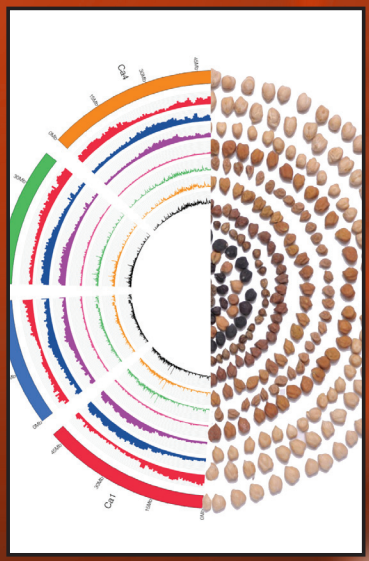
Translational Genomics for Crop Breeding

Volume II: Abiotic Stress, Yield and Quality

Editors: **Rajeev K. Varshney**
Roberto Tuberosa



WILEY Blackwell

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality

Edited by

Rajeev K. Varshney
Roberto Tuberosa

WILEY Blackwell

This edition first published 2013 © 2013 by John Wiley & Sons, Inc.

Editorial offices: 2121 State Avenue, Ames, Iowa 50014-8300, USA
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Translational genomics for crop breeding. Volume II : abiotic stress, yield and quality / edited by Rajeev K. Varshney and Roberto Tuberosa.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-96291-6 (cloth : alk. paper) – ISBN 978-1-118-72837-6 (epub) – ISBN 978-1-118-72842-0 (emobi) – ISBN 978-1-118-72848-2 (ebook) – ISBN 978-1-118-72862-8 (epdf) 1. Crops–Genetic engineering. 2. Plant breeding. 3. Plant genome mapping. 4. Crop improvement. I. Varshney, R. K. (Rajeev K.), 1973– II. Tuberosa, R. (Roberto)
SB123.57.T74 2013
631.5′3–dc23

2013029686

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover images: Background © HelenStocker; Cornfield © jess311; Corn on the cob © Diane Diederich
Cover design by Matt Kuhns

Set in 10/12.5pt Times by Aptara® Inc., New Delhi, India

1 2013

Contents

Foreword

ICRISAT's mission is to reduce poverty, hunger, malnutrition, and environmental degradation in the dryland tropics. To accomplish this, we need to address the challenges presented by population explosion and climate change and their impacts on agriculture.

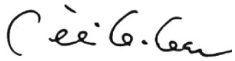
Today about 70% of the food-insecure population lives in developing countries, mostly as small-scale and subsistence farmers, and their population growth is expected to triple by 2100. They eke out a living and feed themselves from food crops cultivated in degraded lands in an unequivocally warmer climate system.

To achieve global food security, the development of crop varieties that produce high yields in harsh climatic conditions will be a key strategy. Although several desirable traits have been developed in crop species through integrated breeding practices, self-sufficiency in food grains and legumes *per se* has remained an elusive dream for poor people in the developing world.

The knowledge generated through advances in genomics during the past two decades has enormous potential in advancing the quest for abiotic stress–tolerant crops in the arid and semi-arid regions of the world. Moreover, DNA-based marker technologies have increased the precision in marker-assisted selection (MAS), thereby reducing the time for improving traits of interest. Genome sequence information for important crops like rice, sorghum, maize, soybean, chickpea, pigeonpea, tomato, and so on is now available to enable greater understanding of traits through comparative studies. It is important to further translate available genome information in crop breeding so that farming communities will be benefitted sooner than later.

Translational Genomics for Crop Breeding: Abiotic Stress, Yield and Quality, edited by Dr. Rajeev K. Varshney, our own ICRISAT scientist, and Professor Roberto Tuberosa from the University of Bologna, Italy, provides a concrete step in this direction. It is the second of two volumes where the eminent editors have carefully selected authors who are experts in translational genomics in crop breeding for different traits in different crop species to write the various chapters of this publication. These chapters provide examples of translational genomics for enhancing tolerance to abiotic stresses and quality traits in a number of crops. I believe that such a book is very timely, informative, and of such quality that it will fill the gap that exists presently between genome science and crop breeding.

Through this publication, we hope the elusive dream of poor people in the developing world will sooner become a reality.



Hyderabad
Date: June 10, 2013

William D. Dar
Director General,
ICRISAT

Preface

Ever since the discovery about four decades ago of the first molecular tools to investigate DNA structure and function, the science of genomics has contributed tremendously to investigate the genetic and functional basis of plant phenotypes and how this variability affects crop productivity. More recently, further advances in high-throughput genotyping technologies and genome sequencing have accelerated gene discovery and allele mining and the application of this knowledge to crop improvement. This aspect of translational genomics has also been referred to as genomics-assisted breeding (GAB). Due to the continuous and coordinated efforts of the global scientific community involved in crop improvement, in a number of cases GAB has become an integral part of crop-breeding programs.

Although substantial efforts have been made in the past century to mitigate the negative effects of abiotic stresses on crop productivity and improve the nutritional quality of crops, the progress achieved through conventional breeding approaches has been limited and will be insufficient to meet the ever-growing needs of humankind in the next decades. Within this daunting scenario, the advances in genomics during the past decades provide new avenues for enhancing our understanding of the genetic basis of complex traits while accelerating crop improvement. While a good number of success stories have been completed or are in progress for resistance to biotic stresses, GAB has limited examples of success stories for the improvement of resistance to abiotic stresses, quality, and yield *per se.* In view of the fast-growing need for developing new cultivars with enhanced tolerance to abiotic stresses, better quality and/or higher yield, this volume compiles reviews from leading authorities in their own fields of expertise while describing the progress and presenting ideas for future applications.

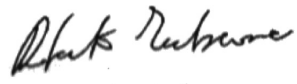
The editors thank all the authors of the different chapters (Appendix I) for the skillful summaries of the research work in their area of expertise and in some cases for sharing unpublished results in order to make the articles as up to date as possible. We also appreciate their cooperation in meeting the deadlines and in revising their manuscripts. While editing this book, the editors also received strong support from several colleagues (Appendix II), who willingly reviewed the manuscripts. Their constructive and critical suggestions have been very useful for improving the quality of manuscripts.

We are also grateful to colleagues and staff from our respective laboratories, who have helped us complete the editing of this volume in parallel with their demanding responsibilities. In particular, Manish Roorkiwal, B. Manjula, Pawan Khera, and Mahendar Thudi helped RKV with the editorial work. RKV is thankful to his wife Monika for her constant encouragement and support, and to Prakhar (son) and Preksha (daughter) for their love and cooperation. Similarly, RT is thankful to his wife Kay for her patience and editorial contributions. RKV would also like to extend his sincerest thanks to Dr. William D. Dar, Director General, ICRISAT,

for his guidance and support in completing this book. The cooperation and help received from Justin Jeffryes, Anna Ehler, Kelvin Matthews, Erin Topp of Wiley Blackwell, and Shikha Sharma of Aptara Corp. during various stages of development and completion of this book are also gratefully acknowledged. RKV would also like to mention that the book was edited during the tenure of RKV as Director, Center of Excellence in Genomics (CEG), ICRISAT, Hyderabad (India), Theme Leader – Comparative and Applied Genomics (CAG), Generation Challenge Programme (GCP) and Adjunct positions at the University of Western Australia, Crops Research Institute of Guangdong Academy of Agricultural Sciences (GAAS), China and BGI-Hongkong Research Institute, China.

We hope that this book will be helpful and useful as a ready guide to students, young researchers, crop specialists, breeders, and policy makers for contributing to the development of new cultivars more resilient to abiotic stresses and with a better nutritional quality. Lastly, we would appreciate if the readers can point out any errors and give comments/suggestions, as this would be useful for the future, revised and updated editions.

Hyderabad, India (Rajeev K. Varshney)
June 09, 2013



Bologna, Italy (Roberto Tuberosa)
June 09, 2013

Chapter 1

Translational Genomics for Crop Breeding: Abiotic Stress Tolerance, Yield, and Quality, An Introduction

Rajeev K. Varshney and Roberto Tuberosa

Abstract

In the context of global climate change and population explosion, feeding the world's population and addressing the issues of malnutrition, especially in developing countries, are daunting tasks before the global scientific community. The yield gains achieved through conventional breeding are not very promising, as several abiotic stresses such as drought, salinity, cold, flooding, submergence, and mineral toxicity have been leading to significant yield losses and reducing the quality of produce. In recent years, advances in genomics research and next generation sequencing (NGS) technologies have largely facilitated understanding and identifying gene networks that are involved in controlling genetic variation for agronomically valuable traits in elite breeding populations. The availability of genome-sequence information, transcriptomic resources, molecular markers, and genetic maps for major crops such as rice, maize, and sorghum have enabled adoption of genomics-assisted breeding (GAB) approaches, including marker-assisted backcrossing (MABC) and marker-assisted recurrent selection (MARS). Nevertheless, during the last decade significant genomic resources were also developed in less-studied crops and efforts are underway in deploying these genomic tools in breeding. Furthermore, the new bioinformatics approaches and decision-support tools developed are able to enhance the precision in selection and complement the success of GAB approaches.

This volume essentially focuses on the research on abiotic stress tolerance and the quality enhancement of agricultural produce. Further, this introductory chapter summarizes the key success stories and lessons learned in the field of genomics tools for crop improvement. In addition, this chapter also emphasizes the essence of deploying genome-wide association mapping and nested association mapping (NAM), as well as genomic selection (GS) approaches for crop improvements, in the context of the availability of a plethora of low-cost and high-throughput sequencing technologies.

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

Introduction

Despite continuous efforts to improve agricultural crops, changes in climate in the past two decades have had a tremendous influence on crop production and productivity. Further, climate changes will have a significant impact on the food security of humankind, especially in developing countries (Lake et al. 2012). It is expected that global temperatures will increase about 3°C by 2100 (Schneider et al. 2007), a change that would drastically curtail global crop production. Additionally, other critical abiotic stresses such as drought, salinity, cold, flooding, submergence, mineral toxicity, and so forth hamper the growth, development, yield, and seed quality of crops. In fact, these abiotic stresses represent the main cause of crop failure worldwide, curtailing average yields of all major crops by more than 50%. Furthermore, the quantitative inheritance and low heritability of resistance to these stresses, coupled with a strong Genotype x Environment x Management interaction of the yield response of crops to abiotic stresses, greatly limit a more accurate dissection and manipulation of such response. Since the world population is increasing at an alarming rate, minimizing these losses is also a major concern for all nations, particularly those with a strong increase in food demand. Besides increasing the production potential, the nutritional quality of the produce needs to be improved to avoid the malnutrition that billions are already facing, particularly in developing countries (Müller and Krawinkel 2005; Bouis et al. 2010).

In the context of rapidly growing demand for the staples of our sustenance, conventional breeding programs are struggling to achieve the yield gain required to adequately meet the burgeoning demand for food and plant-derived products (Tester and Langridge 2010). Accordingly, genomics-assisted breeding (GAB, Varshney et al. 2005) approaches are increasingly being adopted to improve the accuracy and effectiveness of selection while allowing for the dissection of the traits controlling the adaptive response of crops to unfavorable conditions. While availability of molecular markers and genetic maps are the prerequisites for GAB, several orphan crops, neglected at the global level but important for food security at the regional level, until recently lacked the genomic resources and platforms to implement GAB. However, in recent years significant progress has been achieved in the development of genomic resources in a number of orphan crops that have thus become genomic resource-rich crops (Varshney et al. 2009, 2010). As a result, GAB activities including trait mapping and marker-assisted selection (MAS) methods, such as marker-assisted backcrossing (MABC) and marker-assisted recurrent selection (MARS), are increasingly being adopted in breeding programs for major crops and have begun to be deployed in less-studied crops as well (Kulwal et al. 2012; Varshney et al. 2012).

Volume I of this book series presents reviews of genomics applications in crop breeding for biotic stress resistance, while this volume (Volume II) focuses on research endeavors in abiotic stress tolerance and the enhancement of the quality of agricultural produce. Four chapters in Volume II deal with ongoing research on tolerance to submergence, salinity, drought, and phosphorus (P) deficiency in rice. Another chapter discusses work on cloning and molecular breeding work for aluminum toxicity in sorghum. Research on freezing tolerance, an important trait in the Western world, in wheat and barley is summarized in another chapter. One chapter is focused on molecular breeding efforts for stay-green tolerance, an important drought tolerance trait in sorghum. Four chapters in the volume are devoted to quality improvement traits in rice, maize, peanut, and tomato. In the last chapter, authors discuss advances in sugarcane genomics and its applications for enhancing yields and ongoing efforts in genomic selection (GS). Some highlights of these chapters have been summarized in this introductory chapter.

Enhancing Tolerance to Abiotic Stresses in Rice

Submergence stress affects more than 15 Mha in lowland rice-growing areas of South and Southeast Asia. Chapter 2 provides insights into the ongoing efforts at the International Rice Research Institute (IRRI), Manila, Philippines, to improve submergence tolerance in rice. Following the identification of *Sub1* (*Submergence1*) locus and three ethylene responsive factors (ERFs) in rice, Septiningsih and colleagues report the development of eight *Sub1* varieties by the IRRI, six of which are already widely grown in several countries.

More than 444 Mha of global rice-growing area is affected by soil salinization (FAO, 2010). Soil salinization is a major problem in coastal areas of the regions where rice-based farming predominates. Reportedly rice yields are reduced by up to 50% when grown under moderate (6 dS/m) salinity levels (Ren et al. 2005). The losses due to soil salinization can be overcome by soil reclamation or by improving salinity tolerance in the crops. Efforts toward understanding the genetic basis of the trait for crop improvement has revealed that several genes are independently involved in salinity tolerance at different stages of crop cycles. In Chapter 3, Platten and colleagues provide an overview of genomics applications in enhancing salt tolerance in rice.

Drought is the major limiting factor to crop production, and cereals especially experience various kinds of drought stresses, depending on the timing and intensity of the water stress relative to the reproductive stage of the crop. In the case of rice, in Asia about 34 Mha of rain-fed lowland rice and 8 Mha of upland rice (Huke and Huke 1997) are frequently subjected to drought stress. Progress in developing high-yielding, drought-tolerant rice cultivars by conventional breeding has been slow, largely because of difficulties in precisely defining the target environment, complex interactions of drought tolerance with environments and management practices, and lack of appropriate screening methodology. However, during the past decade the availability of large-scale genomic resources and genome sequences have enabled the adoption of various GAB approaches in rice (see Collard et al. 2008). These efforts are summarized in Chapter 4 by Kumar and colleagues.

Sixteen essential elements are required for rice during the crop cycle. The major nutrients such as nitrogen (N), phosphorus (P), and potassium (K) are largely supplied as chemical fertilizers. The excess application of P, owing to its insoluble nature, leads to deficiencies of copper (Cu), iron (Fe), manganese (Mn), and zinc (Zn). Additionally, erosion of P-enriched soils enhances eutrophication in fresh water (Wolf 1996). Most of the rain-fed rice grown in Asia and Africa is cultivated on problematic soils, especially when P becomes unavailable to the crop as it adheres to soil particles. Hence, in this context, development of crops with enhanced efficiency of P utilization and production of higher biomass is essential. Chapter 5 by Heuer and colleagues essentially discusses the issues related to P deficiency in rice production. Further, the authors also highlight the need for the adoption of molecular breeding approaches and summarize the molecular breeding efforts for enhancing P utilization efficiency in rice.

Enhancing Tolerance to Abiotic Stresses in Wheat and Barley

Freezing/cold tolerance in crop plants is most important in the context of global climate change. Freezing tolerance is important in temperate cereals such as wheat (*Triticum* spp.), barley (*Hordeum vulgare*), and rye (*Secale cereale*). Long exposures of winter wheat and barley varieties to non-freezing cold temperatures (Dhillon et al. 2010) will accelerate flowering time (vernalization) and improve freezing tolerance (cold acclimation). In the case of wheat, Kobayashi et al. (2005) reported that *Vrn-Fr1* controls both frost tolerance and vernalization. Chapter 6, by Gabor and colleagues, reports on the developmental plasticity of these *Triticeae* crops

upon onset of the cold stress. This chapter also summarizes the accumulated knowledge of the past 20 years in the area of genetics and genomics of the mechanism of freezing tolerance and the genomic tools available for enhancing freezing tolerance in the *Triticeae* crops.

Enhancing Tolerance to Abiotic Stresses in Sorghum

Aluminum (Al) is the third most abundant element on earth after oxygen and silicon (Ma et al. 2001). It is a light metal that makes up 7% of the earth's crust. Half the arable soils across the globe and especially those in Africa, Asia, and South America are affected by aluminum toxicity (http://www.news.cornell.edu/stories/Aug07/SoilsKochian.kr.html). Chapter 7 by Magalhaes and colleagues deals with the existing diversity with respect to aluminum tolerance in sorghum and maize germplasm accessions as well as the molecular, physiological, and genetic basis of Al tolerance in both crops. The authors provide insights into the structure and functional analysis of membrane transporters such as Al-activated malate transporter (ALMT1) and multidrug and toxic compound efflux (MATE) involved in Al tolerance.

Furthermore, 20 to 30% of production losses in sorghum are due to lodging. Stay-green trait, however, has been an indirect selection criterion used by breeders for enhancing lodging resistance. Stay green is associated with increased grain yield and grain size in the sorghum crop under terminal drought, a common occurrence in arid and semiarid regions across the globe (Jordan et al. 2012). Among several genotypes identified with the stay-green trait, BTX645 has been a useful resource in developing commercial hybrids (Harris et al. 2007). Four major quantitative trait loci (QTLs) and several minor QTLs can enhance stay-green traits and several efforts are underway to introgress these QTLs into various genetic backgrounds. However, Vadez and colleagues in Chapter 8 report that this undertaking has been quite challenging owing to limited polymorphism among the parental lines for this trait. The physiological, genetic, and molecular breeding aspects for the stay-green trait are discussed at length in this chapter.

