. 2015;**20**(6):362-371

[78] Liu F, Marquardt S, Lister C, Swiezewski S. Targeted 3′ processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing. Science. 2010;**327**(5961):94-97

[79] Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, et al. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. Genome Research. 2014;**24**:444-453

[80] Jabnoune M, Secco D, Lecampion C, Robaglia C, Shu QY, et al. A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. The Plant Cell. 2013;**25**:4166-4182

[81] Ding J, Shen J, Mao H, Xie W, Li X, Zhang Q. RNA-directed DNA methylation is involved in regulating photoperiod-sensitive male sterility in rice. Molecular Plant. 2012;**5**:1210-1216

[82] Zhou H, Liu Q, Li J, Jiang D, Zhou L, Wu P, et al. Photoperiod- and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. Cell Research. 2012;**22**:649-660

[83] Wasaki J, Yonetani R, Shinano T, Kai M, Osaki M. Expression of the OsPI1 gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. The New Phytologist. 2003;**158**:239-248

[84] Zubko E, Meyer P. A natural antisense transcript of the *Petunia hybrida* Sho gene suggests a role for an antisense mechanism in cytokinin regulation. The Plant Journal. 2007;**52**:1131-1139

[85] Liu C, Muchhal US, Raghothama KG. Differential expression of TPS11, a phosphate starvation-induced gene in tomato. Plant Molecular Biology. 1997;**33**:867-874

[86] Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, et al. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. Proceedings of the National Academy of Sciences of the United States of America. 2012;**109**:2654-2659

[87] Kornblihtt AR. A long noncoding way to alternative splicing in plant development. Developmental Cell. 2014;**30**:117-119

[88] Jiao Y, Lau OS, Deng XW. Light-regulated transcriptional networks in higher plants. Nature Reviews. Genetics 2007;**8**:217-230

[89] Wang Y, Fan X, Lin F, He G, Terzaghi W, et al. *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. Proceedings of the National Academy of Sciences of the United States of America. 2014;**111**:10359-10364

[90] Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmond SM, Mattick JS. NRED: A database of long noncoding RNA expression. Nucleic Acids Research. 2009;**37**(Database issue):D122-D126

[91] Bardou F, Merchan F, Ariel F, Crespi M. Dual RNAs in plants. Biochimie. 2011;**93**(11):1950-1954

[92] Castrignano T, Canali A, Grillo G, Liuni S, Mignone F, et al. CSTminer: A web tool for the identification of coding and noncoding conserved sequence tags through cross-species genome comparison. Nucleic Acids Research. 2004;**32**:W624-W627

[93] Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research. 2007;**35**:W345-W349

Chapter 7

The CRISPR/Cas9 System for Crop Improvement: Progress and Prospects

Kah-Yung Bernard Leong, Yee-Han Chan,
Wan Muhamad Asrul Nizam Wan Abdullah,
Swee-Hua Erin Lim and Kok-Song Lai

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75024

Abstract

The global demand for high-quality crops is continuously growing with time. Crop improvement techniques have a long history and they had been applied since the beginning of domestication of the first agricultural plants. Since then, various new techniques have and are being developed to further increase the commercial value and yield of crops. The latest crop improvement technique known as genome editing is a technique that enables precise modification of the plant genome via knocking out undesirable genes or enabling genes to gain new function. The variants generated from the genome editing are indistinguishable from naturally occurring variation. It is also less time-consuming and more readily accepted in the market commercially. The usage of genome editing has proven to be advantages and plays a promising role in future crop improvement efforts. Therefore, in this chapter, we aim to highlight the progress and application of genome editing techniques, in particular, the CRISPR/Cas9 system as a powerful genome editing tool for crop improvement. In addition, the challenges and future prospects of this technology for crop improvement will also be discussed.

Keywords: CRISPR/Cas9, crop improvement, genome editing, TALENs, ZFN

1. Introduction

As the agricultural commercial market continues to grow, development of new techniques for crop improvement is always in high demand; conversely, traditional breeding practices for crop improvement are phasing out as some techniques are far too time-consuming and

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

laborious and usually result in little yield. Examples of such traditional breeding are the selective crossbreeding between plants with the desired trait or classical breeding with induced mutation via radiation or chemicals. Currently, a new methodology that involves genetic engineering was developed and had paved the way to improve the quality of plants with high specificity for the attribute of interest; the application of site-specific nucleases (SSNs) [1].

SSNs are programmable nucleases that have the ability to produce DNA single-stranded or double-stranded breaks (DSBs) that activate the endogenous DNA repair pathways of the cells to repair the DNA damage and this usually leads to targeted mutagenesis [1, 2]. This technology empowers plant scientists to precisely regulate any genes in any plant while directly evaluating the function of that specific gene in the plant [3]. As a result, plants that express characteristics such as higher yield, disease resistance or shorter maturation periods can be generated through this methodology [4–6].

SSN-based genome editing system can be classified into three categories that are the zinc finger nuclease (ZFN), transcription-like effector nucleases (TALENs) and, the recently developed, clustered regularly interspaced short palindromic repeats that are associated with the RNA-guided Cas9 double-stranded DNA-binding protein (CRISPR/Cas9) [1]. The main differences between the categories lie in their mechanism of the double-stranded break induction and their efficiency in targeting their desired sequences [7]. These SSN-based genome editing systems are very powerful and they have undoubtedly revolutionized the agriculture industry.

1.1. Site-specific nucleases (SSNs)

SSNs have two major components, which are the engineered nonspecific endonucleases and the sequence-specific DNA-binding domains. The nonspecific nucleases have the ability to produce DSBs in DNA but they are very random as they lack specificity. On the other hand, DNA-binding domains are proteins that can specifically bind to DNA sequences that are complementary to them [8]. Hence, when DNA-binding domains synergize with the endonucleases, the paired components are able to introduce breaks at any specific target site [9, 10].

As shown in **Figure 1**, after the DNA break is induced, it will trigger the native DNA repair mechanisms of the cell to fix the break either by the error-prone nonhomologous end joining (NHEJ) mechanism or by the homology-directed repair (HDR) mechanism [11]. The NHEJ mechanism can occur during any phases of the cell cycle and due to its high erroneous repair rate, it occasionally forms frameshift mutation. Hence, this phenomenon can be exploited to form what is known as "knocking out" of a specific gene where the functionality of the protein encoded by that gene is lost due to random insertion, deletion or inversion. However, if provided an engineered DNA repair template that is homologous to the upstream and downstream of the target sequence, the HDR pathway can then be activated instead. The HDR mechanism functions to repair the break differently from which it will insert a new strand of DNA as a form of a desired mutation that resembles the DNA repair template [7]. This is also known as gene "knocking in" where the genome now has a new gain-of-function to encode a specific protein.