Improving Quality and Yield Through Molecular Breeding in Rice, Maize, Peanut, and Sugarcane

Besides increasing production and productivity, agricultural produce also should fulfill the requirement of consumer acceptance in terms of quality, in order to fetch a good market price. Hence, grain quality improvement also forms the major concern for cereal breeders. In Chapter 9, Hori and Yano describe rice grain quality traits in terms of physical and cooking qualities that are of interest to the consumer. In fact, grain quality is influenced by climate changes, such as high temperatures at the grain ripening stage, and grain components such as amylose, amylopectin, and proteins are greatly affected by such changes. This chapter summarizes the efforts to understand the genetics of grain quality traits and the multiple genes/QTL contributing to grain quality. The chapter also emphasizes the need for developing novel quality evaluation instruments/approaches, such as TILLING (Target Induced Local Lesion IN Genome) that can enhance the qualities related to the cooking and eating of *japonica* rice.

About 190 million children under the age of five years are suffering from malnutrition, especially in the underdeveloped and developing countries in Asia and sub-Saharan Africa (WHO, 2009). Malnutrition can be overcome by supplementing dietary requirements with micronutrients or through promotion of dietary diversification. However, these strategies have been only partially adopted and appear not to have improved the nutrient deficiencies in South African children since 1994 (Labadarios et al. 2005). Nevertheless "biofortification" – breeding staple crops with increased nutritional value – has emerged as a potential long-term strategy to improve nutritional security. Babu and

colleagues in Chapter 10 provide a comprehensive overview on biofortification in maize and highlight two specific cases of genetic improvement in maize that resulted in high nutritional value, particularly with respect to essential amino acid content in the endosperm. Besides emphasizing the molecular marker-assisted QPM (quality protein maize) breeding, the chapter also throws light on its impact in the developing world. Furthermore, the essence of provitamin A, Fe, and Zn, and low-phytate content, and the possibilities for genetically engineered, high-lysine maize are elegantly discussed in the chapter.

Peanut is the most important food legume and oilseed crop cultivated in arid and semi-arid regions of the world. About 45-51% of the dry weight of peanut seeds is oil (Chamberlin et al. 2011). Among fatty acids, oleic and linoleic acids are major fatty acids that determine the oil quality; hence the ratio of the two, the O/L ratio, is critical. In Chapter 11, molecular breeding efforts aimed at improving the oil quality in peanut, undertaken at the Kazusa DNA Research Institute and the Chiba Prefectural Agriculture and Forest Center, both located in Japan, is discussed by Kolikonda and colleagues. This chapter also provides the cost comparisons (costs involved) of conventional and molecular breeding programs.

The cultivated tomato, the most popular vegetable crop in the world, is an important model system for genetics and genomics studies. Marker-assisted selection has been employed extensively in tomato breeding for improving many simple traits. Kinkade and Foolad in Chapter 12 look for QTL analysis approaches and focus on the use of new "omics" technology and its potential use for improving fruit quality in tomato breeding. Progress on reverse genetics approaches, such as TILLING and the bioinformatic workflows to handle high-throughput identification of mutations in candidate genes are discussed.

In sugarcane, up until two decades ago most of the breeding efforts for improvement were purely traditional. Chapter 13, by Gouy and colleagues, highlights recent advances in genomics and its applications for enhancing sugar yields. The chapter also highlights the ongoing efforts on genomic selection (GS) for enhancing yield gains in sugarcane.

Summary and Outlook

This volume presents a number of comprehensive and informative articles written by eminent scientists in the area of crop genomics and molecular breeding. It is important to mention here that the traits and crops discussed in this volume provide just some examples on how genomics can help facilitate the enhancement of tolerance to abiotic stresses and quality in crops.

Volume I of this series offers comprehensive reviews of biotic stress tolerance in a range of crops. As compared to the selected examples of GAB for biotic stress tolerance, it is clear that although rice has made significant progress in GAB for abiotic stress tolerance, most success stories of GAB are related to biotic stresses. This may be attributed to the partially qualitative inheritance and higher heritability of disease resistance as compared to abiotic stress tolerance. Similar to submergence tolerance, if the QTLs contribute higher phenotypic variance, GAB approaches such as MAS and MABC can be deployed in breeding programs. However, in the case of tolerance to abiotic stresses and yield, where several and small-effect QTLs are involved, simple molecular breeding approaches such as MABC and MAS are not as effective. In those cases, MARS (Bernardo and Charcosset 2006) and GS (Heffner et al. 2009, 2010, 2011; Heslot et al. 2012; Nakaya and Isobe 2012) are expected to be the most promising approaches. In this context, and in addition to biparental linkage mapping, mapping approaches such as genome-wide association studies (GWAS; Huang et al. 2010; Zhao et al. 2011; Li et al. 2012; Pasam et al. 2012) and nested association mapping (NAM; Hung et al. 2011; Kump et al. 2011; Cook et al. 2012) can be

implemented due to the availability of low-cost, high-throughput sequencing technologies. While GAB approaches are routinely deployed in the private sector and in developed countries, availability of breeder-friendly decision support tools is required for enhanced adoption of GAB in developing countries. In this context, some tools like integrated system for marker-assisted breeding (ISMAB) (https://www.integrated breeding.net/ib-tools/breeding-decision/marker-assisted-back-crossing-tool), OptiMAS (https://www.integratedbreeding.net/node/1407), GS modules (Pérez-Rodríguez et al. 2012; de Los Campos et al. 2013) and platforms like Integrated Breeding Platform (IBP) (https://www.integratedbreeding.net/) are being developed.

We hope that these two volumes will allow graduate students and young scientists to better appreciate the potential of GAB and will encourage them to pursue careers in this exciting area of crop improvement. In addition, GAB practitioners as well as policy makers should be able to use these volumes for developing the road map for the improvement of target crops in their respective geographical areas.

References

Bernardo R and Charcosset A (2006) Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. Crop Sci 46:614–621.

Bouis HE and Welch RM (2010) Biofortification – a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. Crop Sci 50:S-20–S-32.

Chamberlin KD, Melouk HA, Madden R, Dillwith JW, Bannore Y, El Rassi Z and Payton M (2011) Determining the oleic/linoleic acid ratio in a single peanut seed: A comparison of two methods. Peanut Sci 38:78–84.

Collard BC, Vera Cruz CM, McNally KL, Virk PS, and Mackill DJ (2008) Rice molecular breeding laboratories in the genomics era: Current status and future considerations. Int J Plant Genomics 2008:524847.

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, and Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158:824–834.

de Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, and Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345.

Dhillon T, Pearce SP, Stockinger EJ, Distelfeld A, Li C, Knox AK, Vashegyi I, Vágújfalvi A, Galiba G, and Dubcovsky J (2010) Regulation of freezing tolerance and flowering in temperate cereals: The *Vrn-1* connection. Plant Physiol 153:1846–1858.

Food and Agriculture Organization (FAO), (2010) Report of salt affected agriculture http://www.fao.org/ag/agl/agll/spush/.

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Kleinand R, and Mullet J (2007) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58: 327–338.

Heffner EL, Jannink J-L, Iwata H, Souza E and Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597–2606.

Heffner EL, Lorenz AJ, Jannink J-L, and Sorrells ME (2010) Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci 50:1681–1690.

Heffner EL, Sorrells ME, and Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1–12.

Heslot N, Yang H-P, Sorrells ME, and Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52:146–160.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F, Li J, and Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genetics 42:961–969.

Huke RE and Huke EH (1997) Rice area by type of culture: South, Southeast and East Asia. IRRI, Los Baños, Philippines.

Hung HY, Browne C, Guill K, Coles N, Eller M, Garcia A, Lepak N, Melia-Hancock S, Oropeza-Rosas M, Salvo S, Upadyayula N, Buckler ES, Flint-Garcia S, McMullen MD, Rocheford TR, and Holland JB (2011) The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. Heredity 108:490–499.

Jordan DR, Hunt CH, Cruickshank AW, Borrell AK, and Henzell RG (2012) The relationship between the stay-green trait and grain yield in elite sorghum hybrids grown in a range of environments. Crop Sci 52:1153–1161.

Kobayashi F, Takumi S, Kume S, Ishibashi M, Ohno R, Murai K, and Nakamura C (2005) Regulation by Vrn-1/Fr-1 chromosomal intervals of CBF-mediated Cor/Lea gene expression and freezing tolerance in common wheat. J Exp Bot 56:887–895.

Kulwal PL, Thudi M, and Varshney RK (2011) Genomics interventions in crop breeding for sustainable agriculture. In: Meyers RA (ed) Encyclopedia of sustainability science and technology. New York: Springer, pp. 2527–2540. doi:10.1007/978-1-4419-0851-3.

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, and Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–168.

Labadarios D, Steyn NP, Mgijima C, and Dladla N (2005) Review of the South African nutrition policy 1994-2002 and targets for 2007: Achievements and challenges. Nutrition 21:100–108.

Lake IR, Hooper L, Abdelhamid A, Bentham G, Boxall ABA, Draper A, Fairweather-Tait S, Hulme M, Hunter PR, Nichols G, and Waldron KW (2012) Climate change and food security: Health impacts in developed countries. Environment and health perspectives 120:1520–1526.

Li X, Yan W, Agrama H, Jia L, Jackson A, Moldenhauer K, Yeater K, McClung A, and Wu D (2012) Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). PLoS ONE 7:e29350.

Lopez K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, and Mullet J (2007) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327–338.

Ma JF, Ryan PR, and Delhaize E (2001) Aluminium tolerance in plants and the complexing role of organic acids. Trends Plant Sci 6:273–278.

Müller O and Krawinkel M (2005) Malnutrition and health in developing countries. CMAJ 173:279–286.

Nakaya A and Isobe SN (2012) Will genomic selection be a practical method for plant breeding? Ann Bot 110:1303–1316.

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, and Graner A (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol 12:16.

Pérez-Rodríguez P, Gianola D, González-Camacho JM, Crossa J, Manès Y, and Dreisigacker S (2012) Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3 2:1595–1605.

Ren ZH, Gao J P, Li L-G, Cai X-L, Huang W, Chao D-Y, Zhu M-Z, Wang Z-Y, Luan S, and Lin H-X (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat Genet 37:1141–1146.

Schneider SH, Semenov S, Patwardhan A, Burton I, Magadza CHD, Oppenheimer M, Pittock AB, Rahman A, Smith JB, Suarez A, and Yamin F (2007) Assessing key vulnerabilities and the risk from climate change. In: Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, and Hanson CE (eds) Climate Change 2007: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press, pp. 779–810.

Tester M and Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327:818–822.

Varshney RK, Close TJ, Singh NK, Hoisington DA, and Cook DR (2009) Orphan legume crops enter the genomics era! Curr Opin Plant Biol 12:202–210.

Varshney RK, Glaszmann J-C, Leung H, and Ribaut JM (2010) More genomic resources for less-studied crops. Trends Biotechnol 28:452–460.

Varshney RK, Graner A, and Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630.

Varshney RK, Ribaut JM, Buckler ES, Tuberosa R, Rafalski JA, and Langridge P (2012) Can genomics boost productivity of orphan crops? Nat Biotechnol 30:1172–1176.

WHO (2009) http://www.who.int/whosis/whostat/2009/en/index.html

Wolf B (1996) Diagnostic techniques for improving crop production. New York: Food Products Press.

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, and McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467.

Chapter 2

Applying Genomics Tools for Breeding Submergence Tolerance in Rice

Endang M. Septiningsih, Bertrand C. Y. Collard, Sigrid Heuer, Julia Bailey-Serres, Abdelbagi M. Ismail, and David J. Mackill

Abstract

Flooding stress is one of the most important abiotic stresses constraining rice production, especially in rain-fed lowland areas. The effect of this stress has intensified in past decades and is predicted to increase in the years to come as a result of global climate change. At the International Rice Research Institute (IRRI), breeding for tolerance to submergence imposed by flash flooding during the vegetative stage has been one of the institute's priority objectives for more than three decades. Several tolerant breeding lines have been developed through conventional breeding; however none of those early varieties has been widely accepted by farmers. An important breakthrough was the identification of the major quantitative trait locus (QTL) *SUB1* in the mid-1990s, which led to the identification of three ethylene responsive factors (ERFs), of which *SUB1A* is the primary contributor for tolerance. These findings and the available molecular marker technology have enabled breeders to develop submergence-tolerant varieties through a fast-track marker-assisted backcrossing (MABC) strategy to introgress *SUB1* using mega varieties as recurrent parents. Currently eight Sub1 varieties have been developed in IRRI, six of which have been released in several countries. The success of Sub1 varieties has inspired the development of new breeding products for other stresses using a similar strategy, such as tolerance to anaerobic conditions during germination and stagnant flooding. Recent advances in genomics have tremendously increased the efficiency of marker-assisted breeding, bringing us to the point where rice varieties resilient to multiple stresses can be developed to meet future challenges facing rice production, most of which have intensified with global climate changes.

Introduction

Flooding stress is a widespread problem that adversely affects farmers in rice-growing areas, especially in the flood-prone, rain-fed lowlands of South and Southeast Asia and West Africa. More than 20 million ha of rain-fed lowlands are flood prone, with conditions ranging from flash flood to deepwater, which can be up to several meters and last for months (Mackill et al. 2010). Over thousands of years, rice has developed adaptive mechanisms to grow well in

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

flooded ecosystems, as we can see in today's lowland rice (Pampolino et al. 2008); however, too much water at any stage of development could lead to serious injury or total crop loss. Recently weather patterns have become increasingly irregular as a result of climate change (Wassmann et al. 2009; Jagadish et al. 2012), and unexpected heavy rains can inundate rice fields along riverbanks and in low-lying areas and damage crop production. In recent years, destructive typhoons and heavy rains have caused huge damage in Bangladesh, Cambodia, Myanmar, the Philippines, Laos, Vietnam, Pakistan, and, most recently, Thailand, causing these countries to lose millions of tons of rice.

The most common flooding stress is the "flash flooding" that occurs during the monsoon season with different intensities and durations. During these episodes, water completely submerges the established rice crop for a short period from several days up to two weeks. If rice plants remain submerged for more than five days, they start to die and do not recover after the water recedes. The severity of the stress and the number of days that rice plants can survive underwater depends on environmental conditions, such as temperature, water turbidity, solar radiation, and soil fertility (Setter et al. 1997; Das et al. 2009). In flash flood–prone regions, farmers usually cultivate landraces that are tall and moderately tolerant of submergence but that have low yield. In some other areas where high-yielding but submergence-intolerant rice varieties have been cultivated, farmers often suffer from crop losses or significant yield reduction caused by flash-flood episodes.

It is not uncommon for floodwaters to stay in the field from two weeks to several months in some low-lying rice areas. In this situation, the water level is generally 20 to 50 cm in depth, which is referred to as stagnant flooding (Septiningsih et al. 2009; Mackill et al. 2010; Singh et al. 2011). Even though plants are not completely submerged under these conditions, grain production is greatly reduced due to poor tillering and greater susceptibility to lodging (Tuong et al. 2000; Singh et al. 2011). In some areas, stagnant flooding can also immediately follow a flash-flood period. In this situation, farmers usually rely on landraces that can cope with both stresses, although these tend to be very low yielding. In areas where progressive flood waters reach a depth of several meters for several months, farmers usually cultivate "floating rice" or "deepwater rice" that shows rapid internode elongation with the rising floodwater, maintaining the uppermost leaves and panicles above the surface (Catling 1992). Deepwater rice generally has poor yield due to excessive vegetative growth.

Flash floods can also cause excessive damage during germination and early seedling growth, resulting in poor crop establishment in direct-seeded rice areas. This can occur in both irrigated areas when the land is not level and in flood-prone rain-fed ecosystems when rainfall occurs within a few days following seeding (Ismail et al. 2009; Angaji et al. 2010). Most rice varieties are unable to vigorously germinate, elongate, and survive under complete submergence and, as a consequence, seedling establishment is very poor or completely absent in fields that are flooded. Poor seedling germination is also a problem if fields are not properly leveled, resulting in the formation of puddles in which seedlings are submerged, and in fields with drainage problems, if heavy rains occur directly after seeding. The hazards of flooding during germination can prevent farmers from adopting direct-seeded rice technology or force them to discontinue this practice (Konchan and Kono 1996). Farmers usually wait for the "right" time to sow or broadcast their seeds; however, if an unexpected flood destroys their crops, they reseed or transplant their fields. Apart from being costly and labor demanding, the delay caused by re-planting and thereby late harvest of the first crop can delay the planting of the following season's crop.

Efforts to identify submergence-tolerant rice genotypes were initiated at the International Rice

Research Institute (IRRI) during the 1970s (Vergara and Mazaredo 1975). Among other varieties, FR13A, which was derived from pure-line selection of the landrace "Dhalputtia" originating from Orissa, India, was identified as the most tolerant cultivar (HilleRisLambers and Vergara 1982) and subsequently used extensively as a tolerant donor in breeding programs. Several breeding lines tolerant of submergence have been developed, including IR49830-7-1-2-1, a highly tolerant and high-yielding indica-type line (Mackill et al. 1993), which was released in Cambodia under the name "Popoul" in 1999. However, despite its tolerance of submergence, this variety was not widely adopted by farmers because it lacked several key traits, such as locally preferred grain quality (Neeraja et al. 2007).

A major landmark in the history of submergence-tolerance breeding was the identification of the major quantitative trait locus (QTL) *Submergence 1* (*SUB1)* that controls this trait (Xu and Mackill 1996). This early work ultimately led to the cloning of the *SUB1* region in FR13A and subsequently to the identification of the ethylene-responsive factor (ERF) gene *SUB1A-1* which is necessary and sufficient for submergence tolerance (Xu et al. 2006). Based on the sequence information, *SUB1*-specific molecular markers were developed that facilitated a precise marker-assisted backcrossing (MABC) system that is now successfully being deployed to introgress the *SUB1* QTL into widely grown "mega-varieties" in South and Southeast Asia, as well as in Africa (Neeraja et al. 2007; Septiningsih et al. 2009; Bailey-Serres et al. 2010; Manzanilla et al. 2011; Mackill et al. 2012). As a result of repeated backcrossing to the respective recipient parent, the improved submergence-tolerant varieties carry the FR13A *SUB1* locus but are otherwise identical to the original variety. Most importantly, grain quality and other locally preferred traits are unaltered in these new varieties, which enhance adoption by farmers and variety release in the target countries. These new varieties are called "Sub1" or "Scuba" rice.