http://dx.doi.org/10.5772/intechopen.75024

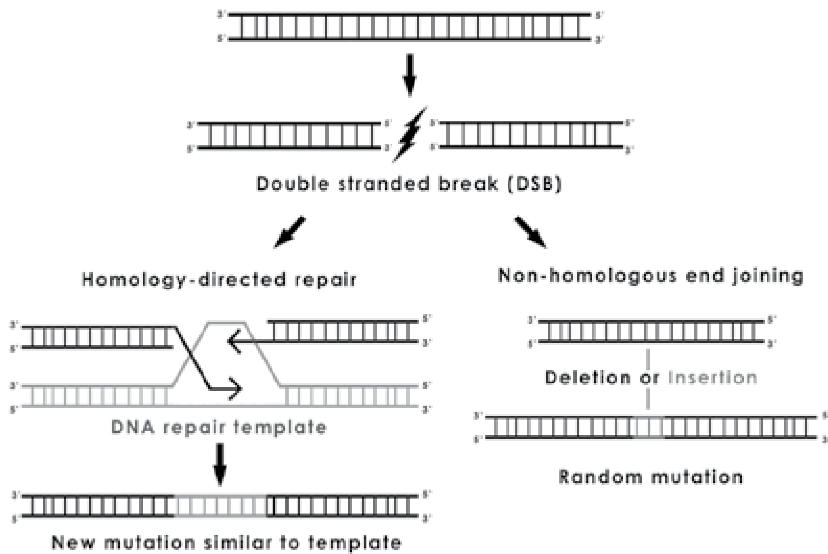


Figure 1. Different DNA repair mechanisms for DSBs. Left: if a DNA repair template is provided, the repair will proceed via the homology-directed repair mechanism where a desired mutation that is based on the template is introduced to the DNA. Right: the double-stranded break can be repaired via error-prone nonhomologous end joining mechanism where random mutation will be introduced to the DNA in the form of deletion or insertion. Adapted from Ott de Bruin et al. [12].

1.1.1. Zinc finger nuclease (ZFN)

One of the earliest SSNs developed for genome editing is known as the zinc finger nuclease (ZFN). ZFN is constructed through the coupling of two major domains, which are the eukaryotic-based DNA-binding domain known as the zinc finger protein (ZFP) and the endonuclease domain of the Fok1 restriction enzyme (**Figure 2**). Consequently, this coupling combines the quality of the DNA-binding specificity of the ZFP and cleaving activity of the Fok1 endonuclease into a single system, thus making ZFN a useful tool for genome editing [10].

Zinc finger proteins (ZFPs) have the ability to specifically bind to a discrete 3-base pairs (bp) sequence of DNA known as codons [12]. In fact, recent studies have shown that each ZFNs can use 3 zinc fingers to bind a 9-bp target sequence, and when made into a ZFN dimer, it can recognize up to 18-bp of a DNA [9, 13]. Furthermore, there are up to 64 possible types of ZFPs that can be produced as there are 64 codon combinations known today [10]. As a result, by determining which type of ZFPs are to be linked into the ZFN, researchers are able to design ZFNs in a way that they would only target desirable sequences. Additionally, studies have also shown that more fingers (up to six per ZFN) can be added to recognize longer and rarer target sequences [10].

Paired with the zinc fingers is the nonspecific Fok1 endonuclease domain, which is an enzyme that can be found naturally in *Flavobacterium okeanokoites* and it has the ability to induce DNA double-stranded breaks [14, 15]. By linking both domains together, they will now be known as the zinc finger nuclease, and the ZFP domain will lead the Fok1 endonuclease domain to the desired DNA sequence to cleave the target site, which is adjacent to it; this equips ZFN with the ability to precisely cut any targeted DNA sequences. In early studies, it has been

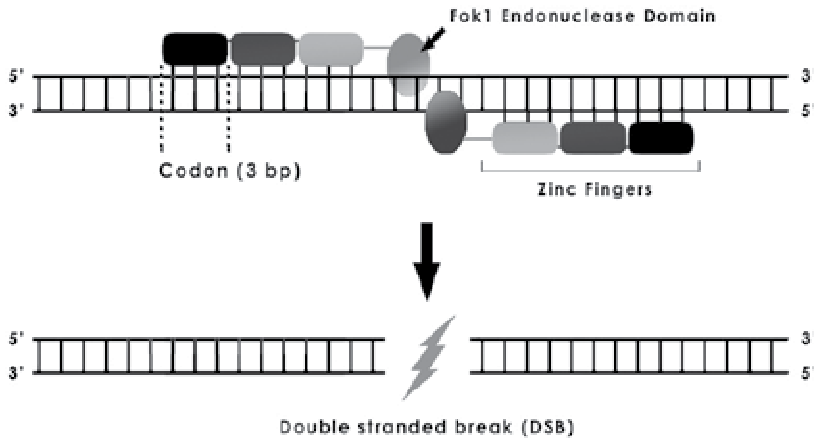


Figure 2. ZFN dimer binding to DNA at each side of the sequence. Different shaded boxes represent the different zinc fingers and each of them binds to a specific codon. Short dotted lines indicate 3-bp long codons. Shaded ellipses indicate the Fok1 endonuclease domain with the ability to induce DNA double-stranded breaks and they are coupled by the zinc fingers to guide them to their cleavage site. The cleavage site will be between the dimer and it is about 4-bp long. Adapted from Carroll [9].

known that the Fok1 endonuclease domain must dimerize to produce the breaks on DNA [13, 16]. However, the interaction between the dimer ZFN can be weak if it is not optimally designed. Thus, the optimal design for ZFN to achieve DNA cleavage is to direct two sets of fingers to neighboring sequences and join each to a Fok1 DNA-cleaving domain. Once the dimers bind to their respective sequences, adjacent cleavage factors will facilitate the dimerization and cleavage of the DNA strand [9]. Nevertheless, the construction of ZFN has proven to be difficult as there are many complicated interference in the interaction between the ZFPs despite many attempts made in the past to simplify them [17–23].

1.1.2. Transcription activator-like effector nucleases (TALENs)

It may seem that ZFN is the most practical method; nevertheless, the challenge lies in the construction as mentioned in addition to less popularity as the success rate for the DNA repair pathway via HDR is still considerably low [7]. Therefore, a more recent genome editing tool known as transcription activator-like effector nucleases (TALENs) had been developed with better modularity [24, 25]. TALENs are quite similar to ZFN in terms of the idea of directing the same nonspecific endonuclease to a specific site with the help of DNA-binding motifs (**Figure 3**).

Unlike ZFN, the DNA-binding domains for TALENs are known as the transcription activator-like effector (TALE) proteins rather than ZFPs and they are found in pathogenic plant bacteria (specifically of genus *Xanthomonas*) instead of in a eukaryotic cell. TALE protein is comprised of a repetitive sequence of a series of 34 amino acid residues, where each TALE protein has the ability to selectively bind to one nucleotide in the DNA target site [1]. Specifically, the pair residue at the 12th and 13th position is the one that determines the nucleotide specificity of the TALE proteins and they are known as the repeat variable domain (RVD).

http://dx.doi.org/10.5772/intechopen.75024

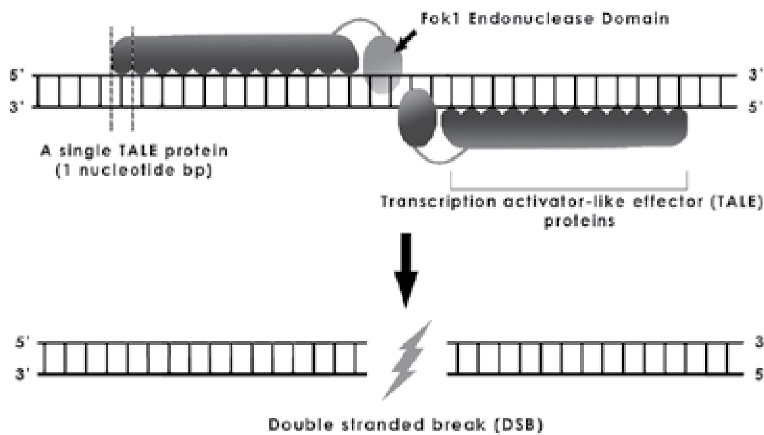


Figure 3. Transcription activator-like effector nuclease (TALEN) dimer bound to DNA. Like the ZFN, it has a Fok1 endonuclease domain that can cleave DNA. Indicated by small dotted lines, each TALE protein can specifically detect a single nucleotide base pair and it can be linked with more TALE proteins to detect longer sequences. By coupling the Fok1 endonuclease domain with the TALE proteins, it is now known as TALEN and it can specifically detect and cleave DNA. Adapted from Gupta [73].