Breeding direct-seeded rice varieties, which requires tolerance of flooding during germination, also began in the past. However, success was very limited mainly due to the lack of tolerant donor varieties, and the complexity of the trait (Yamauchi et al. 1993; Yamauchi and Winn 1996; Biswas and Yamauchi 1997). More recently, after screening thousands of rice accessions, several landraces have been identified that are tolerant of submergence during germination (also referred to as anaerobic germination, AG) (Angaji et al. 2010). The analysis of mapping populations derived from tolerant donor parents has led to the identification of promising QTLs that might be useful for direct-seeded systems, as described in more detail below. As was the case for submergence and AG tolerance, efforts to develop varieties for stagnant flooding and deepwater conditions were made over several decades at IRRI (HilleRisLambers and Seshu 1982). More recently, a concerted effort has started to combine tolerance of stagnant flooding with submergence tolerance (based on *SUB1*). This is particularly important for short rice varieties, for example, Swarna, which was the first mega-variety introgressed with *SUB1* (Swarna-Sub1). One variety that combines tolerance of submergence and stagnant flooding has been developed by conventional breeding based on phenotype selection (IRRI 119), which was released in the Philippines as PSB Rc68 (Septiningsih et al. 2009; Mackill et al. 2010). More recently, breeding lines that are high yielding and perform well under stagnant flooding have also been developed through conventional breeding (Mackill et al. 2010).

This chapter will review the progress made in applying genomics tools to unravel the molecular and physiological basis of different types of submergence tolerance, as well as efforts to use this information to develop improved rice varieties to meet the challenges of the future.

Applying Genomics Tools for Molecular Studies and Breeding

Identification of the QTLs and Genes underlying Tolerance

Genetic dissection of quantitative traits using QTL mapping tools became feasible with the availability of DNA markers about two decades ago. Even though varietal improvement can be achieved using conventional breeding, QTL mapping has tremendous potential to breed varieties in a more effective and efficient way. Once a major QTL for an important trait has been mapped and the markers closely linked to the locus have been identified, the QTL can be used as a target in marker-assisted breeding to rapidly breed an improved variety. Furthermore, once the gene(s) underlying the QTL has(have) been cloned, markers can be developed from the target gene(s) for even more precise marker-assisted breeding. The cloned gene(s) and near-isogenic lines (NILs) that differ only at the locus of interest also provide a starting point for detailed study of molecular, physiological, and developmental mechanisms underlying the trait of interest.

Tolerance to Flash Flood during the Vegetative Stage

FR13A was identified as one of the best submergence-tolerant donors and was used extensively by breeders in the 1970s; however, very little was known about the genetic basis of the tolerance possessed by this variety. Independent studies using the submergence-tolerant variety FR13A identified the major *SUB1* QTL and several minor QTLs (Xu and Mackill 1996; Nandi et al. 1997; Toojinda et al. 2003). It was found that *SUB1* alone contributes 69% of the phenotypic variance (Xu and Mackill 1996) and could provide tolerance to complete submergence for up to two weeks. *SUB1* was then fine-mapped to a region of 0.06 cM using an F_2 segregating population of 2950 individuals (Xu et al. 2000), and the underlying genes were finally cloned as a cluster of three ethylene-responsive factor (ERF) genes, *SUB1A*, *SUB1B*, and *SUB1C* (Xu et al. 2006). It was demonstrated through gene transformation that *SUB1A* was the main contributor for tolerance (Xu et al. 2006); and this finding has been confirmed through a progeny test of recombinants identified within the *SUB1* cluster in several thousand individuals in segregating populations (Septiningsih et al. 2009). It was also shown that *SUB1A* gene expression, which subsequently determines the amount of tolerance, is dosage-dependent. This suggests that, for improvement in hybrid rice, both parents should carry the *SUB1A* gene for maximal effect (Septiningsih et al. 2009). In addition, allelic survey studies showed that, in some cases, the expression of *SUB1A* is a more reliable parameter to use instead of the original allelic determination, that is, the *SUB1A-1*-tolerant allele and *SUB1A-2*-intolerant allele (Singh et al. 2010; Septiningsih et al. 2012).

Tolerance to Anaerobic Conditions during Germination

Tolerance of flooding during seed germination, referred to as anaerobic germination (AG), is one of the most important traits necessary to ensure good seedling establishment in direct-seeded rice in both rain-fed flood-prone and irrigated ecosystems. Several QTLs for AG tolerance were reported on chromosomes 1, 2, 5, and 7 (Jiang et al. 2004, 2006). Our group at IRRI identified several promising QTLs derived from a tolerant donor from Myanmar, Khao Hlan On (Angaji et al. 2010). The QTL with the largest effect was detected on the long arm of chromosome 9 (*qAG-9-2* or *AG1*), having a logarithm of odds (LOD) score of 20.3 and explaining 33.5% of the variation for this trait. Fine mapping of this QTL in the background of IR64 narrowed the locus down to a 58-kb region based on the Nipponbare sequence. Several candidate genes have been identified, and gene validation and

characterization are under way (unpublished data). This QTL is a promising target for marker-assisted selection for direct-seeded rice varieties. Another major QTL was detected on the short arm of chromosome 7 (referred to as *qAG7.1* or *AG2*), derived from the tolerant donor variety Ma-Zhan Red (Septiningsih et al. 2013). NIL development for fine mapping and varietal improvement for this QTL are under way. This QTL can be combined with *AG1* to confer higher tolerance, given that the effect of the two QTLs proved to be additive or synergistic. Studies using other unrelated sources of tolerance are under way to identify additional QTLs involving different tolerance mechanisms.

Deepwater Rice

The most important trait for the survival of deepwater rice is rapid underwater internode elongation, to ensure that upper leaves maintain efficient photosynthesis (Catling 1992; Vergara et al. 1976). A number of studies have identified QTLs for deepwater traits, such as internode elongation and number of elongated internodes (Sripongpangkul et al. 2000; Nemoto et al. 2004; Hattori et al. 2007; Kawano et al. 2008). Even though different donors were used in studies on deepwater rice, QTLs on chromosomes 1, 3, and 12 were repeatedly detected in different mapping populations. Through NIL evaluation, it was confirmed that the QTL on chromosome 12 contributed the most rapid internode elongation in deepwater stress conditions (Hattori et al. 2008). By positional cloning, Hattori and colleagues (2009) identified the genes within this QTL: *SNORKEL1* (*SK1*) and *SNORKEL2* (*SK2*). Like the *SUB1*s, these are ERF genes of subgroup VII, possessing a single AP2 DNA-binding domain. Both *SK1* and *SK2* are missing in the elite recurrent parent genome (Hattori et al. 2009). The presence of genes conferring abiotic stress tolerance have recently been found in donor landraces but not in the reference Nipponbare genome, as in the case of the donors for tolerance of submergence, phosphorus deficiency, and deep water (Xu et al. 2006; Hattori et al. 2009; Heuer et al. 2009).

Development of Sub1 Varieties

Identification of the *SUB1* gene enabled marker-assisted selection (MAS) for submergence tolerance. The *SUB1* QTL has a large effect and the phenotypic difference between tolerant and susceptible types is consistent. A wealth of sequence polymorphisms in and around the gene cluster from both tolerant and susceptible parents provided useful markers for MAS (Xu et al. 2006; Neeraja et al. 2007; Septiningsih et al. 2009; Singh et al. 2010; Iftekharuddaula et al. 2011). Given the popularity of high-yielding varieties with grain quality attributes that lacked submergence tolerance, a large-scale MABC program was undertaken at IRRI using these popular varieties as recurrent parents. Instead of using the original donor FR13A, two FR13A-derived improved lines were used as donors: IR49830-7-1-2-2 (IR49830-7) and IR40931-33-1-3-2 (IR40931-33) (Mackill et al. 1993). This led to successful introgression of the *SUB1* locus into several popular varieties (mega-varieties) in India, Bangladesh, Indonesia, Laos, and the Philippines. By using precision MABC, the high yield and desirable grain and eating qualities of these mega-varieties were retained (Septiningsih et al. 2009; Singh et al. 2009).

Three levels of MABC were applied to ensure high precision in breeding and to reduce the time required for variety development: (1) foreground selection, in which markers tightly linked to *SUB1* are used to select for the locus; (2) recombinant selection, in which closely linked flanking markers are used to minimize the donor chromosomal segment containing *SUB1*; and (3) background selection, in which DNA markers are used to accelerate the recovery of the recurrent parent genome (Figure 2.1; Collard and Mackill 2008). The method of selecting recombinants on both sides of the target

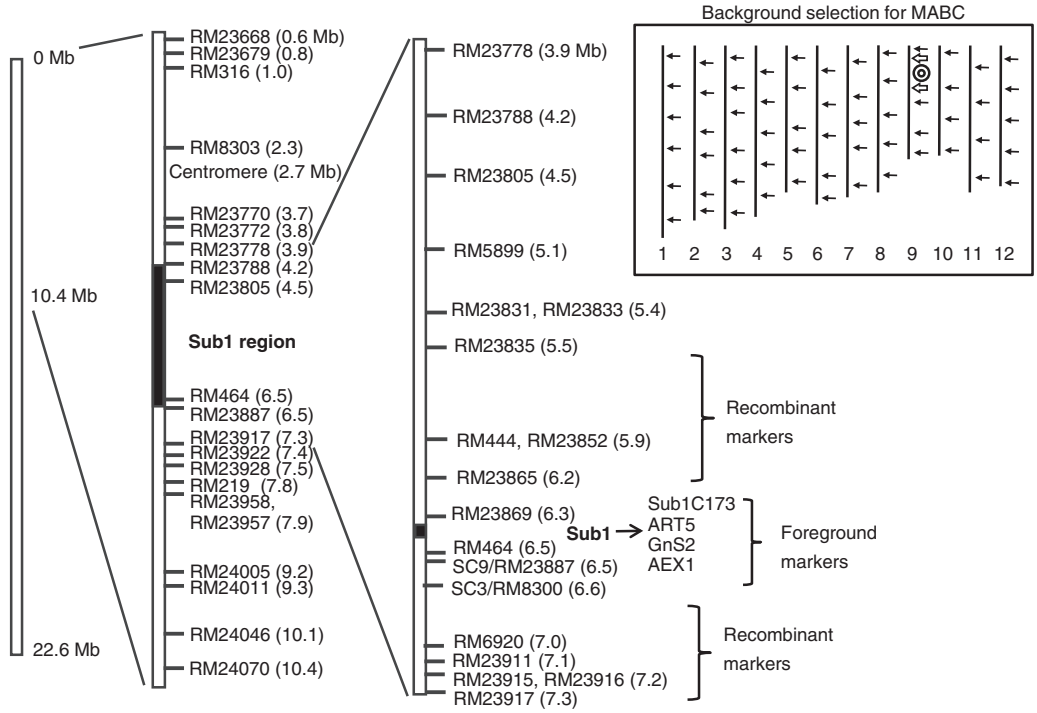


Fig. 2.1. Marker-assisted backcrossing (MABC) for *SUB1*. A number of markers have been identified as useful foreground markers (used to retain the tolerant *SUB1* allele) and recombinant markers (flanking markers used to select for a small *SUB1* introgression). Once foreground and recombinant selection have reduced the population size, background selection is performed to eliminate donor introgressions across the rest of the genome and return to the recurrent parent genome (see inset). For a color version of this figure, please refer to the color plate.

locus in at least two backcross (BC) generations during MABC was first proposed by Young and Tanksley (1989) and applied in rice by Chen and colleagues (2000). Foreground and recombinant polymerase chain reaction (PCR)-based DNA markers, such as simple sequence repeat (SSR), cleaved amplified polymorphic sequences (CAPS), insertion/deletion (INDEL), and mismatched single nucleotide polymorphism (SNP) markers were generated to facilitate introgression of the *SUB1* allele from the FR13A-derived lines in the background of popular varieties through MABC (Neeraja et al. 2007; Septiningsih et al. 2009). The most common markers used for foreground selection are summarized in Table 2.1.

In the first stage, six varieties were enhanced with *SUB1* using MABC (Neeraja et al. 2007; Septiningsih et al. 2009; Iftekharuddaula et al. 2011). In 2011, with the addition of Ciherang-Sub1 and PSB Rc18-Sub1, a total of eight Sub1 mega-varieties have now been developed (Table 2.2; Figure 2.2). Evaluation of these new cultivars has indicated that there is no negative effect of *SUB1* on other traits, and no linkage drag, especially as recombinant selection was performed. In many cases, the development of these new Sub1 varieties will make it easier to incorporate *SUB1* in the future because recombinant and background selection may not need to be as rigorous because the donor parents are highly adapted and possess many desirable agronomic characters. For example, a single backcross has been used for the development of Ciherang-Sub1 using IR64-Sub1 as the donor parent; this donor is closely related to the

Table 2.1. The most common markers used in the development of Sub1 mega-varieties

Primer	Sequence	Tm (°C)	Position	Type of marker	Accession	Expected size (bp)
RM8300 (SC3)F	GCTAGTGCAGGGTTGACACA	60	~300 kb upstream of *SUB1A*	SSR	NB[a]	200
RM8300 (SC3)R	CTCTGGCCGTTTCATGGTAT	60				
GnS2F	CTTCTTGCTCAACGACAACG	60	exon of *SUB1A*	CAPS (*AluI/PvuII*)	Teqing	242
GnS2R	TCGATGGGGTCTTGATCTCT	60			26D17[b]	132 & 110
					NB[a]	No product
AEX1F	AGGCGGAGCTACGAGTACCA	62	non-synonymous SNP for *SUB1A*	mismatch	26D17[b]	231
AEX1R	GCAGAGCGGCTGCGA	62		specific for tolerance	Teqing	No product
					NB[a]	No product
ART5F	CAGGGAAAGAGATGGTGGA	60	*SUB1C* promoter	15 bp insertion in NB/93-11	NB[a]/93-11	217
ART5R	TTGGCCCTAGGTTGTTTCAG	60			118k20[b]	202
Sub1C173F	AACGCCAAGACCAACTTCC	60	exon of *SUB1C*	9 bp deletion in NB/93-11	NB[a]/93-11	164
Sub1C173R	AGGAGGCTGTCCATCAGGT	60			118k20[b]	173

[a]Nipponbare
[b]derived from IR40931-26

Table 2.2. Current status of Sub1 lines developed through marker-assisted backcrossing (MABC)

Recurrent parent	Country of origin	Donor parent	Generation	Introgression size (Mb)	IRRI designation	Country where released	Year of release
Swarna	India	IR49830	BC_3F_2	2.3–3.4	IR05F102	India & Indonesia, Bangladesh, Nepal & Myanmar	2009 2010 2011
IR64	Philippines	IR40931	BC_2F_2	6.5–7.8	IR07F102	Philippines & Indonesia	2009
Samba Mahsuri	India	IR49830	BC_2F_2	6.5–9.2	IR07F101	Nepal, India	2011 2013
TDK1	Laos	IR40931	BC_3F_2	1.5–2.5	IR07F289	N/A[a]	
BR11	Bangladesh	IR40931	BC_2F_2	0.3–2.6	IR07F290	Bangladesh	2010
CR1009	India	IR40931	BC_2F_3	2.7–6.4	IR07F291	N/A[b]	
Ciherang	Indonesia	IR64-Sub1	BC_1F_2	6.5–7.8	IR09F436	Indonesia, Bangladesh	2012 2013
PSB Rc18	Philippines	IR64-Sub1	BC_1F_2	6.5–7.8	IR09F437	N/A[c]	

[a]Not applicable; under advanced evaluation in Lao PDR
[b]Not applicable; released proposal submitted in India
[c]Not applicable; under evaluation in the Philippines

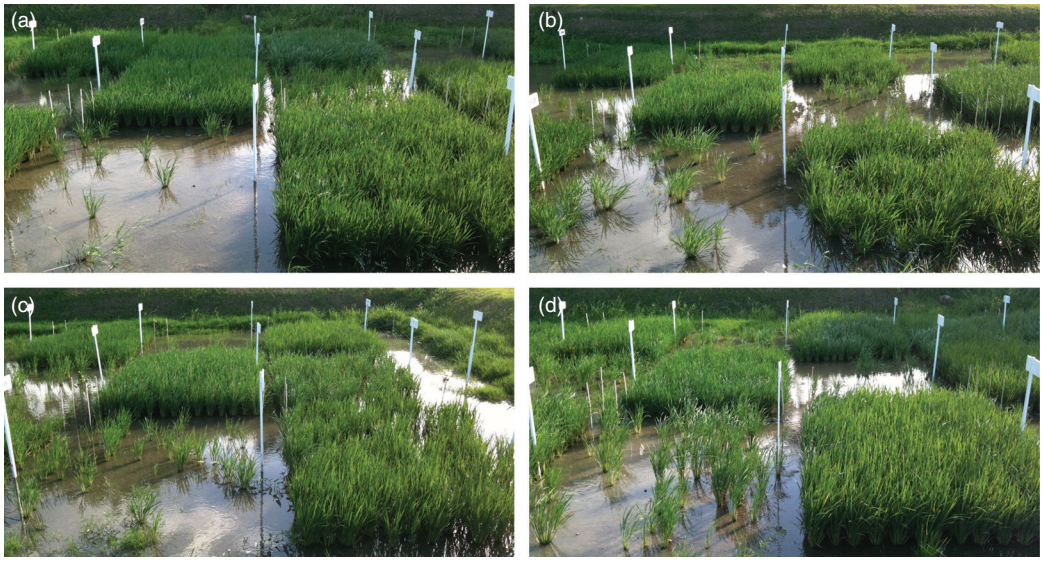


Fig. 2.2. Photographs of Sub1 varieties compared with the original varieties (foreground) at 2 months after 16 days of submergence in an IRRI field trial. (a) Swarna vs Swarna-Sub1; (b) BR11 vs BR11-Sub1; (c) Samba Mahsuri vs Samba Mahsuri-Sub1; and (d) Ciherang vs Ciherang-Sub1. For a color version of this figure, please refer to the color plate.

recurrent parent Ciherang (i.e., Ciherang is derived from a cross with IR64).