TALE proteins have a different method to detect DNA sequences. They recognize a single nucleotide instead of DNA triplets. In fact, TALE proteins can also be assembled in a way to recognize virtually any DNA sequence that is desired by their users [26]. This attribute enables TALENs to have a higher range of target sites to suit user specifications, making them significantly more flexible and generally more straightforward. Like ZFN, the DNA-cleaving domain for TALENs is the nonspecific Fok1 endonuclease domain, which is highly dependent on the DNA-binding domains to achieve higher specificity of DSBs [1]. In a nut shell, both ZFNs and TALENs function as genome editing tools but they are considerably complex, a challenge to be constructed and less efficient compared to newly developed genome editing tool like CRISP/Cas9 system.

1.1.3. CRISPR/Cas9

The most recently developed SSN in genome editing technology is known as the CRISPR/Cas9 system. The CRISPR/Cas9 system was developed based on the bacterial type-II CRISPR/Cas adaptive immune system that is deployed by the hosts to recognize and eliminate any invading phage or plasmid DNA [27–29]. The three major components of the immune system are the protospacer-containing CRISPR RNA (crRNA), transactivating crRNA (tracrRNA) and the Cas9 endonuclease [27, 30].

During the invasion of foreign DNA, the bacterial type-II CRISPR/Cas system will integrate a short fragment of the foreign DNA, called "spacers," into the CRISPR genomic loci [31]. These spacers act as a form of an acquired immunity memory for the host. Then, the spacers will be transcribed and processed in the form of crRNA. Thus, the CRISPR genome loci act as the library where it can store information to enable the bacteria CRISPR/Cas9 system to retarget any known foreign DNA. In case of another attack by the invader, the crRNAs will then bind to the tracrRNAs that have the ability to trigger the direction of the Cas9 endonuclease

to the target site of the foreign DNA. After that, it will induce DSBs on the foreign DNA that carries the same protospacer sequence, which is accompanied by the protospacer adjacent motif (PAM) to disable the virus [7, 30]. As a matter of fact, recent studies have shown that the specificity of Cas9 protein is highly dependent to the PAM sequence as it licenses the landing of the crRNA-tracrRNA-Cas9 complex [32–34].

The CRISPR/Cas9 system has been redesigned to work more efficiently by fusing the Cas9 endonuclease with an artificial crRNA-tracrRNA chimera known as guide RNA (gRNA) [35–37]. The gRNA maintains the function of both crRNA and tracrRNA where just by itself is able to recognize the PAM-containing target sequence to direct the Cas9 protein for DNA cleaving activity. In fact, studies have shown that the gRNA works more efficiently compared to the combination of crRNA and tracrRNA [38, 39].

Compared to ZFNs and TALENs, the CRISPR/Cas9 system is potentially more efficient and effective with three main advantages [40], which are:

1. *Target design simplicity*: the DNA sequence targeting system of the CRISPR/Cas9 system is based on a formation of RNAs rather than protein or DNA recognition. RNA is much more accessible as it is simple and readily produced to be used to target any sequence in a genome.
2. *Efficiency*: the CRISPR/Cas9 system is a straightforward tool for genome editing. Modification of the target genome can be carried out by directly introducing RNAs that encode the Cas9 protein and gRNA to the host. This method allows researchers to skip all the long and laborious processes of the classical homologous recombination techniques.
3. Multiplexed mutation: several targeted mutations can be carried out in multiple genes at different locations at once by introducing multiple gRNAs at once. For example, Dr. Yang H. and his team have successfully simultaneously introduced mutation in five different genes in mouse embryonic stem cells [41–43].

2. Mechanism of CRISPR/Cas9 system

The CRISPR/Cas9 system is a type II adaptive immune system in bacteria and archaea, protecting them against invading nucleic acid such as virus by cleaving the foreign DNA through specific sequence recognition [29]. The immunity is acquired through the integration of short fragment of the foreign DNA as spacer between two adjacent repeats at the proximal end of a CRISPR locus [29]. The bacterial CRISPR/Cas9 system involves three stepwise processes, namely acquisition, biogenesis and interference.

2.1. Bacterial adaptive immunity: acquisition, biogenesis and interference

The defense mechanism of CRISPR/Cas9 system can be divided into three stages, which are spacer acquisition or adaptation, crRNA biogenesis and interference (**Figure 4**).

Conceptually, the spacer acquisition process can be further divided into two steps, which are protospacer selection and integration of the spacer into the CRISPR array to synthesize new repeat [44]. The protospacer selection step of the spacer acquisition stage is guided by a conserved DNA sequence element, namely protospacer adjacent motif (PAM), that is located downstream of the DNA target, which has the sequence of 5′-NGG-3 [44]. Next, the selected protospacer will be integrated into the leader-repeat boundary of CRISPR array as new spacer and is duplicated in order to synthesize new repeats. Each repeat is a 29 nucleotide sequence, and repeats are interspaced by five intervening 32-nucleotide nonrepetitive sequences [30].

Then, the CRISPR will be transcribed into a long precursor crRNA (pre-crRNA) with the help of Cas proteins (Cas1, Cas2, Cas9 and Cas4/casn2) or simply known as biogenesis. During this process, the tracrRNA will be transcribed from the bacterial genome as it is needed for processing of the pre-crRNA into a mature guide crRNA. Next, the antirepeat sequence of tracrRNA will enable the complementary base pairing with each pre-crRNA repeats, resulting in the formation of a crRNA-tracrRNA duplex [45].

During the target interference stage, the crRNA-tracrRNA duplex recognizes the PAM sequence located downstream of the protospacer in the foreign DNA and triggers the ‘non–self-activation,’ which prevents self-targeting of CRISPR array. After that, the duplex guides the Cas9 endonuclease to bind and cleave the DNA target, resulting in the formation of DSBs. Additionally, the DNA target specificity duplex is provided by what is called the ‘seed sequence,’ which is located at approximately 12 bases upstream of the PAM sequence that matches the RNA with the DNA target [29] (**Figure 4**).

2.2. Engineered CRISPR/Cas9 in plant genome editing

The CRISPR/Cas9 system for plant genome editing comprises four steps. First, a gene-specific gRNA is designed and constructed by fusing crRNA and tracrRNA. Many online tools have been developed for computer-based design of gRNA [46]. However, the in silico design of gRNA has not been fully adapted for plants, and large-scale data collection and systematic study of gRNA efficiencies in plant cells are needed to increase the accuracy of computational gRNA selection [46]. Both gRNA and Cas9 expression cassettes are constructed separately. The expression of gRNA is driven by U3 or U6 small nuclear RNA gene promoters with defined initiation and termination site, facilitating the transcription of gRNA by RNA polymerase III. The first nucleotide in the guide sequence is a ‘G’ if U6 promoter is used or an ‘A’ if U3 promoter is used [47]. Guide sequence should match the target, except for the first nucleotide (5′ G or A) that does not have to match [47]. In Cas9 expression, nuclear localization of Cas9 requires fusion of a single or dual nuclear localization signal (NLS) to the Cas9 coding sequence, which is 4107-bp in length. Both expression cassettes are then assembled into a vector.