Performance of Sub1 Varieties

Performance of Sub1 Varieties under Controlled Flooding

The aim of evaluating Sub1 varieties under controlled non-flooded conditions was to determine whether the incorporation of *SUB1* had any other effects on grain yield or grain quality of any particular popular variety used as a recipient. This question is important because submergence normally occurs once every 2-5 years, and any reduction in yield or changes in grain quality of these popular varieties will ultimately lead to their rejection by farmers and/or millers. A set of trials was conducted under controlled flooding in the field to determine the extent to which *SUB1* can enhance survival and yield in field conditions and whether it can work in several genetic backgrounds and under variable environments. These trials used the first set of three popular varieties introgressed with *SUB1*: Swarna, IR64, and Samba Mahsuri. When grown under controlled non-flooded conditions, the pairs of NILs (pairs with and without *SUB1*) were almost identical in their growth, flowering and maturity duration, and yield, as well as in all aspects of grain quality. These findings suggest that *SUB1* does not have other effects on growth and yield in the absence of stress (Singh et al. 2009).

The same NILs were then evaluated under complete submergence for either 12 or 17 days in the field (Singh et al. 2009, 2011), where substantial differences were observed between the Sub1 lines and their recurrent parents. The survival of both groups of lines decreased following submergence, but the survival of the Sub1 lines was considerably higher. For instance, following 17 days of submergence, the survival of Samba Mahsuri and IR64 was only 7% and 11%, while that of Samba Mahsuri-Sub1 and IR64-Sub1 lines was 83% and 85%, respectively. This improvement in survival of the tolerant lines was associated with substantial suppression of shoot elongation and higher biomass, with more non-structural carbohydrates in the shoot, and chlorophyll retention in leaves after submergence, all of which were previously associated with

tolerance of complete submergence in rice (Ella et al. 2003; Das et al. 2005; Sarkar et al. 2006). Recovery after submergence was also faster in the tolerant lines, resulting in greater production of early tillers that subsequently produced panicles, unlike the later tillers produced by the sensitive lines, which mostly remained vegetative through maturity. Consequently, the yield of the tolerant lines was consistently and considerably higher than that of the sensitive parents, ranging from 1 t ha^{-1} to over 3.5 t ha^{-1} yield enhancement over sensitive lines under field conditions (Sarkar et al. 2009; Singh et al. 2009). All grain quality aspects of the Sub1 lines were also either similar to or better than those of the recurrent parents after submergence. In addition, Sub1 introgression lines also suffer less delay in flowering and maturity than their parental lines (Singh et al. 2009, 2011). Together, these data provided sufficient evidence that the *SUB1* gene can effectively provide protection against flash floods and can work across many genetic backgrounds and environments. These conclusions were further confirmed after additional Sub1 pairs became available.

Sub1 Performance in Farmers' Fields

A participatory approach involving farmers was used to extensively evaluate the new Sub1 varieties in South and Southeast Asia. This approach provides information on the adaptation of these varieties to local conditions and their reaction to common pests and diseases, and also provides direct feedback from farmers on their acceptance of these varieties based on their local preferences. The data generated through these trials are being used as part of the variety release process in some countries, which also speeds up the notification and commercialization of the new varieties (Manzanilla et al. 2011; Paris et al. 2011). In most of these trials, comparisons were made between Sub1 NILs and usually involving one or more varieties common among local farmers as checks. At sites where there was no flooding during the season, yield of both varieties was similar. However, in areas that experienced incidences of flooding, Sub1 varieties always had higher yields, exceeding twice the yield of the non-Sub1 types and checks in some cases, particularly when the stress was severe. For example, in one of the first set of trials conducted with these lines in 2007, Swarna and Swarna-Sub1 were compared at 32 sites in Uttar Pradesh, India. Yields of the two varieties were the same at sites that did not experience submergence (averaging 5.5 t ha^{-1}). However, at most of the sites, submergence of more than 5 days was experienced, and the average yield of Swarna-Sub1 was 3.67 t ha^{-1} compared with 2.34 t ha^{-1} for Swarna. Under more severe flooding of 15 days, the yield of Swarna-Sub1 was more than double that of Swarna (2.7 t ha^{-1} vs. 1.3 t ha^{-1}; Mackill et al. 2012). This trend was consistently experienced in subsequent trials conducted over hundreds of sites in Asia every year. In almost all trials, genotypes carrying *SUB1* had higher survival, with yield advantages of $1 - 3$ t ha^{-1} over the sensitive NILs based on the duration of the flood and conditions of the floodwater.

Molecular and Physiological Mechanisms underlying Tolerance

Tolerance of Transient Flooding during the Vegetative Stage: The *SUB1* Mode of Action

As mentioned above, the *SUB1* locus contains a variable number of ERF genes. In all analyzed *japonica* rice varieties, including Nipponbare, and in some intolerant *indica* varieties, two ERF genes (*SUB1B*, *SUB1C*) are present in the *SUB1* locus (Figure 2.3a). The third gene, *SUB1A*, is present only in submergence-tolerant *aus*- and *indica* rice varieties, and in some intolerant indica varieties (Xu et al. 2006; Singh et al. 2010). Two different alleles of the *SUB1A* gene have been identified (*SUB1A-1*, *SUB1A-2*) and the *SUB1A-1* allele has been shown to be the major determinant of submergence tolerance (Septiningsih et al. 2009; Xu et al. 2006).

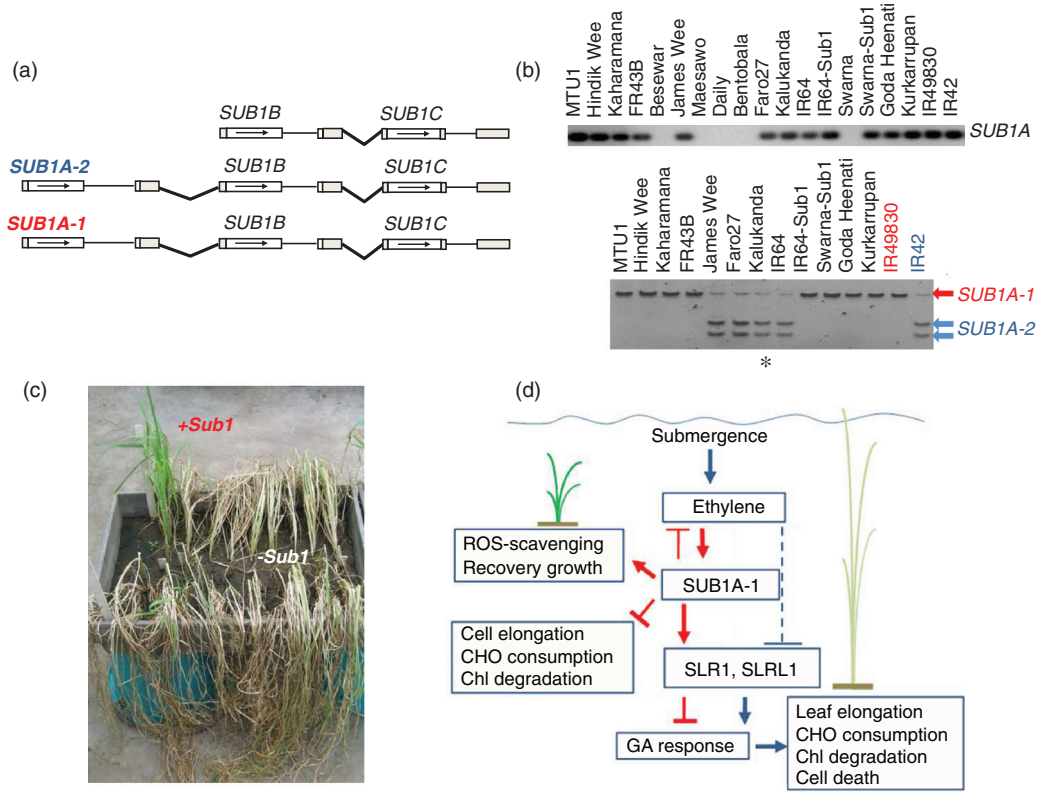


Fig. 2.3. ***SUB1*-specific markers and the molecular basis of submergence tolerance.** The *SUB1* locus contains either two (*SUB1B*, *SUB1C*) or three (*SUB1A*, *SUB1B*, *SUB1C*) ERF genes (**a**). Two different indica-/aus-specific *SUB1A* alleles are known, of which the *SUB1A-1* allele is present in submergence-tolerant varieties. Molecular markers distinguish between varieties with and without *SUB1A* (**b**, top panel) and between the *SUB1A-1* and *SUB1A-2* allele (**b**, bottom panel). The visible effect of the *SUB1* locus is the suppression of growth during submergence and plant recovery within about 2 weeks after de-submergence (**c**). Under submergence, ethylene induces the gibberellic acid (GA)-dependent escape response, which is suppressed by *SUB1A-1*-mediated maintenance of GA repression via the GA repressor proteins SLR1 and SLRL1 (**d**). ROS, reactive oxygen species; CHO, carbohydrate, CHL, chlorophyll, GA, gibberellic acid, SLR1, SLENDER RICE 1, SLRL1, SLENDER RICE LIKE 1. For a color version of this figure, please refer to the color plate.

SUB1A-1 is also the allele present in the *SUB1* donor variety FR13A and FR13A-derived lines that are now being used in Sub1 breeding programs (see above). The molecular markers used for Sub1 breeding enable breeders to distinguish between plants with and without the *SUB1A* gene, and to discriminate between the alleles of *SUB1A* (Figure 2.3b).

It has recently been suggested that the *SUB1A-2* allele can also confer tolerance of submergence since high expression of the *SUB1A-2* allele was observed in some varieties with intermediate tolerance (Singh et al. 2010; Septiningsih et al. 2012). However, the most tolerant *SUB1A-2* variety (James Wee) had a submergence-survival rate of 44%, which is significantly lower than *SUB1A-1* varieties (51–65%) (Singh et al. 2010). It remains to be shown whether the higher tolerance mediated by *SUB1A-1* compared with *SUB1A*-2 is related to the level of transcription or post-translational modification, for example, phosphorylation at the putative MAP-kinase target site that is specific to the *SUB1A-1* allele (Xu et al. 2006).

The main effect of *SUB1A-1* is the suppression of the "escape" response, which, in

intolerant plants, leads to leaf elongation growth in the attempt to grow to the water surface, that is, toward the light and oxygen (Figure 2.3c). In contrast, plants with the tolerant *SUB1* locus and transgenics that overexpress *SUB1A-1* (Xu et al. 2006) enter into a "quiescent" state that preserves carbohydrate reserves and limits anaerobic metabolism (Fukao et al. 2006; Barding et al. 2012). This tolerance strategy limits the energy crisis caused by carbohydrate starvation or reduced mitochondrial ATP generation under flooded, that is, low-oxygen conditions.

Leaf elongation under submergence is triggered by increased synthesis and entrapment of ethylene (Bailey-Serres and Voesenek 2008) and a subsequent decrease in abscisic acid (ABA) due to altered synthesis or enhanced turnover. This triggers an increase in levels of or sensitivity to gibberellins (GA) and a GA-dependent growth response (Figure 2.3d). Comparative analyses of intolerant and tolerant NILs and transgenics revealed that *SUB1A* has no effect on the degradation of ABA but alters GA-dependent elongation growth under submergence. This process is regulated by two GA-signaling repressor proteins, SLENDER RICE-1 (SLR1) and SLR1 LIKE-1 (SLRL1). In Sub1 rice and in *SUB1A-1* overexpression lines, SLR1 and SLRL1 transcript and protein accumulation was higher in submerged tissues, thereby maintaining inhibition of GA-mediated growth responses (Fukao and Bailey-Serres 2008). Numerous other genes, including many transcription factors, are additionally differentially regulated during submergence as was shown by a comparative microarray gene expression analysis (Jung et al. 2010). As a result of the repression of the GA response, Sub1 plants are significantly shorter than intolerant plants once floodwaters recede. Because of the repressed growth and additional physiological adaptations (Fukao et al. 2006), tolerant plants retain sufficient energy reserves for growth following submergence. In contrast, the elongated and chlorophyll-deficient leaves of intolerant plants lodge and generally do not renew growth (Figure 2.3c).

Additionally, upon de-submergence, plants experience severe oxidative stress because they are exposed to atmospheric oxygen and natural (high) light. This leads to the formation of high concentrations of cell-toxic reactive-oxygen species (ROS) and cell death. It has been shown that, in *SUB1A-1* genotypes, accumulation of superoxide and hydrogen peroxide as well as lipid peroxidation is lower than in intolerant genotypes (Fukao et al. 2011). In agreement with this, the authors have shown that abundance of transcripts encoding ROS-scavenging enzymes (ascorbate peroxidase, superoxide dismutase, catalase) is higher in Sub1 genotypes. Furthermore, *SUB1A-1* transcript abundance increased after treatment with methyl viologen (paraquat), which stimulates ROS production in chloroplasts (Fukao et al. 2011), suggesting that *SUB1A-1* is directly responsive to ROS. Taken together, the data suggest that *SUB1A* protects plants in submerged fields in two ways: (1) inhibition of GA-induced elongation growth, thereby preventing exhaustion of carbohydrates and an energy crisis during submergence; and (2) up-regulation of the ROS-scavenging system, thereby providing protection against cell damage upon de-submergence.

Recently it was demonstrated in *Arabidopsis* that the five ERF genes of the subfamily VII are substrates of an oxygen-regulated branch of the N-end rule pathway of targeted proteolysis (Gibbs et al. 2011). Members of this gene family in rice include *SUB1A* as well as the *SNORKEL* genes. The N-end rule pathway of targeted proteolysis is important for sensing of low oxygen and it regulates the expression of hypoxia-responsive genes in *Arabidopsis*. The substrates of turnover are constitutively synthesized ERF genes, including *RAP2.12*. Under aerated conditions, *RAP2.12* is stabilized by interaction with the membrane-localized acyl-CoA-binding protein (ACBP) 1 and 2 (Licausi et al. 2011). Under low-oxygen conditions, RAP2.12 is released from the plasma membrane and moves into the nucleus, where it positively regulates the expression of hypoxia-responsive genes. Under aerated

conditions, including upon re-oxygenation, the terminal methionine of RAP2.12 is removed by a methionine amino peptidase (MetAP). This exposes the second amino acid, a cysteine (C_2) within the consensus sequence MCGGAII, to oxidation, followed by conjugation of an arginine residue to the amino-terminus in a reaction catalyzed by arginine transferase (ATE), triggering ubiquitination by the ligase PROTEOLYSIS 6 (PRT6), and targeting RAP2.12 for proteasome degradation (Licausi et al. 2011). It was further shown that another member of the ERF subfamily VII, HYPOXIA RESPONSIVE 2 (HRE2), is stabilized under aerated conditions in the *prt6* mutant, resulting in constitutive accumulation of hypoxia-responsive genes (Gibbs et al. 2011). However, as has been shown by Gibbs and colleagues (2011) using an *in vitro* assay that SUB1A-1 protein is not degraded by this pathway even after the N-terminal conserved N-end rule target motif (MCGGEVI) was modified (MCGGAVI) to better match the consensus sequence. Resistance of SUB1A-1 to degradation prior to severe oxygen deficiency and upon re-oxygenation might indeed be very important for survival of submergence as well as protection against cell damage by ROS as outlined above.

Interestingly, it has been shown that *SUB1A* enhanced the survival of rice seedlings after drought stress in a pot experiment and that genes related to drought tolerance *(DREB1A*, *DREB1E*, *AP59*) are induced in M202-Sub1 (Fukao et al. 2011). It therefore seems that tolerance of submergence and drought stress is conferred by similar and partially overlapping pathways. This is currently being validated under different environmental conditions. It has been shown that *SUB1A* is specifically expressed in the growing parts of leaves (leaf base and leaf collar) and in the shoot apex of young rice seedlings (Singh et al. 2010; unpublished data). These tissues are the actively growing parts of rice plants and are therefore critical for growth and regeneration after exposure to stress. Detailed studies are now under way to specify a putative role of *SUB1A* in maintaining and protecting meristematic cells under stress.

Tolerance of Flooding during Germination

Seeds of most cereal crops (maize, wheat, barley, oats, sorghum) are extremely sensitive to low-oxygen stress during germination, and they normally fail to germinate in flooded or even saturated soils (Perata et al. 1997; Vartapetian and Jackson 1997). Unlike other cereals, rice is capable of germination under flooded conditions (Taylor 1942; Yamauchi et al. 1993; Ella and Setter 1999; Angaji et al. 2010), but only through elongation of the coleoptiles with failure to form roots and leaves (Biswas and Yamauchi 1997). However, substantial genetic variation in ability to germinate and establish in flooded soil was found in rice after screening a large number of germplasm accessions and breeding lines of different origins (Yamauchi et al. 1993; Biswas and Yamauchi 1997; Jiang et al. 2004; Angaji et al. 2010). A few rice genotypes with greater tolerance of flooding during germination and early seedling growth were identified and subsequently characterized to assess the physiological and molecular bases of tolerance (Ismail et al. 2009). Germinating embryos in flooded soils can suffer from hypoxia or even anoxia in severe cases, and this hinders further growth as oxygen is necessary for functioning of the enzymes involved in the breakdown and mobilization of stored carbohydrates and the oxidative pathways to generate energy for the growing embryos (Drew 1990; Greenway and Setter 1996).

Germinating rice seeds tolerate oxygen deficiency in flooded soils through various growth and metabolic adjustments. One of the most spectacular adaptive growth features is the excessive growth of the coleoptile, which can elongate even faster at low O_2 concentrations than in air (Alpi and Beevers 1983), and with considerable variation among genotypes in elongation

rate (Turner et al. 1981). Fast coleoptile elongation could facilitate contact with air in waterlogged or flooded soils and maintain adequate aeration of the growing embryo. This accelerated elongation was independent of ethylene synthesis (Pearce et al. 1992), but rather dependent on the extent of alcohol fermentation and ethanol synthesis (Setter and Ella 1994), which emphasizes the importance of anaerobic metabolism during germination and early seedling growth. However, this rapid growth as a mechanism of avoiding oxygen stress in anaerobic soils contrasts with tolerance of transient flooding at later vegetative stages of growth, when submergence injury is worsened by rapid growth and use of carbohydrate reserves (Das et al. 2005; Bailey-Serres et al. 2010). Recent studies showed that it is possible to combine tolerance of anaerobic conditions during germination and early growth with that during the vegetative stage conferred by the *SUB1* gene, with the latter causing inactivation of shoot elongation (Mackill et al. 2010). The regulatory mechanisms that mediate these opposing responses are yet to be determined, but seem to be stage-specific, and probably mediated by sugar and light signaling, as the seedlings shift from being dependent on storage reserves to carbohydrates supplied through photosynthesis.