The next step involves the transformation of protoplasts with CRISPR in which the activity of gRNA is best validated before being used in genome editing [46]. This step is followed by the selection of active CRISPR using polymerase chain reaction (PCR) or restriction enzyme digestion. Then, the CRISPR/Cas9 system in the form of vector is delivered into the target

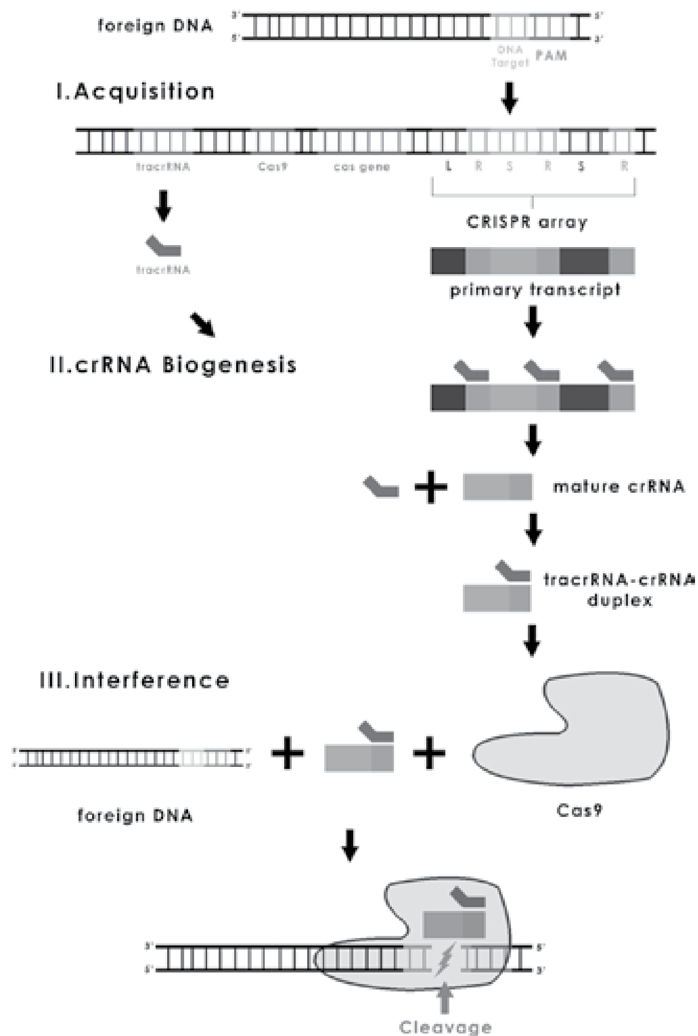


Figure 4. Type II adaptive immunity system by CRISPR/Cas9 in bacteria. The type II adaptive immunity system by CRISPR/Cas9 involves three stages, which are acquisition, crRNA biogenesis and interference to cleavage the DNA target, resulting in the formation of DSBs.

plant cell. The CRISPR/Cas9 system can be delivered via *Agrobacterium*-mediated transformation or particle bombardment [46]. Up to this stage, the expression cassettes are stably integrated into plant genome. Finally, the transformed plants with targeted mutations are screened by polymerase chain reaction (PCR) genotyping and confirmed by sequencing [46].

2.3. RNA-guided DNA cleavage by Cas9

The Cas9 protein consists of six domains, which are REC I, REC II, Bridge Helix, PAM interacting, HNH and RuvC [47]. It remains inactive in the absence of gRNA. The gRNA binds to Cas9 protein and induces a conformational change to form a riboprotein complex. This results

in the activation of the Cas9 protein from a non–DNA-binding conformation into an active DNA-binding conformation. Once the Cas9 protein is activated, it recognizes DNA target by binding with DNA sequence that matches its PAM. Once the Cas9 protein finds a potential target sequence with complementary PAM, it melts the bases immediately upstream of the PAM and pairs them with the complementary sequence on the gRNA. RuvC and HNH nuclease domains will then cleave the DNA target after the third nucleotide base upstream of the PAM to generate blunt-ended DSBs.

2.4. Orthologues of CRISPR/Cas9

To date, the most common Cas9 protein used in plants is from the bacteria *Streptococcus pyrogens* (SpCAs9), which recognizes the NGG-type PAM [46]. Even though this PAM sequence is widely distributed across plant genomes, it does not cover the entire genome of the plant [46]. Many naturally occurring Cas9 orthologues impose distinct crRNA-tracrRNA duplex and PAM requirements [30]. Therefore, exploring orthologous Cas9 proteins with cognate gRNA and PAM sequences would greatly expand possible target sequences in a given genome and add new Cas9 orthologues with unique properties into the CRISPR/Cas arsenal [30]. As the alternative orthologous Cas9 requires different PAM sequences, the total number of possible target sites within a plant genome can be increased. Hence, the CRISPR/Cas9 system is modified by cointroducing multiple Cas9 orthologue-based platforms with different effectors such as nuclease, transcription activator or repressor into the same cell, where they are guided by a specific group of gRNAs to carry out multiplex and complex manipulation of gene activities [30]

3. Application and recent advances

Today, the application of the CRISPR/Cas9 system-based genome editing for crop improvement has already begun its movement despite still being in its infancy. As a matter of fact, there are already numerous publications that reported that this technology has been successfully implemented in a broad range of plant species (as shown in **Table 1**).

The applications of the CRISPR/Cas9 system in plants can be classified into three types, which are the gene disruption, gene insertion and gene regulation.

3.1. Gene disruption

Gene disruption or simply known as gene knockout is a genetic technique that turns one of the genes in an organism to become inoperative. This technique is very powerful as it can inactivate any potential harmful or nonbeneficial gene that downgrades the quality of a plant. Gene disruption is the most applied technique as it can knockout genes by simply introducing small deletion or insertion via NHEJ repair mechanism in CRISPR/Cas9 system [5, 7].

A good example that employed full use of the gene knockout mechanism is the *Waxy* (*WX1*) gene of a maize plant. The maize *WX1* gene encodes a starch-synthesizing protein that is

Species name	Target gene(s)	Gene function	Description	Mode of action	Ref.
Arabidopsis thaliana	*BR11*, *JAZ1*, *GAI*	Growth regulators	Transgenic plants displayed retarded growth after being subjected to targeted mutagenesis	Gene disruption	[51]
Brassica oleracea	*BolC.GA4.a*	Gibberellin biosynthesis	Transgenic plants displayed dwarf phenotype after being subjected to targeted mutagenesis	Gene disruption	[52]
Citrus sinensis	*CsPDS*	Carotenoid biosynthesis	Transgenic plants displayed albinism expression after being subjected to targeted mutagenesis	Gene regulation	[53]
Cucumis sativus	*eIF4E*	Translation initiation factor	Transgenic plants developed resistance toward a broad range of virus	Gene disruption	[54]
Glycine max	*Bar*, *GmFE11*, *GmFE12*, etc.	Root hair growth factors	Transgenic plants displayed higher root hair growth induction after being subjected to targeted mutagenesis	Gene regulation	[55]
Hordeum vulgare	*HvPM19*	Grain dormancy regulator	Transgenic plants displayed signs of dormancy after being subjected to targeted mutagenesis	Gene disruption	[52]
Marchantia polymorpha	*ARF1*	Auxin response factor	Transgenic plants showed no response toward auxins after being subjected to targeted mutagenesis	Gene disruption	[56]
Medicago truncatula	*GUS*	Fluorescence	Transgenic plants displayed no signs of staining after being subjected to targeted mutagenesis	Gene disruption	[57]
Nicotiana benthamiana	*NbPDS*	Carotenoid biosynthesis	Transgenic plants displayed albinism expression after being subjected to targeted mutagenesis	Gene insertion	[58]
Nicotiana tabacum	*NtPDS*	Carotenoid biosynthesis	Transgenic plants displayed albinism expression after being subjected to targeted mutagenesis	Gene disruption	[59]