Another important and well-characterized survival mechanism of anaerobic conditions is the primary shift from aerobic metabolism to anaerobic fermentation to generate the energy needed to sustain growth of the germinating embryo (Guglielminetti et al. 1995). Unlike other cereals, rice seeds sustain their ability to break down starch into readily fermentable carbohydrates when germinating under hypoxic or even anoxic conditions (Atwell and Greenway 1987). Rice seeds are known to have all the enzymes needed for degradation of starch and its use for the growing embryo, but the activities of these enzymes are affected by oxygen availability. The importance of α-amylases in starch degradation when O_2 is limiting was reported, in which both mRNA and protein accumulation were observed under anoxia (Perata et al. 1993; Guglielminetti et al. 1995; Perata et al. 1997; Hwang et al. 1999). Recently, Ismail and colleagues (2009) reported genetic variation in the ability of rice to break down starch reserves, with genotypes tolerant of anoxia showing better ability to mobilize and use stored reserves. The process was also associated with higher expression of *RAMY3D*, a member of the amylase gene family known to be expressed under anaerobic conditions.

Germinating rice seeds can therefore degrade starch to soluble sugars under anaerobic conditions. The next bottleneck is the use of these sugars as substrates to generate ATP, as oxygen is the terminal electron accepter in aerobic respiration. This is achieved through switching to anaerobic metabolism when oxygen is lacking (Avadhani et al. 1978; Jackson et al. 1982; Gibbs et al. 2000). The key enzymes for anaerobic fermentation are pyruvate decarboxylase (PDC), which catalyzes the decarboxylation of pyruvate to yield carbon dioxide and acetaldehyde, and alcohol dehydrogenase (ADH) that reduces acetaldehyde to ethanol, with the reduced form of nicotinamide adenine dinucleotide (NADH) being oxidized in the process to maintain glycolysis. We observed genetic variation in the extent of anaerobic respiration in rice cultivars contrasting in tolerance of anaerobic conditions during germination and early seedling growth (Ismail et al. 2009). Activities of PDC and ADH increased shortly after imbibition in both tolerant and sensitive genotypes, but were substantially higher in the tolerant genotypes. However, expression of *PDC1* and *PDC2* and *ADH1* and *ADH2* was simultaneously high in both tolerant and sensitive genotypes. The data support the notion that anaerobic respiration is important for the survival of rice seeds germinating under anaerobic conditions, and that tolerant lines are capable of greater induction of this pathway than sensitive lines; however, this induction seems to be regulated post-transcriptionally. Further studies are ongoing to identify and validate the traits that are associated with the tolerant phenotype

within rice for further analysis and use in breeding tolerant varieties.

Escape Strategy under Longer-term Partial Flooding

Water of 20 to 50 cm depth can accumulate and persist for most of the season in flood-affected rain-fed areas, and is referred to as medium-deep or stagnant floods (SF), and this sometimes occurs following short-term complete submergence, when the effects are even more devastating (Mackill et al. 2006; Singh et al. 2011). Modern rice varieties are not adapted to these conditions and their yield can decline markedly. This type of flooding received less attention in the past despite the enormous yield losses and the vast areas affected. In recent years, efforts have been devoted to screening large sets of diverse rice germplasm to identify genotypes with reasonable tolerance. Out of more than 700 accessions screened based on survival, yield, and yield components, fewer than one-third showed some tolerance (unpublished data). SF resulted in poor survival and low yield. Variation in elongation of the internodes of the tolerant genotypes was also observed in a manner dependent on the water depth (facultative elongation) just to keep pace with the rising water. Survival and yield under SF are largely dependent on the extent of this partial elongation, together with lesser carbohydrate depletion, thicker culms for increased internal oxygen transport, and high tillering underwater with more productive panicles with greater spikelet fertility.

Our recent studies highlighted the need to combine both *SUB1* and SF tolerance for areas where both stresses are experienced (Singh et al. 2011). As *SUB1* promotes survival of submerged plants by hindering shoot elongation to conserve energy reserves, its incorporation into shorter varieties makes them more sensitive to partial stagnant flooding. However, responses to SF were independent of *SUB1* introgression but more dependent on the background of the recipient genotype, with better performance of genotypes that are inherently taller. Varieties combining both tolerance of prolonged SF and *SUB1* are needed for broader adaptation in almost all flood-prone areas.

Under more extreme conditions, water can reach in excess of 100 cm to as much as a few meters and stagnate for several months, with the result referred to as deepwater or floating rice (Catling 1992). Elongation ability of leaves and internodes under these conditions is essential to keep up with the rising water levels and escape complete submergence. This will ensure an oxygen supply to roots via the continuum of aerenchyma tissue, to avoid anoxia and gain access to CO_2 and light to maintain the energy supply for such excessive growth (Setter et al. 1997). Traditional varieties adapted to these environments can elongate by more than 20 cm per day, but have poor grain quality and are low yielding due to their low tillering ability, long droopy leaves, and susceptibility to lodging. Areas under deepwater rice are now shrinking because of the lower productivity of the system and the spread of dry-season rice with the use of shallow tube wells, as in Bangladesh and Thailand, or due to better flood control, as in Vietnam. Internode elongation in deepwater and floating rice is induced by ethylene and GA, in a manner contrasting to that of the *SUB1* mode of action, and, recently two genes (*SNORKEL1* and *SNORKEL2*) that are responsible for this internode elongation under deepwater conditions were cloned (Hattori et al. 2009).

Future Prospects

Flooding during Germination

Development of Varieties Tolerant of Flooding during Germination

We have now introduced the major QTL *AG1* derived from Khao Hlan On (KHO) variety into IR64 and IR64-Sub1. The introgression of this QTL has also begun in other genetic backgrounds such as Ciherang-Sub1, PSB Rc18-Sub1, and PSB Rc82. We will introduce this QTL into more

popular varieties in the future. The introgression of the major QTL *AG2* derived from Ma-Zhan Red into several popular varieties has also begun.

It is interesting to note that most of the major QTLs controlling tolerance of flooding during germination or anaerobic germination derived from different tolerant donors identified so far are located in different positions in the rice genome (Angaji et al. 2010; Septiningsih et al. 2013; unpublished data). However, the contribution of each QTL to AG tolerance ranged from small to medium. This suggests that, unlike the case with the *SUB1* gene, the introduction of one QTL only into popular varieties or breeding lines will not give the level of tolerance needed. Assuming that their contributions are additive and there are no yield and quality penalties, the best strategy to increase the tolerance of AG stress is QTL pyramiding. The active pipeline of our research in discovering novel major QTLs from diverse tolerance donors will give us more opportunities in searching for the best combination of multiple QTLs for different target environments. Ultimately, these pyramided lines could be used directly as improved varieties or could be combined with tolerance of different types of flooding or some other key traits for direct-seeded environments or other abiotic and biotic stresses according to the needs of the target environments.

Unraveling the Genetic and Molecular Mechanisms underlying Tolerance of Flooding during Germination

Tolerance of AG is independent of tolerance of submergence during the vegetative stage, conferred by the *SUB1* QTL. In this case, the donor for the *SUB1* QTL, FR13A, is susceptible to AG. It has been reported that CIPK15 [calcineurin B-like (CBL)–interacting protein kinase] plays an important role in tolerance of AG (Lee et al. 2009). Additionally, transcript profiling of coleoptiles under anaerobic conditions (Lasanthi-Kudahettige et al. 2007) and germinating embryos under anoxia or anoxia followed by reoxygenation allowed the identification of several genes that are strongly regulated under AG stress (Narsai et al. 2009). However, these studies are just beginning to unravel the complex mechanisms underlying tolerance of AG. The genetic and molecular basis of tolerance of AG needs to be further investigated so that advances in knowledge can be applied to rice improvement. Tolerant donors identified at IRRI provide an excellent opportunity to further unravel the genetic and molecular basis of tolerance and to develop more resilient rice varieties that can overcome some of the current obstacles associated with rice cultivation.

Exploring the Genetic Control of Tolerance of Stagnant Flooding

Despite the development of several IRRI varieties and many elite breeding lines with tolerance of stagnant flooding, there is still a complete lack of genetic information regarding the control of stagnant flooding tolerance and extremely limited information regarding the physiology of this trait. Accordingly, elucidating the genetic control of stagnant flooding tolerance is a major objective of the breeding program. Efforts are under way to exploit shortcut QTL-mapping methods such as selective genotyping and early-generation mapping populations (e.g., F_2-derived bulks) in order to identify QTLs using classical QTL-mapping approaches in a relatively short time frame. A recombinant inbred population is also being developed for mapping QTLs from the IRRI 154 source. Plans are under way to evaluate a large range of elite and adapted IRRI material. The International Rice Germplasm Collection (IRGC) has also been exploited for identifying novel sources of tolerance. The use of backcross-derived mapping populations would be the most effective way to integrate QTL mapping with the development of new breeding material.

Alternatively, association-mapping approaches could be used. Given that considerable variation exists in elite breeding material of

the submergence breeding program, association mapping or family-based QTL approaches could be effective for exploiting routinely available breeding populations. The main constraint from the breeding program's perspective is the cost *per sample* of using high-throughput SNP assays as well as the expense in field trials for phenotyping. As the ideal number of markers for association mapping in rice is > 5,000, based on the level of linkage disequilibrium, high-throughput marker assays are essential (Mather et al. 2007). Therefore, using association-mapping methods with candidate gene markers and scanning specific regions delimited by QTLs are probably the most realistic methods that can be implemented within the breeding program, at least in the short term. Genome-wide association studies offer great potential for identifying new QTLs in germplasm collections, but this research cannot be performed by public-sector breeding programs.

Beyond the *SUB1* Gene

Efficient Development of Additional Sub1 Varieties

Currently several IRRI and other associated projects are under way to develop additional enhanced Sub1 varieties for Bangladesh, India, Nepal, Pakistan, Cambodia, Vietnam, and Africa. Furthermore, the *SUB1* gene has been incorporated into a wide range of IRRI's elite breeding material, across several programs, as a way to develop a broad range of varieties for different ecosystems across many countries, as an "insurance" against short-term flooding, even in areas where flooding does not normally occur.

Despite its success, MABC is a labor-intensive and resource-demanding exercise. To date, all of the Sub1 varieties have been developed using SSR markers. The advent of new high-throughput SNP genotyping platforms promise to improve efficiency in the future (Thomson et al. 2012), because the time required for background selection would decrease markedly. Although the cost per SNP data point is cheaper than SSRs, the cost per sample is still high (around US$40/sample). Therefore, in the majority of cases, SNP genotyping for background selection would need to be combined with SSR genotyping for foreground and/or recombinant selection. Alternatively, SSR genotyping could be adopted as a "first round" of background selection to reduce sample numbers for a "second round" of SNP genotyping, because simulation studies suggest that a small number of carefully selected markers can be used for accurate background selection in early backcross generations (Hospital et al. 1992; Visscher 1996).

The high SNP multiplex level of these new platforms could enable recombinant and background selection steps to be combined, potentially saving a considerable amount of labor, but this scheme would obviously depend on the funds available for the MABC program. Other medium through-put SNP genotyping platforms (24- to 96-plex) such as the Fluidigm EP1 system provide lower costs per sample, and may provide several new options for MABC. For example, combinations of genotyping with different SNP platforms could be implemented, or in different generations (M. Thomson, IRRI, pers. comm.). In general, computer simulations can provide useful guidelines regarding the use of combined low- and high-throughput marker systems (Herzog and Frisch 2011).

Development of Sub1^{plus} Varieties: Increasing Submergence Tolerance

Generally, the effect of *SUB1* is tolerance of submergence during flash flooding of up to 14 days. In some regions of Asia, flash flooding may occur for longer periods of time – up to 21 days (or even longer). The identification of additional QTLs for submergence tolerance is now a major objective of the IRRI submergence breeding program so that varieties with submergence tolerance of >14 days can be developed. Furthermore, the incorporation of additional QTLs could increase the survival and vigor of regeneration after submergence stress.

Compared to other abiotic stresses such as salinity or drought, relatively few QTLs (apart from the *SUB1* locus) for submergence tolerance in rice have been reported. In one of the most detailed studies, Toojinda and colleagues (2003) reported QTLs on chromosomes 1, 2, 5, 7, 10, and 11 using five phenotypic measurements and two donors (FR13A and Jao Hom Nin). In an earlier study, selective genotyping was used to identify QTLs on chromosomes 6, 7, 11, and 12 (Nandi et al. 1997). However, since primarily amplified fragment length polymorphism (AFLP) markers were used, it is difficult to determine QTL positions accurately. A summary of previously detected QTLs for submergence tolerance is shown in Figure 2.4. Apart from searching for and validating minor QTLs from FR13A, the research focus at IRRI is currently on using highly adapted sources, including IR72, in which a major QTL was identified on the long arm of chromosome 1 (Septiningsih et al. 2012), IR64, and IRRI 105 (PSB Rc18), since more

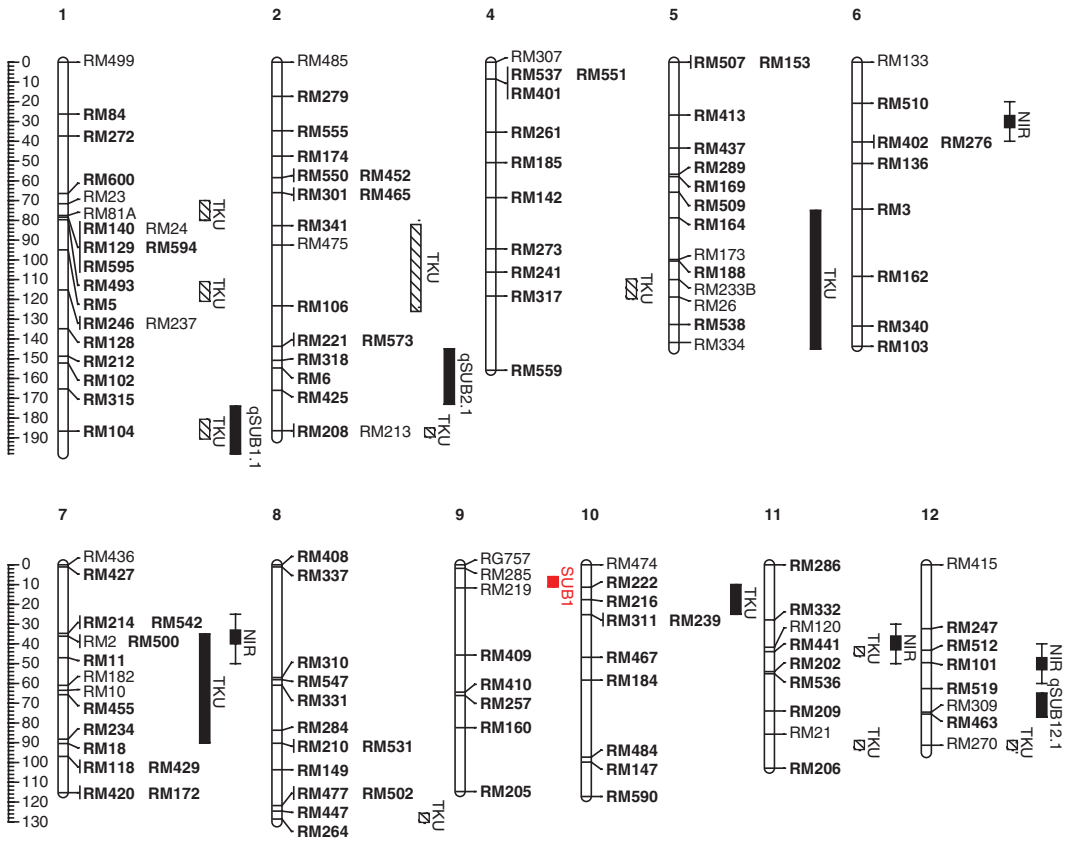


Fig. 2.4. Consensus QTL map for submergence tolerance. Linkage maps are derived from "anchor" SSR markers (i.e., common to both maps) from the "Cornell SSR 2001" and "CIAT SSR 2006" (www.gramene.org; Temnykh et al., 2000; Orjuela et al., 2010). Chromosome 3 is not shown because no QTLs have been detected on this chromosome. Anchor SSRs are indicated in bold and larger font; other markers are from the Cornell SSR 2001 map only, using genetic map distances. *SUB1* is shown on chromosome 9. Approximate QTL positions are represented from Toojinda et al. (2003; "TKU"), Nandi et al. (1997; "NIR"), and Septiningsih et al. (2012). Filled QTL bars indicate QTLs that were detected in multiple populations and/or component traits, whereas hatched QTL bars indicate minor QTLs detected in only single population or using a single component trait. Box plot QTL bars are indicated for QTLs from Nandi et al. (1997). QTL positions could only be estimated due to the low number of SSRs used in this study. Possible QTLs overlapping between studies were detected on chromosomes 7 and 11. The figure was produced using MapChart 2.1 (Voorrips, 2002). For a color version of this figure, please refer to the color plate.

rapid gains could be exploited in breeding using these sources (in contrast to unadapted sources such as landraces). Field evaluations across multiple sites and years for the latter two varieties indicate a moderate tolerance presumably due to the presence of minor or moderate-effect QTLs that are probably unlinked to *SUB1*.

Alternative QTL detection methods such as "marker-evaluated selection" (Steele et al. 2004), which are based on the principle of genotyping selected breeding material to identify genomic regions under selection, could also be useful for preliminary mapping of QTLs associated with this trait, because submergence is usually screened in the F_2 and F_3 generations. Germplasm evaluations have been undertaken to identify potentially new sources by phenotypic and genotypic screening of germplasm to identify novel sources of tolerance that could potentially be combined with *SUB1* (Singh et al. 2010).