http://dx.doi.org/10.5772/intechopen.75024

Species name	Target gene(s)	Gene function	Description	Mode of action	Ref.
Oryza sativa	*OsPDS*, *OsMPK2*, *OsBADH2*, *etc.*	Carotenoid biosynthesis, growth regulator	Transgenic plants displayed albinism and dwarfism after being subjected to targeted mutagenesis	Gene disruption	[60]
Petunia hybrid	*PDS*	Carotenoid biosynthesis	Transgenic plants displayed albinism expression after being subjected to targeted mutagenesis	Gene disruption	[61]
Populus tomentosa	*PtoPDS*	Carotenoid biosynthesis	Transgenic plants displayed albinism expression after being subjected to targeted mutagenesis	Gene disruption	[62]
Solanum lycopersicum	*SlAGO7*	Involved in RNA biogenesis regulation	Transgenic plants displayed needle-like or lacking lamina leaves after being subjected to targeted mutagenesis	Gene disruption	[63]
Solanum tuberosum	*StALS1*	Acetolactate biosynthesis	Transgenic plants showed increased resistance on herbicides after being subjected to targeted mutagenesis	Gene insertion	[64]
Sorghum bicolor	*DsRED2*	Fluorescence	Transgenic plants showed signs of red fluorescence after being subjected to targeted mutagenesis	Gene insertion	[65]
Triticum aestivum	*TaINOX*, *TaPDS*	Inositol metabolism and carotenoid biosynthesis	Transgenic plants displayed albinism expression after being subjected to targeted mutagenesis	Gene disruption	[66]
Vitis vinifera	*IdnDH*	Tartaric acid biosynthesis	Transgenic plants showed no signs of tartaric acid in their fruits after being subjected to targeted mutagenesis	Gene disruption	[67]
Zea mays	*ZmIPK*	Phytic acid biosynthetic pathway catalyst	Transgenic plants showed reduction of phytic acid level after being subjected to targeted mutagenesis	Gene disruption	[68]

Table 1. List of CRISPR/Cas9 system-based genome-edited plants.

involved in the kernel maintenance [48]. Today, there is a known mutant maize that has a deletion in the coding sequence of the WX1 allele [49, 50] that causes it to have an altered grain starch composition [51]. Waxy corns are highly sought after in the commercial market because it provides a variety of benefit such as improved uniformity, stability and texture despite its lower yield compared to elite corns [52]. Up until recently, there have been attempts to introduce the mutant WX1 allele by crossbreeding a nonelite Waxy corn with an elite plant with excellent agronomic qualities. However, this method was unsuccessful as some of the nonelite alleles near the mutant *WX1* gene may be carried along during the introgression process in addition to increased time requirements [52].

Recently, an agricultural company known as DuPoint took this matter with an alternative solution through gene disruption by using the CRISPR/Cas9 system [53]. The gene disruption via CRISPR/Cas9 system is cheap, fast and, most importantly, precise as *WX1* deletions can now be generated directly in the genome of the elite plant to overcome the imperfections that are associated with trait introgression. The gene disruption via CRISPR/Cas9 system works by deleting the entire *WX1* gene with the usage of two Cas9-gRNAs. Each of the Cas9-gRNAs will target two sites, which are the upstream of the transcriptional start site and the downstream of the stop codon. Then, the region is excised and the remaining DNA damage is repaired through the NHEJ, which will bring about the *WX1* null allele with the Waxy phenotype [52].

Another study that utilized the ability of gene disruption of CRISPR/Cas9 system was carried out in wheat, an important staple food in many parts of the world [54]. The team reported that the *inositol oxygenase* (*INOX*) and *phytoene desaturase* (*PDS*) gene of the wheat plant was successfully deactivated at the same time, making it a multiplex mutagenesis. The application of the CRISPR/Cas9 system to the gene causes the gene to have random insertion into its sequence, resulting in gene disruption. Consequently, the phenotype of the wheat changes to express albinism or etiolated leaves.

3.2. Gene insertion

Gene insertion or addition is another famous technique where more than one nucleotide base pairs are added into a DNA sequence. The newly inserted sequence can be designed in such a way where it can specifically encode proteins that bring crucial benefits. However, designing the inserted-to-be DNA sequence is not an easy task as imperfection could pose a risk to the health of the host cell or organism, or it can be simply nulled as it fails to function as predicted [55]. In the terms of gene addition for crop improvement, the desired goal is usually the addition of beneficial traits such as pest resistance, high yield or quality.

In fact, adding multiple genes that confer different trait improvement in a single plant is a common practice to produce elite cultivar. For instance, the TC1507 maize that contains both the *Bacillus thuringiensis* (*Bt*) gene that confers insect resistance and *acetyltransferase* gene that is herbicide tolerance [56]. To date, the most common and efficient method used is to collocate all the desired genes into a single molecular stack, whereby all of the genes will now behave as a single locus or better known as gene stacking [52]. The colocation is no easy task as it has two major limitations. First, each of the genes may potentially affect each other if they are placed adjacently too close [57]. Second, once the transgenes are collocated, they can no longer be moved as they are placed too close to each other to the point where it acts as a single locus.

There is another unconventional substitute to molecular stacks where it can only be generated through the CRISPR/Cas9 system. That substitute is known as complex trait loci (CTL) or quantitative trait loci (QTL) and where transgenes can also be genetically collocated [58, 59]. An example of CTLs is constructed through the CRISPR/Cas9 system by specifically inserting the transgenes into the desired region in the genome through HDR. To start, the transgenes in the CTL can be separated by a larger distance (50 kb to more than 1 Mb) compared to the molecular stacks (few hundred or thousand bp) while retaining their genetic linkage [60]. The changes of distance solves both the limitations of the molecular stacks as adjacent transgene will no longer affect each of their function and they can now be individually moved and swapped.

Similarly, with the help of the CRISPR/Cas9 system, the DsRED2 gene, which encodes a protein that expresses red fluorescence, was also successfully inserted into the genome of an immature sorghum embryo [61]. As a result, the plant now displays red fluorescence.

3.3. Gene regulation

Gene regulation is a technique whereby the gene encoding for its transcription factors is altered to induce changes in its gene expression level [62]. Consequently, plant traits such as the fruit color, size and shape can be controlled and adjusted according to the consumer demands.

The CRISPR/Cas9 system can also be used to regulate the expression of genes for plants [63]. It was carried out by the usage of a catalytically inactive Cas9 known as dead Cas9 (dCas9). The deactivation occurs when rare bacteriophages with anti-CRISPR protein AcrIIA4 binds to the Cas9 of a gRNA that causes its cleaving activity to be disabled [64]. Consequently, the dCas9 is unable to cleave DNAs but it can still bind to specific DNA sequences with gRNA. To be used in gene regulation, the dCas9 must be fused with either a transcriptional activator or a repressor.

For transcriptional activation, dCas9 will be fused with a transcription activator domain such as VP64. For example, there is a study that reported that the paired dCas9-VP64 couple successfully activates the *anthocyanin pigment 1* (*AtPAP1*) gene from *Arabidopsis thaliana,* which encodes the protein involved in the production of anthocyanin pigment 1 [65]. Meanwhile, for transcriptional repression, dCas9 will be fused with a transcription repressor domain such as SRDX instead. Consistently, a study had reported the usage of dCas9-SRDX pair to successfully repress the *A. thaliana cleavage stimulating factor 64* (*AtCSTF64*) gene of a plant of the same species. This technique is still new compared to the previously mentioned gene disruption and gene addition techniques.

3.4. CRISPR/Cas9 system-based genome-edited plants

As the aforementioned plants are successfully genetically modified in the lab, there are actually some of them that are almost readily available in the commercial market. These plants may be new to the market but it is undeniable that they will eventually be able to monopolize the market as they have much more improved traits compared to their relative wild-type plants. As shown in **Table 2**, most of the plants such as the wheat and Ranger Russet potato are important food staples in many parts of the world and this proves that the CRISPR/Cas9 system-based genome editing for crop improvement is definitely on its way to revolutionize the agriculture industry.

Crop	Trait(s) improved	Status	Name of organization	Ref.
White button mushroom	Browning resistant	Submitting for review to Food and Drug Administration (FDA)	Yinong Yang; Penn State College of Agricultural Science	[74]
Waxy corn	Disease resistant Drought tolerant	To be marketed within 5–10 years, pending field trials and applicable regulatory review	DuPont Pioneer	[75]
Wheat	Produce gluten-free wheat by eliminating gliadins in wheat	Working with *gliadin* genes that are still present	Institute for Sustainable Agriculture	[76]
Soybean	Produce healthier oil with reduced unsaturated fat content by increasing the percentage of oleic acid	Inactivation of two genes in soybean	Institute for Basic Research (IBS)	[77]
Ranger Russet Potato	Longer freshness because it does not accumulate sweet sugars at typical cold storage temperature Does not produce acrylamide (carcinogen) when fried	To be grown and sold in 2019	Dan Voytas; Cellectis Plant Sciences	[78]

Table 2. List of CRISPR/Cas9 system-based genome-edited plants that are making their way to the commercial market.

4. Social acceptance and regulation

Genome editing with engineered nucleases (GEEN) has evolved as a highly specific and efficient tool for crop improvement with the potential to rapidly generate useful novel phenotypes. This leads to the emergence of new plant breeding technologies such as to allow the investigation of gene functions and inducing variations for crop improvement. Among these, CRISPR/Cas9 system is now one of the trending applications in plant breeding. Besides the CRISPR/Cas9 system, there are also other plant-breeding technologies that involve cis-genesis and intra-genesis such as transgenic development, whereby unspecific mutagenesis is induced by radiation or chemicals that are much faster and efficient than the conventional breeding method [66]. A question arises as to how genetically edited plants with desired traits will be received by the public and regulated within legislation on genetically modified organism (GMO). According to a recent survey comparing scientist and citizen views on a range of science, engineering and technology issues [67], the most pronounced difference obtained from the study was found on the question addressing the safety of consuming genetically engineered crops; whereby 37% of the public at large responded that GM foods are generally safe to eat, whereas 88% of scientists interviewed recognized GM foods as generally safe [66].

There are two sides to this discussion. Those who take the view that new plant breeding technique (NPBT) such as CRISPR/Cas9 system should be exempted from GMO legislation argued that the products are similar to the products generated from conventional breeding methods. The opponents contend that the process used to generate the plants is in fact genetically modified. As stated in the European Law, the definition of GMO means an organism with the

exception of human beings, in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination.' Europe Commission (EC) has stressed that the decision to include or exclude a technique from the scope of Directives 2001/18/EC and 2009/41/EC depends only on the interpretation of the definition of genetically modified organisms and genetically modified microorganisms and of the conditions for exemption provided for in the two directives [68]. There are regulators such as the German Consumer Protection Association, or also known as Verbraucherzentrale Bundesverband (VZBV) and Swedish scientists that call for the exclusion of such 'gene editing' from GMO regulation as long as such crops do not contain any 'foreign DNA' [69]. The evaluation is sophisticated because the definition of GMO under European Union (EU) law refers both to the characteristics of organisms and to the techniques used. To date, a clarifying legal opinion of the EC is still pending. Until the legal opinion is released, the legal status of living organisms as well as products deriving from NPBT approaches is unclear [70].

In the United State, the Coordinated Framework for Regulation of Biotechnology (CFRB) determined that it is the final product of genetic engineering that potentially poses a risk to human health and the environment, not the process by which the product is made [71]. The engineered products could be channeled to and handled by regulatory net involving Environmental Protection Agency (EPA), Food and Drug Administration (FDA) and US Department of Agriculture (USDA) depending on what category it falls into [71]. In April 2016, a CRISPR edited, nonbrowning mushroom emerged as the first CRISPR-derived product to be approved by USDA [72].

5. Future prospects

With all the studies done so far, it is undeniable that the CRISPR/Cas9 system is on its way to change the pace and course in the agriculture industry. Perfect plants that have high yield, quality and resistance toward any disease and pests will no longer be impossible with the dawn of this technology. Moreover, CRISPR/Cas9-based gene editing for plants will also be developed to the point where it can be used to replace any defective gene with a normal allele at its natural location. Consequently, all plants will now no longer need to be in danger from any traditional diseases as long as this technique is present and approved for human consumption.

There are still many uncertainties on the usage of plant genome editing. Therefore, in-depth studies are required to ensure this technology will have zero risks while gaining maximum benefits. Besides that, the idea of genome editing might also raise ethical questions from the public; these need to be adequately addressed by researchers and scientists that are well adept in genome engineering. Educational talk or workshop on genome editing should be given to nonscientists to ensure they understand the basics and benefits of this technology. More laws and regulations will also be required for the implementation to ensure CRISPR/Cas9 system is used responsibly without slowing down its development and research. Only when the CRISPR/Cas9 system is well understood and regulated, it will be possible for the application of this technology to be maximized to its fullest potential to achieve previously envisioned ideas in plant science.

Acknowledgements

The authors thank Putra Grant (GP-IPM/2015/9450800) from Universiti Putra Malaysia and Fundamental Research Grant Scheme (FRGS/1/2015/ST03/UPM/02/5) from Ministry of Higher Education, Malaysia, for the support.

Author details

Kah-Yung Bernard Leong[1], Yee-Han Chan[1], Wan Muhamad Asrul Nizam Wan Abdullah[2], Swee-Hua Erin Lim[3,4] and Kok-Song Lai[2]*

*Address all correspondence to: laikoksong@upm.edu.my

1 Faculty of Science, University Tunku Abdul Rahman, Jalan Universiti, Bandar Barat, Perak, Malaysia

2 Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, Selangor, Malaysia

3 Perdana University-Royal College of Surgeons in Ireland, Perdana University, MAEPS Building, Selangor, Malaysia

4 Health Sciences Division, Abu Dhabi Women's College, Higher Colleges of Technology, Abu Dhabi, United Arab Emirates

References

[1] Abdallah NA, Prakash CS, McHughen AG. Genome editing for crop improvement: Challenges and opportunities. GM Crops Food. 2015;**6**(4):183-205. DOI: 10.1080/21645698.2015.1129937

[2] Bibikova M, Golic M, Golic KG, Carroll D. Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. Genetics. 2002;**161**(3):1169-1175

[3] Kim H, Kim J-S. A guide to genome engineering with programmable nucleases. Nature Reviews. Genetics. 2014;**15**(5):321-334. DOI: 10.1038/nrg3686

[4] Puchta H, Fauser F. Synthetic nucleases for genome engineering in plants: Prospects for a bright future. Plant Journal. 2014;**78**:727-741. DOI: 10.1111/tpj.12338