Pyramiding of SUB1 with Other Abiotic and Biotic Stress-tolerance Traits

The identification of major tolerance genes and QTLs for important abiotic (e.g., drought, salinity, phosphorus deficiency) and biotic stresses will permit the use of marker-assisted pyramiding to combine multiple genes/QTLs in individual rice varieties. Bacterial blight and other diseases are prevalent in flood-prone regions and are exacerbated by flooding (Nino-Liu et al. 2006). As major genes for disease resistance have been identified, this provides opportunities for pyramiding bacterial blight, blast, and virus resistance with *SUB1*.

Rain-fed rice areas exposed to regular flash floods and stagnant flooding can also experience periods of low rainfall causing drought stress. Large-effect QTLs that confer tolerance of drought at the reproductive stage have been identified (Bernier et al. 2007; Vikram et al. 2011) and are available for pyramiding with *SUB1*. The pyramiding of submergence and salinity tolerance is also important, especially for coastal areas where floodwaters are often saline. A major QTL called *Saltol* has been identified on chromosome 1 and has been characterized recently (Thomson et al. 2010). However, multiple QTLs are probably required to achieve sufficient salinity tolerance in the field (G. Gregorio, IRRI, pers. comm.) and additional QTLs for seedling- and reproductive-stage salinity tolerance may be needed to provide protection from salinity stress throughout the season.

The opportunity to apply molecular marker technologies as a means of combining multiple tolerance genes/QTLs into individual rice varieties provides an unprecedented opportunity for breeders to rapidly develop tolerant cultivars for targeted environments. However, today the number of large-effect QTLs is still small and many reported QTLs are not sufficiently characterized, that is, data on the effect of QTLs in different genetic backgrounds and environments are not available. Without this information, it will be difficult for breeders to use QTLs in their breeding programs. With the rapid progress made in genome sequencing technologies and high-throughput marker technologies, it can be expected that more large-effect QTLs will be identified in the future. This will be even more important in light of climatic changes that adversely affect rice production, especially in rain-fed lowland ecosystems.

References

Alpi A, Beevers H (1983) Effect of O_2 concentration on rice seedlings. *Plant Physiology* 71:30–34.

Angaji SA, Septiningsih EM, Mackill DJ, Ismail AM (2010) QTLs associated with tolerance of flooding during germination in rice (*Oryza sativa* L.). *Euphytica* 172:159–168.

Atwell BJ, Greenway H (1987) Carbohydrate-metabolism of rice seedlings grown in oxygen deficient solution. *Journal of Experimental Botany* 38:466–478. doi:10.1093/jxb/38.3.466

Avadhani PN, Greenway H, Lefroy R, Prior L (1978) Alcoholic fermentation and malate metabolism in rice germinating at low oxygen concentrations. *Australian Journal of Plant Physiology* 5:15–25.

Bailey-Serres J, Fukao T, Ronald P, Ismail A, Heuer S, Mackill D (2010) Submergence tolerant rice: SUB1's

journey from landrace to modern cultivar. *Rice* 3:138–147.

Bailey-Serres J, Voesenek LACJ (2008) Flooding stress: Acclimations and genetic diversity. *Annual Review of Plant Biology* 59:313–339.

Barding GA, Jr., Fukao T, Beni S, Bailey-Serres J, Larive CK (2012) Differential metabolic regulation governed by the rice *SUB1A* gene during submergence stress and identification of alanylglycine by (1)H NMR spectroscopy. *Journal of Proteome Research* 11:320–330.

Bernier J, Kumar A, Ramaiah V, Spaner D, Atlin G (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Science* 47:507–518.

Biswas JK, Yamauchi M (1997) Mechanism of seedling establishment of direct-seeded rice (*Oryza sativa* L.) under lowland conditions. *Botanical Bulletin of Academia Sinica* 38:29–32.

Catling D (1992) Rice in deep water. International Rice Research Institute, Manila, Philippines.

Chen S, Lin XH, Xu CG, Zhang QF (2000) Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. *Crop Science* 40:239–244.

Collard BCY, Mackill DJ (2008) Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 363:557–572.

Das KK, Panda D, Sarkar RK, Reddy JN, Ismail AM (2009) Submergence tolerance in relation to variable floodwater conditions in rice. *Environmental and Experimental Botany* 66:425–434.

Das KK, Sarkar RK, Ismail AM (2005) Elongation ability and non-structural carbohydrate levels in relation to submergence tolerance in rice. *Plant Science* 168:131–136.

Drew MC (1990) Sensing soil oxygen. *Plant, Cell and Environment* 13:681–693

Ella ES, Kawano N, Yamauchi Y, Tanaka K, Ismail AM (2003) Blocking ethylene perception enhances flooding tolerance in rice seedlings. *Functional Plant Biology* 30:813–819.

Ella ES, Setter TL (1999) Importance of seed carbohydrates in rice seedling establishment under anoxia. In: *Proceedings of the Sixth Symposium on Stand Establishment and ISHS Seed Sympoisum,*pp 209-216

Fukao T, Bailey-Serres J (2008) Submergence tolerance conferred by *Sub1A* is mediated by *SLR1* and *SLRL1* restriction of gibberellin responses in rice. *Proceedings of the National Academy of Sciences of the United States of America* 105:16814–16819.

Fukao T, Xu K, Ronald PC, Bailey-Serres J (2006) A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* 18:2021–2034.

Fukao T, Yeung E, Bailey-Serres J (2011) The submergence tolerance regulator *SUB1A* mediates crosstalk between submergence and drought tolerance in rice. *Plant Cell* 23:412–427.

Gibbs DJ, Lee SC, Isa NM, Gramuglia S, Fukao T, Bassel GW, Correia CS, Corbineau F, Theodoulou FL, Bailey-Serres J, Holdsworth MJ (2011) Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. *Nature* 479:415–U172.

Gibbs J, Morrell S, Valdez A, Setter TL, Greenway T (2000) Regulation of alcoholic fermentation in coleoptiles of two rice cultivars differing in tolerance to anoxia. *Journal of Experimental Botany* 51:785–796.

Greenway H, Setter TL (1996) Is there anaerobic metabolism in submerged rice plants? A view point. In: Singh VP, Singh RK, Singh BB, Zeigler RS (eds) *Physiology of stress tolerance in rice*. Narendra Deva University of Agriculture and Technology and International Rice Research Institute, Manila, Philippines, pp 11–30.

Guglielminetti L, Yamaguchi J, Perata P, Alpi A (1995) Amylolytic activities in cereal seeds under aerobic and anaerobic conditions. *Plant Physiology* 109:1069–1076.

Hattori Y, Miura K, Asano K, Yamamoto E, Mori H, Kitano H, Matsuoka M, Ashikari M (2007) A major QTL confers rapid internode elongation in response to water rise in deepwater rice. *Breeding Science* 57:305–314.

Hattori Y, Nagai K, Furukawa S, Song XJ, Kawano R, Sakakibara H, Wu JZ, Matsumoto T, Yoshimura A, Kitano H, Matsuoka M, Mori H, Ashikari M (2009) The ethylene response factors *Snorkel1* and *Snorkel2* allow rice to adapt to deep water. *Nature* 460:1026–1031.

Hattori Y, Nagai K, Mori H, Kitano H, Matsuoka M, Ashikari M (2008) Mapping of three QTLs that regulate internode elongation in deepwater rice. *Breeding Science* 58: 39–46.

Herzog E, Frisch M (2011) Selection strategies for marker-assisted backcrossing with high-throughput marker systems. *Theoretical and Applied Genetics* 123:251–260.

Heuer S, Lu XC, Chin JH, Tanaka JP, Kanamori H, Matsumoto T, De Leon T, Ulat VJ, Ismail AM, Yano M, Wissuwa M (2009) Comparative sequence analyses of the major quantitative trait locus phosphorus uptake 1 (*Pup1*) reveal a complex genetic structure. *Plant Biotechnology Journal* 7:456–471.

HilleRisLambers D, Seshu DV (1982) Some ideas on breeding procedures and requirements for deepwater rice improvement. In: *Proceedings of the 1981 International Deepwater Rice Workshop*. International Rice Research Institute, Los Baños, Philippines, pp 29–44.

HilleRisLambers D, Vergara BS (1982) Summary results of an international collaboration on screening methods for flood tolerance. In: *Proceedings of the 1981 International Deepwater Rice Workshop*. International Rice Research Institute, Manila, Philippines, pp 347–353.

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. *Genetics* 132:1199–1210.

Hwang YS, Thomas BR, Rodriguez RL (1999) Differential expression of α-amylase genes during seedling development under anoxia. *Plant Molecular Biology* 40:911–920.

Iftekharuddaula KM, Newaz MA, Salam MA, Ahmed HU, Mahbub MAA, Septiningsih EM, Collard BCY, Sanchez DL, Pamplona AM, Mackill DJ (2011) Rapid and high-precision marker assisted backcrossing to introgress the *SUB1* QTL into BR11, the rainfed lowland rice mega variety of Bangladesh. *Euphytica* 178:83–97.

Ismail AM, Ella ES, Vergara GV, Mackill DJ (2009) Mechanisms associated with tolerance to flooding during germination and early seedling growth in rice (*Oryza sativa*). *Annals of Botany* 103:197–209.

Jackson MB, Herman B, Goodenough A (1982) An examination of the importance of ethanol in causing injury to flooded plants. *Plant, Cell and Environment* 5:163–172.

Jagadish SVK, Septiningsih EM, Kohli A, Thomson MJ, Ye C, Redoña E, Kumar A, Gregorio GB, Wassmann R, Ismail AMS, R. K. (2012) Advances in adapting rice to a rapidly changing climate. *Journal of Agronomy and Crop Science* 198:360–373.

Jiang L, Liu S, Hou M, Tang J, Chen L, Zhai H, Wan J (2006) Analysis of QTLs for seed low temperature germinability and anoxia germinability in rice (*Oryza sativa* L.). *Field Crops Research* 98:68–75.

Jiang L, Ming-yu HW, Ming C, Jian-min W (2004) Quantitative trait loci and epistatic analysis of seed anoxia germinability in rice (*Oryza sativa* L.). *Rice Science* 11:238–244.

Jung K-H, Seo Y-S, Walia H, Cao P, Fukao T, Canlas PE, Amonpant F, Bailey-Serres J, Ronald PC (2010) The submergence tolerance regulator *SUB1A* mediates stress-responsive expression of AP2/ERF transcription factors. *Plant Physiology* 152:1674–1692.

Kawano R, Doi K, Yasui H, Mochizuki T, Yoshimura A (2008) Mapping of QTLs for floating ability in rice. *Breeding Science* 58:47–53.

Konchan S, Kono Y (1996) Spread of direct seeded lowland rice in Northeast Thailand: Farmers' adaptation to economic growth. *Southeast Asian Studies* 33:523–546.

Lasanthi-Kudahettige R, Magneschi L, Loreti E, Gonzali S, Licausi F, Novi G, Beretta O, Vitulli F, Alpi A, Perata P (2007) Transcript profiling of the anoxic rice coleoptile. *Plant Physiology* 144:218–231.

Lee K-W, Chen P-W, Lu C-A, Chen S, Ho T-HD, Yu S-M (2009) Coordinated responses to oxygen and sugar deficiency allow rice seedlings to tolerate flooding. *Science Signaling* 2.

Licausi F, Kosmacz M, Weits DA, Giuntoli B, Giorgi FM, Voesenek LACJ, Perata P, van Dongen JT (2011) Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. *Nature* 479:419–422.

Mackill DJ, Amante MM, Vergara BS, Sarkarung S (1993) Improved semidwarf rice lines with tolerance to submergence of seedlings. *Crop Science* 33:749–753.

Mackill DJ, Coffman WR, Garrity DP (1996) Rainfed lowland rice improvement. International Rice Research Institute, Los Baños, Philippines.

Mackill DJ, Ismail AM, Pamplona AM, Sanchez DL, Carandang JJ, Septiningsih EM (2010) Stress tolerant rice varieties for adaptation to a changing climate. *Crop, Environment and Bioinformatics* 7:250–259.

Mackill DJ, Ismail AM, Singh US, Labios RV, Paris TR (2012) Development and rapid adoption of submergence-tolerant (Sub1) rice varieties. *Advances in Agronomy* 115:303–356.

Manzanilla DO, Paris TR, Vergara GV, Ismail AM, Pandey S, Labios RV, Tatlonghari GT, Acda RD, Chi TTN, Duoangsila K, Siliphouthone I, Manikmas MOA, Mackill DJ (2011) Submergence risks and farmers' preferences: Implications for breeding Sub1 rice in Southeast Asia. *Agricultural Systems* 104:335–347.

Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177:2223–2232.

Nandi S, Subudhi PK, Senadhira D, Manigbas NL, Sen-Mandi S, Huang N (1997) Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. *Molecular and General Genetics* 255:1–8.

Narsai R, Howell KA, Carroll A, Ivanova A, Millar AH, Whelan J (2009) Defining core metabolic and transcriptomic responses to oxygen availability in rice embryos and young seedlings. *Plant Physiology* 151:306–322.

Neeraja CN, Maghirang-Rodriguez R, Pamplona A, Heuer S, Collard BCY, Septiningsih EM, Vergara G, Sanchez D, Xu K, Ismail AM, Mackill DJ (2007) A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theoretical and Applied Genetics* 115:767–776.

Nemoto K, Ukai Y, Tang DQ, Kasai Y, Morita M (2004) Inheritance of early elongation ability in floating rice revealed by diallel and QTL analyses. *Theoretical and Applied Genetics* 109:42–47.

Nino-Liu DO, Ronald PC, Bogdanove AJ (2006) *Xanthomonas oryzae* pathovars: Model pathogens of a model crop. *Molecular Plant Pathology* 7:303–324.

Orjuela J, Garavito A, Bouniol M, David Arbelaez J, Moreno L, Kimball J, Wilson G, Rami J-F, Tohme J, McCouch SR, Lorieux M (2010) A universal core genetic map for rice. *Theoretical and Applied Genetics* 120:563–572.

Pampolino MF, Laureles EV, Gines HC, Buresh RJ (2008) Soil carbon and nitrogen changes in long-term continuous lowland rice cropping. *Soil Science Society of America Journal* 72:798–807.

Paris TR, Manzanilla DO, Tatlonghari GT, Labios RV, Cueno A, Villanueva D (2011) Guide to participatory varietal selection for submergence-tolerant rice. International Rice Research Institute, Los Baños, Philippines.

Pearce DME, Hall KC, Jackson MB (1992) The effects of oxygen, carbon-dioxide and ethylene on ethylene biosynthesis in relation to shoot extension in seedlings of rice (*Oryza sativa*) and barnyard grass (*Echinochloa oryzoides*). *Annals of Botany* 69:441–447.

Perata P, Geshi N, Yamaguchi J, Akazawa T (1993) Effect of anoxia on the induction of alpha-amylase in cereal seeds. *Planta* 191:402–408.

Perata P, Guglielminetti L, Alpi A (1997) Mobilization of endosperm reserves in cereal seeds under anoxia. *Annals of Botany* 79:49–56.

Sarkar RK, Panda D, Reddy JN, Patnaik SSC, Mackill DJ, Ismail AM (2009) Performance of submergence tolerant rice (*Oryza sativa*) genotypes carrying the *Sub1* quantitative trait locus under stressed and non-stressed natural field conditions. *Indian Journal of Agricultural Sciences* 79:876–883.

Sarkar RK, Reddy JN, Sharma SG, Ismail AM (2006) Physiological basis of submergence tolerance in rice and implications for crop improvement. *Current Science* 91:899–906.

Septiningsih EM, Ignacio JCI, Sendon PMD, Sanchez DL, Ismail AM, Mackill DJ (2013) QTL mapping and confirmation for tolerance of anaerobic conditions during germination derived from the rice landrace Ma-Zhan Red. *Theoretical and Applied Genetics* 126:1357–1366.

Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara GV, Heuer S, Ismail AM, Mackill DJ (2009) Development of submergence-tolerant rice cultivars: The *Sub1* locus and beyond. *Annals of Botany* 103:151–160.

Septiningsih EM, Sanchez DL, Singh N, Sendon PMD, Pamplona AM, Heuer S, Mackill DJ (2012) Identifying novel QTLs for submergence tolerance in rice cultivars IR72 and Madabaru. *Theoretical and Applied Genetics* 124:867–874.

Setter TL, Ella ES (1994) Relationship between coleoptile elongation and alcoholic fermentation in rice exposed to anoxia. 1. Importance of treatment conditions and different tissues. *Annals of Botany* 74:265–271.

Setter TL, Ellis M, Laureles EV, Ella ES, Senadhira D, Mishra SB, Sarkarung S, Datta S (1997) Physiology and genetics of submergence tolerance in rice. *Annals of Botany* 79:67–77.

Singh N, Dang TTM, Vergara GV, Pandey DM, Sanchez D, Neeraja CN, Septiningsih EM, Mendioro M, Tecson-Mendoza EM, Ismail AM, Mackill DJ, Heuer S (2010) Molecular marker survey and expression analyses of the rice submergence-tolerance gene *SUB1A*. *Theoretical and Applied Genetics* 121:1441–1453.

Singh S, Mackill DJ, Ismail AM (2009) Responses of *SUB1* rice introgression lines to submergence in the field: Yield and grain quality. *Field Crops Research* 113:12–23.

Singh S, Mackill DJ, Ismail AM (2011) Tolerance of longer-term partial stagnant flooding is independent of the *SUB1* locus in rice. *Field Crops Research* 121:311–323.

Sripongpangkul K, Posa GBT, Senadhira DW, Brar D, Huang N, Khush GS, Li ZK (2000) Genes/QTLs affecting flood tolerance in rice. *Theoretical and Applied Genetics* 101:1074–1081.

Steele KA, Edwards G, Zhu J, Witcombe JR (2004) Marker-evaluated selection in rice: Shifts in allele frequency among bulks selected in contrasting agricultural environments identify genomic regions of importance to rice adaptation and breeding. *Theoretical and Applied Genetics* 109:1247–1260.

Taylor DL (1942) Influence of oxygen tension on respiration, fermentation and growth of wheat and rice. *American Journal of Botany* 29:721–738.