[5] Puchta H. The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. Journal of Experimental Botany. 2005;**56**(409):1-14. DOI: 10.1093/jxb/eri025

[6] Voytas DF. Plant genome engineering with sequence-specific nucleases. Annual Review of Plant Biology. 2013;**64**(1):327-350. DOI: 10.1146/annurev-arplant-042811-105552

[7] Gaj T. ZFN, TALEN and CRISPR/Cas based methods for genome engineering. Trends in Biotechnology. 2013;**31**(7):397-405. DOI: 10.1016/j.tibtech.2013.04.004

[8] Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th ed. New York: Garland Science; 2002. Site Specific Recombination. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26845/

[9] Carroll D. Genome engineering with zinc-finger nucleases. Genetics. 2011;**188**(4): 773-782. DOI: 10.1534/genetics.111.131433

[10] Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nature Reviews. Genetics. 2010 Sep;**11**(9):636-646. DOI: 10.1038/nrg2842

[11] Wyman C, Kanaar R. DNA double-strand break repair: All's well that ends well. Annual Review of Genetics. 2006;**40**(1):363-383. DOI: 10.1146/annurev.genet.40.110405.090451

[12] Pavletich N, Pabo C. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 A. Science (80-). 1991;**252**(5007):809-817

[13] Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I. FokI dimerization is required for DNA cleavage. Biochemistry. 1998;**95**:10570-10575

[14] Li L, Wu LP, Chandrasegaran S. Functional domains in Fok I restriction endonuclease. Proceedings of the National Academy of Sciences of the United States of America. 1992;**89**(10):4275-4279

[15] Sugisaki H, Kanazawa S. New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). Gene. 1981 Dec;**16**(1-3):73-78

[16] Smith J. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Research. 2000;**28**(17):3361-3369

[17] Carroll D, Morton JJ, Beumer KJ, Segal DJ. Design, construction and in vitro testing of zinc finger nucleases. Nature Protocols. 2006;**1**(3):1329-1341. DOI: 10.1038/nprot.2006.231

[18] Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, et al. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nature Protocols. 2006;**1**(4):1637-1652. DOI: 10.1038/nprot.2006.259

[19] Gonzalez B, Schwimmer LJ, Fuller RP, Ye Y, Asawapornmongkol L, Barbas CF. Modular system for the construction of zinc-finger libraries and proteins. Nature Protocols. 2010;**5**(4):791-810. DOI: 10.1038/nprot.2010.34

[20] Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nature Methods. 2011;**8**(1):67-69. DOI: 10.1038/nmeth.1542

[21] Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Molecular Cell. 2008;**31**(2):294-301. DOI: 10.1016/j.molcel.2008.06.016

[22] Kim S, Lee MJ, Kim H, Kang M, Kim J-S. Preassembled zinc-finger arrays for rapid construction of ZFNs. Nature Methods. 2011;**8**(1):7-7. DOI: 10.1038/nmeth0111-7a

[23] Gupta A, Christensen RG, Rayla AL, Lakshmanan A, Stormo GD, Wolfe SA. An optimized two-finger archive for ZFN-mediated gene targeting. Nature Methods. 2012;**9**(6):588-590. DOI: 10.1038/nmeth.1994

[24] Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science (80-). 2009;**326**(5959):1509-1512. DOI: 10.1126/science.1178811

[25] Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science (80-). 2009;**326**(5959):1501-1501. DOI: 10.1126/science.1178817

[26] Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. Nature Biotechnology. 2012;**30**(5):460-465. DOI: 10.1038/nbt.2170

[27] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-). 2012;**337**(6096):816-821. DOI: 10.1126/science.1225829

[28] Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science (80-). 2014;**346**(6213):1258096-1258096. DOI: 10.1126/science.1258096

[29] Bortesi L, Fischer R. The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnology Advances. 2015;**33**(1):41-52. DOI: 10.1016/j.biotechadv.2014.12.006

[30] Zhang D, Li Z, Li JF. Targeted gene manipulation in plants using the CRISPR/Cas technology. Journal of Genetics and Genomics. 2016;**43**(5):251-262. DOI: 10.1016/j.jgg.2016.03.001

[31] Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012;**482**(7385):331-338. DOI: 10.1038/nature10886

[32] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier EA. Programmable Dual-RNA—Guided DNA endonuclease in Adaptice bacterial immunity. Science. 2012;**337**(August):816-822. DOI: 10.1126/science.1225829

[33] Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014;**156**(5):935-949. DOI: 10.1016/j.cell.2014.02.001

[34] Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;**507**(7490):62-67. DOI: 10.1038/nature13011

[35] Miller JC, Holmes MC, Wang J, Guschin DY, Lee Y-L, Rupniewski I, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nature Biotechnology. 2007;**25**(7):778-785. DOI: 10.1038/nbt1319

[36] Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;**339**(6121):823-826. DOI: 10.1126/science.1232033

[37] Cho SW, Kim S, Kim JM, Kim J-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nature Biotechnology. 2013;**31**(3):230-232. DOI: 10.1038/nbt.2507

[38] Miao J, Guo D, Zhang J, Huang Q, Qin G, Zhang X, et al. Targeted mutagenesis in rice using CRISPR-Cas system. Cell Research. 2013 Oct 3;**23**(10):1233-1236. DOI: 10.1038/cr.2013.123

[39] Zhou H, Liu B, Weeks DP, Spalding MH, Yang B. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Research. 2014 Sep 29;**42**(17):10903-10914. DOI: 10.1093/nar/gku806

[40] Yeadon J. Pros and cons of ZNFs, TALENs, and CRISPR/Cas [Internet]. 2014. Available from: https://www.jax.org/news-and-insights/jax-blog/2014/march/pros-and-cons-of-znfs-talens-and-crispr-cas [Accessed: 2017-11-01]

[41] Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nature Protocols. 2014 Jul 24;**9**(8):1956-1968. DOI: 10.1038/nprot.2014.134

[42] Yan W, Chen D, Kaufmann K. Efficient multiplex mutagenesis by RNA-guided Cas9 and its use in the characterization of regulatory elements in the AGAMOUS gene. Plant Methods. 2016;**12**:23. DOI: 10.1186/s13007-016-0125-7

[43] Ma X, Liu Y-G. CRISPR/Cas9-based multiplex genome editing in monocot and dicot plants. In: Current Protocols in Molecular Biology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2016. pp. 31.6.1-31.6.21. DOI: 10.1002/cpmb.10

[44] Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie. 2015;**117**:119-128. DOI: 10.1016/j.biochi.2015.03.025

[45] Ma X, Zhu Q, Chen Y, Liu YG. CRISPR/Cas9 platforms for genome editing in plants: Developments and applications. Molecular Plant. 2016;**9**(7):961-974. DOI: 10.1016/j.molp.2016.04.009

[46] Yin K, Gao C, Qiu J-L. Progress and prospects in plant genome editing. Nature Plants. 2017;**3**(8):17107. DOI: 10.1038/nplants.2017.107

[47] Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V. Editing plant genomes with CRISPR/Cas9. Current Opinion in Biotechnology. 2015;**32**:76-84. DOI: 10.1016/j.copbio.2014.11.007

[48] Nelson OE, Rines HW. The enzymatic deficiency in the waxy mutant of maize. Biochemical and Biophysical Research Communications. 1962 Oct 31;**9**(4):297-300

[49] Collins G. A New Type of Indian Corn From China. 1st ed. Bureau of Plant Industry, editor. Washington: Government Printing Office; 1909. 1-30 p