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 100:697–712.

Thomson MJ, de Ocampo M, Egdane J, Akhlasur Rahman M, Godwin Sajise A, Adorada DL, Tumimbang-Raiz E, Blumwald E, Seraj ZI, Singh RK, Gregorio GB, Ismail AM (2010) Characterizing the *Saltol* quantitative trait locus for salinity tolerance in rice. *Rice* 3:148–160.

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung C-W, Reynolds A, Scheffler B, Eizenga G, McClung A, Kim H, Ismail AM, Ocampo Md, Mojica C, Reveche MY, Dilla-Ermita CJ, Mauleon R, Leung H, Bustamante C, McCouch SR (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding* 29:875–886.

Toojinda T, Siangliw M, Tragoonrung S, Vanavichit A (2003) Molecular genetics of submergence tolerance in rice: QTL analysis of key traits. *Annals of Botany* 91:243–253.

Tuong TP, Pablico PP, Yamauchi M, Confesor R, Moody K (2000) Increasing water productivity and weed suppression of wet seeded rice: Effect of water management and rice genotypes. *Experimental Agriculture* 36:71–89.

Turner FT, Chen CC, McCauley GN (1981) Morphological development of rice seedlings in water at controlled oxygen levels. *Agronomy Journal* 73:566–570.

Vartapetian BB, Jackson MB (1997) Plant adaptations to anaerobic stress. *Annals of Botany* 79:3–20

Vergara BS, Jackson MB, De Datta SK (1976) Deepwater rice and its response to deep-water stress. In: *Climate and Rice*. International Rice Research Institute, Los Baños, Philippines, p 301.

Vergara BS, Mazaredo A (1975) Screening for resistance to submergence under greenhouse conditions. In: *Proceeding of the International Seminar on Deep-water Rice*. Bangladesh Rice Research Institute, Joydepbur, Dhaka, pp 67–70.

Vikram P, Swamy BPM, Dixit S, Ahmed HU, Cruz MTS, Singh AK, Kumar A (2011) *qDTY(1.1)*, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. *BMC Genetics* 12. doi:8910.1186/1471-2156-12-89.

Visscher PM (1996) Proportion of the variation in genetic composition in backcrossing programs explained by genetic markers. *Journal of Heredity* 87:136–138.

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* 93:77–78.

Wassmann R, Jagadish SVK, Heuer S, Ismail A, Redona E, Serraj R, Singh RK, Howell G, Pathak H, Sumfleth

K (2009) Climate change affecting rice production: The physiological and agronomic basis for possible adaptation strategies. *Advances in Agronomy* 101:59–122.

Xu K, Mackill DJ (1996) A major locus for submergence tolerance mapped on rice chromosome 9. *Molecular Breeding* 2:219–224.

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442:705–708.

Xu K, Xu X, Ronald PC, Mackill DJ (2000) A high-resolution linkage map of the vicinity of the rice submergence tolerance locus *Sub1*. *Molecular and General Genetics* 263:681–689.

Yamauchi M, Aguilar AM, Vaughan DA, Seshu DV (1993) Rice (*Oryza sativa* L.) germplasm suitable for direct sowing under flooded soil surface. *Euphytica* 67:177–184.

Yamauchi M, Winn T (1996) Rice seed vigor and seedling establishment in anaerobic soil. *Crop Science* 36:680–686.

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. *Theoretical and Applied Genetics* 77:353–359.

Chapter 3

Genomics Applications to Salinity Tolerance Breeding in Rice

J. Damien Platten, Michael J. Thomson, and Abdelbagi M. Ismail

Abstract

Salinity tolerance is a complex trait, and it is typical for 5 to 8 significant quantitative trait loci (QTLs) to be identified in most mapping populations, and for many of these loci to control a relatively small proportion (15–30%) of the total phenotypic variance. Comparisons of QTLs across multiple populations and from multiple donors show that many QTL regions are shared between traditional donors, yet to date only a single gene (*OsHKT1.5*) has been shown to confer a QTL phenotype (*qSKC1*) in rice through map-based cloning. The slow progress in cloning these QTLs is partly due to large QTL intervals, a consequence of the expense of mapping larger populations, and partly due to the large number of genes that could potentially confer salinity tolerance annotated at each locus. However, the availability of cheap high-throughput sequencing and SNP genotyping technologies stands to revolutionize this process by providing unprecedented marker density. The high level of marker density should make association mapping feasible, thus freeing QTL discovery from the need to produce mapping populations. Second, QTL intervals derived from association mapping are much smaller, thus greatly aiding in narrowing the list of candidate genes. Third, the availability of whole-genome de novo or resequencing data makes it feasible to sort through candidate genes found in even quite large QTL intervals. The combined effect of these factors should greatly reduce the time and effort required to investigate the genetics of salinity tolerance in rice, with the corollary that the focus of research efforts is likely to shift towards producing accurate, detailed phenotype data and towards validation efforts for identified loci. At the same time, major QTLs for tolerance can be rapidly transferred through marker-assisted backcrossing (MABC) to develop improved varieties for salt-stressed environments. Likewise, as more genomics data lead to improved characterization of the genes and alleles controlling salinity tolerance, progress towards successful molecular breeding can be accelerated.

Introduction

The majority of domesticated crops did not evolve in areas affected by salt stress and do not naturally have strong tolerance mechanisms against salinity (Flowers et al. 2010). Salinity is a historic problem for agriculture, and as

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

cultivation intensifies and spreads to ever more marginal areas, the impact of salt stress continues to grow. Rice (*Oryza sativa* L.) is considered a relatively salt-sensitive species. However, considerable genetic variation for salinity tolerance exists in rice (Akbar et al. 1972; Moradi et al. 2003), and the interest in rice breeding for salt tolerance is gradually being renewed as a result of the latest developments in modern breeding tools and confidence in exploiting this variation.

The effects of salinity on plant growth are varied, but can be broken into a few broad categories: toxicity of the Na^+ ion, osmotic effects of high salt concentrations, toxicity of the Cl^- ion, secondary effects on mineral nutrition, and secondary effects on plant growth (Ismail et al. 2007). First and foremost is toxicity due to the Na^+ ion. Excess Na^+ has a general inhibitory effect on metabolism, and many tolerance mechanisms aim at either reducing the uptake of Na^+ under conditions when it is in excess or sequestering the Na^+ in areas where its inhibitory effects are less pronounced. The most well-known tolerance mechanisms fall into this category. For example, the *SOS1* protein acts as an Na^+/H^+ antiporter, using the H^+ gradient established across the cell plasma membrane to actively pump Na^+ that enters the cytoplasm out of the cell (Zhu et al. 1998; Shi et al. 2000, 2002). Likewise, a related Na^+/H^+ antiporter, NHX1, actively pumps cytoplasmic Na^+ into the vacuole, sequestering it where its effects on enzyme activity are less pronounced (e.g., Fukuda et al. 2004, 2011). Both of these pumps are tied to transmembrane H^+ concentration (pH) gradients. These gradients are in turn established and maintained in large part by the H^+-translocating ATPase and pyrophosphatase pumps. These energy pumps use chemical energy from phosphate-phosphate bonds to actively transport H^+ against its concentration gradient across the membrane, maintaining this gradient and thus energizing the aforementioned Na^+/H^+ antiporter pumps (e.g., Zhu 2003).

Another class of Na^+ transporters that has received considerable attention is the HKT gene family, as two members of this family (*HKT1;5* and *HKT1;4*, Platten et al. 2006) are to date the only transporters yet positively identified as the cause of naturally occurring variation in salinity tolerance in agriculturally important species (Ren et al. 2005; Byrt et al. 2007; Huang et al. 2006). This family actually facilitates the transport of Na^+ *into* the cell. This may appear to be a counter-adaptive feature, and indeed when members of the family are overexpressed they do result in increased Na^+ uptake and usually result in reduced salinity tolerance (Møller et al. 2009; Platten et al. unpubl.). However, the wild-type genes typically show extremely tissue- and treatment-specific expression profiles (Ren et al. 2005), and the action of *HKT1;4* and *HKT1;5* in particular seems to involve sequestering Na^+ in xylem parenchyma cells, where it presumably has fewer toxic effects than in photosynthetic leaf tissue. Both the HKT and NHX mechanisms are related to the number and volume of cells; consequently, varieties with increased biomass have greater reservoirs for sequestration of Na^+, and are known to have comparatively greater salt tolerance (Yeo et al. 1990; Flores et al. 2005).

The second broad effect of salinity is the simple osmotic stress imposed by high salt concentration (Munns et al. 1995; Bahaji et al. 2002; Castillo et al. 2007). Even if salinity concentrations change gradually (avoiding an osmotic shock), it still becomes increasingly difficult for plants to extract water and therefore nutrients as the salt concentration progressively increases in the soil solution. Nutrient imbalances can also occur due to displacement of some elements, such as potassium, under salt stress (Peng and Ismail 2004; Netondo et al. 2004; Nemati et al. 2011). Osmotic stress is not a consequence of the NaCl per se, but simply the high concentration of solutes; NaCl just happens to be the most common solute that occurs in excess. Tolerance of the osmotic component of the stress is conferred by increasing the osmotic potential of the internal fluids in the plant (Ismail et al. 2007; Nemati et al. 2011). This is achieved by increasing the concentration of various "compatible solutes" that

in themselves are either beneficial or have no inhibitory effects on metabolism. Examples are soluble sugars, free amino acids such as proline, and K^+ ions. The first two are synthesized organically, and thus up-regulation of the enzyme systems that produce them is known to confer tolerance. However, synthesis obviously incurs a significant metabolic cost and, in addition, the accumulation of organic compounds does not inherently prevent the accumulation of Na^+ to toxic concentrations. These "compatible solutes" are probably more important when the harmful ions such as Na^+ are effectively compartmented in the vacuoles, with the buildup of these solutes in the cytoplasm preventing internal dehydration, and consequently decreasing the osmotic potential of the root tissue to facilitate water and nutrient uptake (Ismail et al. 2007).

Third, high NaCl concentrations also result in high Cl^- levels. The effect of Cl^- is somewhat controversial. In many species, it is believed to have a significant toxic effect, whereas, in others, such as rice, it may not exert harmful effects or might even contribute to osmotic adjustment (Garthwaite et al. 2005; Diedhiou and Golldack 2006; Teakle et al. 2007; Brumos et al. 2009; Tavakkoli et al. 2010, 2011). Fourth, high salinity also causes a wide range of secondary effects. For example, high Na^+ in solution interferes with the solubility of other metal ions such as Ca^{2+} and many micronutrients. This can result in deficiencies, which can exacerbate the toxic effects of Na^+ and, in some situations, the toxic effects may partially result from these deficiencies. Application of exogenous Ca^{2+} is well known to reduce the severity of Na^+ stress (Tobe et al. 2003; An et al. 2004; Shabala et al. 2006; Zhang et al. 2011).

Another general consequence of salinity stress (whether ion toxicity, osmotic stress, or nutrients stress) is the effect on metabolism. Photosynthesis is adversely affected, through both stomatal closure and direct effects on the photosynthetic machinery (Yeo et al. 1985; Khan et al. 1997; Moradi and Ismail 2007). A significant increase in the concentrations of reactive oxygen species (ROS, i.e., hydrogen peroxide, hydroxyl radicals, and the like), which are highly toxic, was frequently reported as the consequence of salt stress (Moradi and Ismail 2007; Yamane et al. 2009; Ghosh et al. 2011). Many pathways exist for detoxification of ROS and up-regulation of these pathways often raises tolerance of salinity stress (Badawi et al. 2004; Sunkar et al. 2006; Moradi and Ismail, 2007; Hou et al. 2009; Lu et al. 2011). ROS are also required for cell expansion and as signaling molecules in various responses, and up-regulation of these pathways can sometimes lead to growth inhibition (Rodriguez et al. 2004; Taleisnik et al. 2009; Bernstein et al. 2010). Finally, Na^+ stress is directly sensed by the plant and can induce various hormonal and growth responses (Zhu 2003), though these sensing mechanisms and the sensors/receptors are still not known.

Mapping of Loci Associated with Salinity Tolerance in Rice

Over the past 15 years, numerous QTL studies have investigated the genetic basis for salinity tolerance in rice in a number of donors, resulting in the identification of many genetic loci controlling various physiological mechanisms related to tolerance. The most widely reported – though by no means the first – QTL study is that of Lin et al. (2004). They studied an F_2 population derived from a cross between the highly tolerant Indian *indica* landrace Nona Bokra and the highly sensitive *japonica* variety Koshihikari. A total of 11 QTLs from six non-overlapping loci were identified controlling root and shoot Na^+ and K^+ traits, and survival, and one QTL was subsequently fine-mapped and identified as the *OsHKT1.5* gene (Ren et al. 2005).

Comparisons of these results with other previously and subsequently published studies (Prasad et al. 2000; Gong et al. 2001; Bonilla et al. 2002; Takehisa et al. 2004; Ismail et al. 2007; Lee et al. 2007; Zang et al. 2008; Sabouri et al. 2009; Thomson et al. 2010; Alam et al. 2011) and our recent unpublished studies show

that the majority of these QTLs are robust, being identified in multiple populations from multiple donor parents. Thus, the major *SKC1* QTL is seen in nearly all populations derived from donors that originate from India/Bangladesh, that is, Nona Bokra, Pokkali, FL478 (derived from Pokkali), Kala Rata, Chikiram Patnai, and Cheriviruppu (Figure 3.1). The same is true for most QTLs; they are typically shared between several donors (Table 3.1). For example, regions around 11 Mb (*Saltol*/*SKC1*) and 39 Mb on chromosome 1, 28 Mb on chromosome 2, 5 Mb and 30 Mb on chromosome 3, 22 Mb on chromosome 6, and 8 and 25 Mb on chromosome 12 have been identified from numerous populations. In some cases, there is some doubt as to QTL position, as the markers used are difficult to accurately map to the reference genomes or to other genetic maps. In other cases, there may be more than one QTL present in the interval (e.g., chr01L and chr03L). Nevertheless, in the majority of these cases, the same genes are probably responsible for these overlapping QTLs.

The high incidence of overlap in QTL regions probably reflects the shared origins of many donors in India and Bangladesh, and their shared physiological mechanisms (dominated by Na^+ exclusion). Those few studies involving donors with no significant attachment to this geographic region (such as IR64 × Azucena and IR64 × Binam: Prasad et al. 2000; Zang et al. 2008) have a much higher proportion of unique QTLs. In

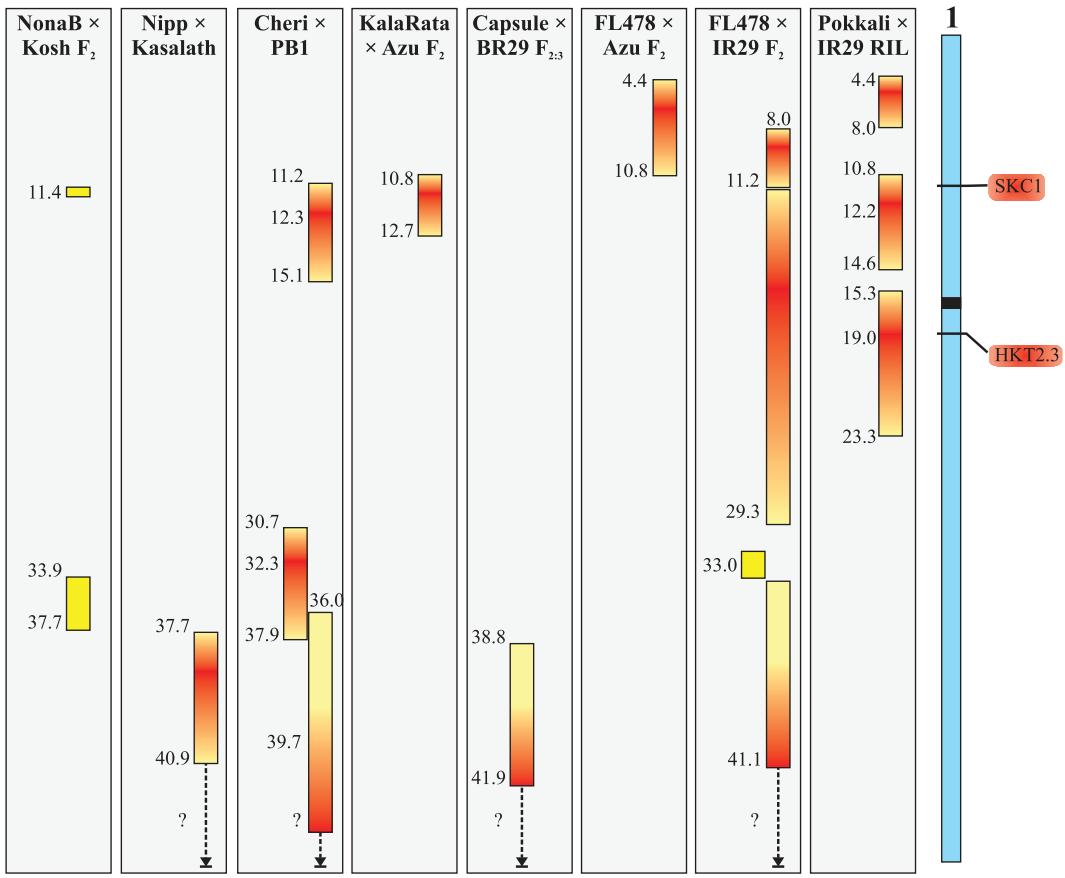


Fig. 3.1. Comparison of QTLs on a) chromosome 1 and b) chromosome 3 identified in various mapping populations. For a color version of this figure, please refer to the color plate.