[50] Wessler S, Tarpley A, Purugganan M, Spell M, Okagaki R. Filler DNA is associated with spontaneous deletions in maize. Proceedings of the National Academy of Sciences of the United States of America. 1990 Nov 1;**87**(22):8731-8735

[51] Sprague GF, Brimhall B, Hixon RM. Some effects of the waxy gene in corn on properties of the endosperm Starch1. Agronomy Journal. 1943;**35**(9):817

[52] Chilcoat D, Liu ZB, Sander J. Use of CRISPR/Cas9 for Crop Improvement in Maize and Soybean. Progress in Molecular Biology and Translational Science. Vol. 149. 1st ed. Elsevier Inc.; 2017. 27-46 p. DOI: 10.1016/bs.pmbts.2017.04.005

[53] Steenhuyse L Van. DuPont Pioneer's Next Generation of Waxy Corn Shows the Green Side of CRISPR-Cas9 [Internet]. 2017. Available from: https://biovox.eu/insights/detail/dupont-pioneer-rsquo-s-next-generation-of-waxy-corn-shows-the-green-side-of-crispr-cas9 [Accessed: 2017-11-14]

[54] Upadhyay SK, Kumar J, Alok A, Tuli R. RNA-guided genome editing for target gene mutations in wheat. G3 (Bethesda). 2013 Dec 9;**3**(12):2233-2238. DOI: 10.1534/g3.113.008847

[55] Papapetrou EP, Schambach A. Gene insertion into genomic safe harbors for human gene therapy. Molecular Therapy. 2016 Apr;**24**(4):678-684. DOI: 10.1038/mt.2016.38

[56] Baktavachalam GB, Delaney B, Fisher TL, Ladics GS, Layton RJ, Locke ME, et al. Transgenic maize event TC1507: Global status of food, feed, and environmental safety. GM Crops Food. 2015 Apr 3;**6**(2):80-102. DOI: 10.1080/21645698.2015.1054093

[57] Eszterhas SK, Bouhassira EE, Martin DIK, Fiering S. Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. Molecular and Cellular Biology. 2002 Jan 15;**22**(2):469-479. DOI: 10.1128/MCB.22.2.469-479.2002

[58] Holland JB. Genetic architecture of complex traits in plants. Curr Opin Plant Biol. 2007 Apr 1;**10**(2):156-161

[59] Collard BCY, Jahufer MZZ, Brouwer JB, ECK P. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. 2005:169-196. DOI: 10.4236/ajps.2014.59125

[60] Lowe K, Wu E, Wang N, Hoerster G, Hastings C, Cho M-J, et al. Morphogenic regulators baby boom and Wuschel improve monocot transformation. The Plant Cell. 2016 Sep 6;**28**(9):1998-2015. DOI: https://doi.org/10.1105/tpc.16.00124

[61] Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Research. 2013 Nov 1;**41**(20):e188-e188. DOI: 10.1093/nar/gkt780

[62] Liu X, Wu S, Xu J, Sui C, Wei J. Application of CRISPR/Cas9 in plant biology. Acta Pharmaceutica Sinica B. 2017;**7**(3):292-302. DOI: 10.1016/j.apsb.2017.01.002

[63] Piatek A, Ali Z, Baazim H, Li L, Abulfaraj A, Al-Shareef S, et al. RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. Plant Biotechnology Journal. 2015;**13**(4):578-589. DOI: 10.1111/pbi.12284

[64] Shin J, Jiang F, Liu J-J, Bray NL, Rauch BJ, Baik SH, et al. Disabling Cas9 by an anti-CRISPR DNA mimic. Science Advances. 2017 Jul;**3**(7):e1701620. DOI: 10.1126/sciadv.1701620

[65] Lowder LG, Zhang D, Baltes NJ, Paul JW, Tang X, Zheng X, et al. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiology. 2015 Oct 1;**169**(2):971-985. DOI: 10.1104/pp.15.00636

[66] Wolt JD, Wang K, Yang B. The Regulatory Status of Genome-edited Crops. 2016;**4**(October 2015):510-518. DOI: 10.1111/pbi.12444

[67] Funk C, Lee R. Public and Scientists' Views on Science and Society | Pew Research Center [Internet]. 2015. Available from: http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/ [Accessed: 2017-12-17]

[68] Laaninen T. New Plant-Breeding Techniques Applicability of GM Rules [Internet]. 2016. Available from: http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/582018/EPRS_BRI(2016)582018_EN.pdf [Accessed: 2017-12-17]

[69] Ammann K. How Should Europe Regulate CRISPR, New Breeding Technologies (NBTs [Internet]). 2016. Available from: https://geneticliteracyproject.org/2016/01/15/europe-regulate-crispr-new-breeding-technologies-nbts/ [Accessed: 2017-12-15]

[70] Sprink T, Eriksson D, Schiemann J, Hartung F. Regulatory hurdles for genome editing: Process- vs. product-based approaches in different regulatory contexts. Plant Cell Reports. 2016;**35**(7):1493-1506. DOI: 10.1007/s00299-016-1990-2

[71] Kuzma J. Policy: Reboot the debate on genetic engineering. Nature. 2016;**531**(7593):165-167. DOI: 10.1038/531165a

[72] Waltz E. Gene-edited CRISPR mushroom escapes US regulation. Nature. 2016;**532**(7599):293-293. DOI: 10.1038/nature.2016.19754

[73] Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. Journal of Clinical Investigation. 2014 Oct 1;**124**(10):4154-4161

[74] Gill C. Penn State Developer of Gene-Edited Mushroom Wins "Best of What"s New' award | Penn State University [Internet]. Penn State News. 2016 [cited 2018 Mar 3]. Available from: http://news.psu.edu/story/432734/2016/10/19/academics/penn-state-developer-gene-edited-mushroom-wins-best-whats-new?utm_source=newswire&utm_medium=email&utm_term=433007_HTML&utm_content=10-19-2016-21-27&utm_campaign=Penn+State+Today

[75] Sander J, Jeschke M. CRISPR-Cas Advanced Plant Breeding [Internet]. DuPont Pioneer. 2016 [cited 2018 Mar 3]. Available from: https://www.pioneer.com/home/site/us/agronomy/library/crispr-cas/

[76] Dormehl L. CRISPR's Latest Achievement? Taking the Gluten Out of Wheat [Internet]. Digital Trends. 2017 [cited 2018 Mar 3]. Available from: https://www.digitaltrends.com/cool-tech/crispr-gene-editing-low-gluten/

[77] Kim H, Kim S-T, Ryu J, Kang B-C, Kim J-S, Kim S-G. CRISPR/Cpf1-mediated DNA-free plant genome editing. Nat Commun. 2017 Feb 16;**8**:14406

[78] Regalado A. A Potato Made with Gene Editing - MIT Technology Review [Internet]. MIT Technology Review. 20115 [cited 2018 Mar 3]. Available from: https://www.technologyreview.com/s/536756/a-potato-made-with-gene-editing/

Following the green revolution and transgenic crop development, another revolutionary progress has been experienced in plant breeding in the last decade with the application of marker-assisted selection (MAS), next-generation sequencing (NGS), and gene editing techniques together with omic technologies, including genomics, transcriptomics, proteomics, and phenomics. Thus, this book is structured into two sections: “Marker-Assisted Breeding” and “RNA-seq and Gene Editing in Plants,” which aim to provide a reference for students, instructors, and scientists on recent innovative advances in plant-breeding programs to cultivate crops for tomorrow.

Published in London, UK

© 2018 IntechOpen
© Tods0859 / iStock

IntechOpen