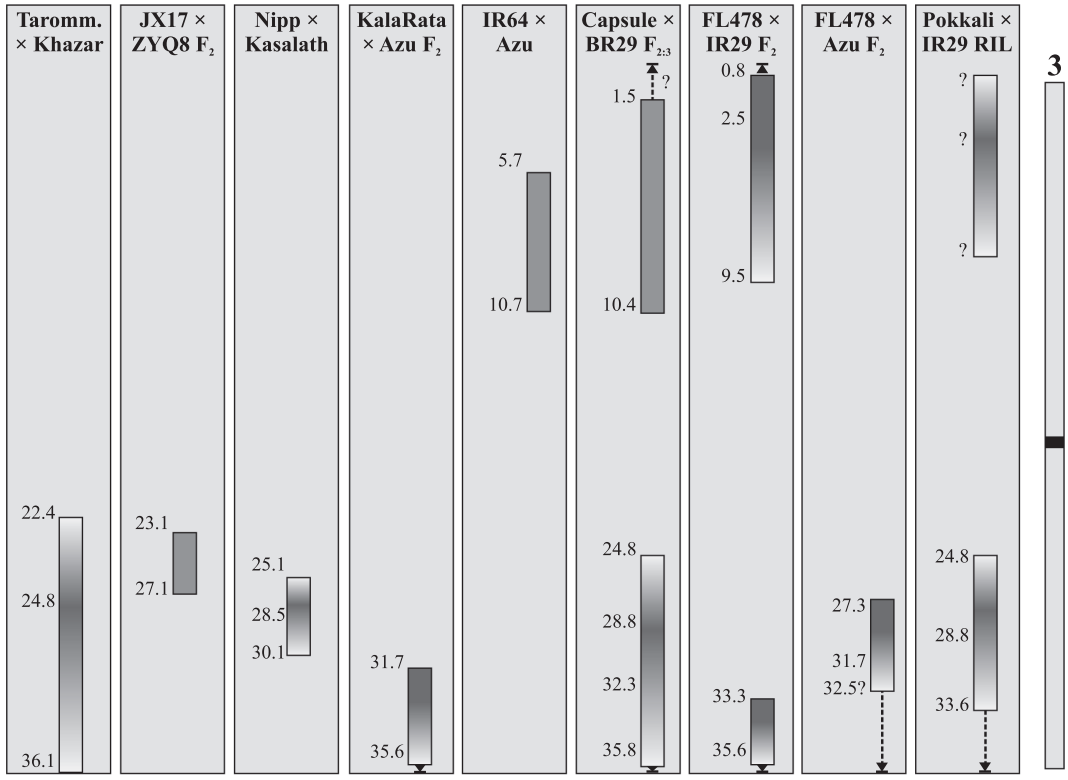


Fig. 3.1. (*Continued*)

addition, there do not appear to be any QTL studies conducted on physiological traits apart from general survival (e.g., duration of survival and/or visual injury assessed using standard evaluation scores) and traits related to Na^+ and K^+ content. Thus, it would appear that there is significant scope for the identification of additional QTLs through both screening for additional donor

Table 3.1. QTLs common to multiple mapping populations

Chromosome	Position (Mb)	Tolerant donors
1	11	Pokkali, Nona Bokra, Chikiram Patnai, Kala Rata, FL478, Cheriviruppu
1	39	Nona Bokra, FL478, Capsule, Cheriviruppu, Chikiram, Patnai, Gihobyeo, Nipponbare
2	28	Capsule, Kala Rata, FL478, IR64, Kasalath
3	5	Pokkali, IR64, FL478, Chikiram Patnai, Capsule
3	30	Pokkali, Kasalath, Kala Rata, FL478, Cheriviruppu, Capsule
4	23	Nona Bokra, Boilam, Chikiram Patnai, IR64, JX17
5	24	Capsule, Kala Rata, IR64
6	5	FL478, IR64, JX17
6	22	Nona Bokra, Pokkali, FL478, JX17, Tarommahalli
7	4	Nona Bokra, Nipponbare, IR64, Boilam
9	15	Nona Bokra, Pokkali, IR64
12	8	Pokkali, Kala Rata, FL478, Capsule
12	25	Pokkali, JX17, FL478

germplasm outside of India and Bangladesh and for additional physiological mechanisms such as ROS scavenging, compartmentation in older tissue, and tissue tolerance (Ismail et al. 2007).

Marker-assisted Backcrossing to Use Salt Tolerance QTLs for Breeding

As more QTLs are identified, they present opportunities for breeding programs to improve the salinity tolerance of rice varieties by precisely transferring QTLs using marker-assisted backcrossing (MABC). Although conventional backcrossing can transfer desired genes into recipient varieties, the use of markers can accelerate the process by speeding up background selection for recurrent parent alleles and reducing the size of the target introgression, thus reducing the chances for negative linkage drag (Young and Tanksley 1989; Collard and Mackill 2008; Ismail and Thomson 2011). An excellent example is the MABC transfer of the *SUB1* QTL for submergence tolerance into numerous popular rice varieties, leading to a substantial increase in yield after recovery from variable durations of submergence, ranging from a few days to more than 2 weeks (Septiningsih et al. 2009; Mackill et al. 2012). A MABC program has also begun for improving the salinity tolerance of rice, starting with the *Saltol* QTL being transferred into popular varieties such as IR64, BR11, BRRI dhan28 (BR28), and BRRI dhan29 (BR29), using FL478 as the donor of the tolerant Pokkali allele (Thomson et al. 2010). In contrast to the *SUB1* story, however, the *Saltol* locus provides an intermediate increase in tolerance on its own – evidently multiple salinity tolerance QTLs will need to be pyramided in the same genetic background to provide adequate tolerance for stress-prone environments (our unpublished data). Thus, the development of NILs for different QTLs and the combining of multiple tolerance QTLs will need to be pursued to obtain the desired tolerance for target environments. This pyramiding of QTLs is required for both seedling- and reproductive-stage tolerance, since tolerance at the two stages is weakly associated (Moradi et al. 2003).

Cloning of QTLs Associated with Salinity Tolerance in Rice

Despite many advances in molecular biology, in both technical and theoretical understanding, only one salinity tolerance QTL from rice has been cloned so far (*qSKC1* = *OsHKT1.5*; Ren et al. 2005). This in large part reflects the difficulty of QTL cloning, and it is instructive that in this case the cloning was carried out essentially via ultra-fine-mapping of a single major QTL controlling almost 50% of the total phenotypic variance, from a simple F_2 mapping population. Other mapping population types (RIL, doubled-haploid, etc.) provide significant advantages in terms of power and repeatability, but make fine-mapping much harder. The majority of QTL regions are very poorly mapped, the QTL often being described by just 2 to 3 marker positions and often spanning 10 Mb or more of the Nipponbare reference genome. Thus, as might be expected, hundreds of candidate genes are typically found within any given QTL interval. In addition, most salinity tolerance QTLs explain much smaller proportions of the total phenotypic variance, which limits the power of QTL detection and makes definition of QTL limits more difficult.

To some degree, comparison of QTLs identified between different mapping populations can help to narrow QTL limits. For example, on the long arm of chromosome 1 (Figure 3.1a), a QTL explaining a significant proportion of the phenotypic variance is observed in many populations, with a position varying from 29.8 Mb through to the chromosome end (approx. 43.5 Mb). However, comparison of QTLs identified between mapping populations suggests that the QTL is probably located in the region 37-42 Mb, and including a QTL from the Nipponbare × Kasalath population (for which Nipponbare was the donor; Takehisa et al. 2004) helps pin a peak at around 39-40 Mb. However, Pokkali

appears to have a different QTL peak at 31-36 Mb, and indeed this is thought to be derived from the sensitive parent in that population, IR29 (Thomson et al. 2010). Comparison of this putative QTL with other populations tends to support the hypothesis of two QTLs in the region, as QTLs for multiple traits that were not fully overlapping were also observed in Nona Bokra, Cheriviruppu, and one FL478 population, while the "upper" QTL alone was observed in Kala Rata × Azucena and JX17 × ZYQ8 populations (Thé et al. unpubl.; Gong et al. 2001). This comparison also tends to suggest that the lower QTL is primarily responsible for affecting biomass-related traits, with possibly secondary effects on Na^+ concentrations. However, even after taking into account information from all available populations, most QTL intervals are still on the order of 7–10 Mb, relative to the Nipponbare reference genome. This appears to be largely due to an inherent limitation in QTL studies: that QTL limits are defined by both population size and marker density, and both are typically minimized to save time and reduce costs. In addition, some marker types are difficult to map onto the reference genome, making precise definition of QTL limits difficult or impossible. Thus, comparison of QTLs from multiple populations is very powerful in helping to determine the prevalence and effect of a QTL, but typically cannot narrow QTL limits sufficiently to allow unambiguous determination of the underlying genes, aside from the expense and time required to develop these multiple populations.

Another complicating factor in the use of QTLs in plant breeding is the phenomenon of epistasis, often attributed to and described as the "genetic background" effect. Epistasis is the case in which the effect of an allele (*A*) at one locus (A) masks the effect of genetic variation at another locus (B) (*B/b*); a different allele at the first locus (*a*) may unmask the effects at locus (B). Consensus is growing that epistasis plays a very important role in determining the extent of salinity tolerance displayed by rice plants. For example, a minor QTL around the centromere of chromosome 1 from Pokkali, FL478, and IR64 may be caused by a 1-bp deletion in the coding region of *OsHKT2.3*, a gene related to other genes known to increase shoot Na^+ content. The presence of a functional copy of this gene would thus counteract the effects of tolerant alleles at the *Saltol* locus. Extending this further, it can be hypothesized that specific combinations of genes may mask the effects of a gene responsible for a QTL, and thus a QTL may not be seen in certain genetic backgrounds even when present. This becomes important when considering the use of marker-assisted backcrossing, especially if "anonymous," non-gene-specific markers such as flanking SSRs (simple sequence repeats) are used for this purpose. Recipient lines for such marker-assisted backcrossing are chosen due to their lack of the QTL phenotype, i.e., salinity tolerance, and it is assumed that they lack significant loci conferring this phenotype. Markers are then chosen for their polymorphism between the donor and recipient lines, and the assumption is made that the marker haplotype (forward and flanking markers) from recipient lines denotes a sensitive haplotype. However, this is not necessarily true. Only those markers polymorphic between parents are chosen, and the combination of marker genotypes (the marker haplotype) is specific for one or the other parent. But unless one or more markers is gene-specific, the marker haplotype of the recipient parent does not necessarily indicate that the recipient parent (or progeny that have that haplotype) lacks the QTL; this is an assumption based on the phenotype. It is thus perfectly possible to use MABC to introgress a QTL into a recipient line that already contains that QTL, but whose effects are masked due to genetic background. This may explain some of the difficulties encountered in the MABC of *Saltol* into the variety BR28: the MABC was performed without any problem, but the resulting BR28-Saltol lines have limited improvement in salinity tolerance. Preliminary sequence evidence of *OsHKT1.5* from BR28 suggests that BR28 already contains an allele identical to that from Nona Bokra, and that

would presumably be active (our unpublished data). Yet, BR28 is manifestly sensitive, raising the possibility that the BR28 genetic background may have additional loci suppressing the effects of *Saltol*, at least on whole-plant performance. More data are needed to explore the *SKC1* alleles in other recipient varieties and further investigate the usefulness of the Pokkali *Saltol* QTL for breeding applications.

Thus, cloning of QTLs is of significant value in plant breeding, both in terms of developing perfect markers for MABC and in accurate identification of potential recipient lines. Yet, the cloning of QTLs still relies on one of two techniques, both costly and labor-intensive. Either the QTL must be fine-mapped in a large population to narrow down the limits, or researchers must sort through many candidate genes, testing such parameters as expression changes under treatment, sequence polymorphisms (cloning), and eventually transgenic validation. Both approaches have been used successfully in salinity research (fine-mapping of *SKC1*, Ren et al. 2005; the candidate gene approach for wheat *Nax1* and *Nax2*, Byrt et al. 2007; Huang et al. 2006), but neither is simple or "high-throughput" in nature. Yet, the large number – and relatively small effect – of many salinity-tolerance QTLs means that, to achieve high tolerance, multiple QTLs will need to be introgressed into any potential recipient germplasm. This will multiply the number of genes that must be identified, while at the same time making it more and more likely that the mutations responsible for tolerance alleles will be subtle in nature and possibly difficult to identify. Thus, it remains challenging to identify the genes and causal mutations responsible for salinity-tolerance QTLs. In some respects, approaches that do not rely on QTL isolation, such as marker-assisted recurrent selection and genomic selection, may prove useful as alternative molecular breeding strategies in the future (Bernardo 2008). However, the benefits of isolating the genes underlying QTLs makes the challenge worthwhile, since it enables more efficient allele mining, development of perfect markers, and characterization of the molecular mechanisms underlying salinity tolerance.

Next-generation Sequencing: Advances and Limitations

The significant advances made in genomics techniques, from microarrays to very-high-throughput sequencing and SNP detection, provide enormous opportunities to sidestep and overcome these limitations in QTL mapping, candidate gene identification, and marker-assisted breeding. Much of the utility of next-generation genomics techniques revolves around the revolution in sequencing technology that started with MPSS (massively-parallel signature sequencing; Brenner et al. 2000) and was brought to its first widely useful applications in the 454 pyrosequencing (454 Life Sciences, www.my454.com; now part of Roche Diagnostics) and Illumina (formerly Solexa, www.illumina.com) sequencing platforms. Other platforms include sequencing-by-ligation techniques such as those used by the Applied Biosystems SOLiD platform and the Complete Genomics system, the IonTorrent system, Pacific Biosciences' single-molecule real-time (SMRT) sequencing, and methods still under development, such as Oxford's Nanopore sequencing technique.

Many of these technologies are based around principles similar to the dideoxy terminator technique used in traditional Sanger sequencing. However, they differ in a few key details. The Sanger process involves a single, static reaction using a single, defined primer on a single, pure template, and produces a complex mixture of products that are subsequently resolved via a high-resolution electrophoresis system. This allows the Sanger process to produce long reads of high-quality contiguous sequence, but it limits the process to sequencing pure, defined templates (e.g., PCR product or cloned DNA) at defined locations (determined by the sequencing primer). This makes it difficult, laborious, and

expensive to sequence whole genomes, as these must first be fragmented and cloned to produce sequencing libraries, which must then be amplified and extracted (per clone), and sequenced via primer walking until a contiguous sequence is produced.

Next-generation sequencing approaches (such as the Illumina and Roche systems, as the most common systems in current use), on the other hand, employ techniques that sidestep many of these difficulties. The key innovation in both is that sequencing products are no longer resolved and read off linear sequencing gels. Rather, the sequencing template is immobilized on a solid, 2-dimensional substrate (the sequencing chip), and the sequencing results "read" by taking ultra-high-resolution images of this substrate. Each image is capable of reading only a single base (or type of base, for 454) of sequencing at a time, so these techniques no longer involve irreversible termination of the sequencing synthesis step. Rather, they either employ reversible termination (Illumina) or simply add only a single type of nucleotide at a time (454). Thus, sequencing proceeds by only one or a few base-pairs at a time, and so is conducted in cycles. Results are read after each cycle and built up incrementally, and finally put together in post-sequencing processing (details can be found on the respective websites, www.illumina.com and www.my454.com).

Although this may appear laborious, it offers two related key advantages. First, since the sequencing occurs on templates immobilized on a 2-dimensional substrate, the original template does not have to be pure. The template is "purified" by ensuring that each region of the substrate (sequencing chip) receives only a single template molecule and is thus pure. Second, since each region of the sequencing chip may have a different template molecule, sequencing must proceed using universal primers. This necessitates a prior fragmentation and ligation step in which the primer complementary sequence is ligated to template fragments, but the result is that sequencing reactions are now no longer dependent on the template sequence, that is, the results come close to being template-independent. Thus, it opens the possibility of massive multiplexing, limited essentially by the spatial resolution of the imaging system (and storage capacity of computers acquiring the data). Thus, these systems have become characterized by extremely high-throughput results, producing tens (and now hundreds) of millions of reads per sequencing chip.

But, this has come at a cost: both systems rely on reactions that are not 100% accurate, and accuracy declines as the number of cycles increases. Thus, both technologies started off producing only very short reads (as short as 25 nucleotides), and have gradually extended this to approximately 400-500 bp for 454 sequencing and 100-150 bp for Illumina. The short reads, combined with the fact that the template used in each sequencing region can be sequenced at most only twice (once from each end), makes it far more challenging to use these technologies to assemble genomes de novo, that is, without reference to a previously sequenced genome (e.g., Alkan et al. 2011). On the other hand, the exceptionally large number of reads produced – particularly by the Illumina system – does make them extremely useful at resequencing a known genome, or one fairly closely related to a known genome. The large number of reads produces very high coverage per base, resulting in very high confidence of the sequence at each position, even with relatively modest coverage such as 20×. In addition, the high coverage and relative template-independence of the reads also permit the identification of heterozygous sites. Thus, one of the first applications of these technologies has been in SNP discovery and associated applications in association mapping.

Application of Next-generation Sequencing Technologies to Salinity Tolerance Research

SNP Discovery and QTL Identification

The advantages and utility of next-generation sequencing technologies in single nucleotide

polymorphism (SNP) identification have already been outlined. Uptake of these technologies for this purpose is by now pushing rice research well beyond the SNP diversity knowledge generated by the OryzaSNP project (McNally et al. 2009). For example, the OryzaSNP project identified approximately 160,000 SNPs among the 20 varieties sequenced, and ~100 Mb of reference genome covered in the resequencing project. This dataset has been used for a variety of purposes, including the design of high-density SNP genotyping arrays (Zhao et al. 2011). On the other hand, the availability of high-throughput, low-cost Illumina sequencing has allowed the resequencing of approximately 70 varieties of *O. sativa*, 14 accessions from the *O. rufipogon – O. nivara* complex, 7 landraces of *O. glaberrima*, and 8 accessions of *O. barthii* through the Rice SNP Consortium (www.ricesnp.org). The International Rice Research Institute (IRRI) is also in the process of sequencing 3,000 accessions of *O. sativa*, chosen for their diversity and relevance to research, as the initial set in a larger project to resequence about 10,000 rice accessions (K. McNally pers. comm.).

The low cost of this sequencing also allows even single laboratories on modest budgets to generate significant quantities of useful information. For example, one service provider (Macrogen Inc., Korea) currently offers sufficient sequencing data to provide 24× coverage of the *O. sativa* genome for approximately US$3,000 on the Illumina platform. Using this service, we have resequenced six varieties relevant to salinity tolerance breeding (Pokkali, Capsule, FL478, Hasawi, and IR29 from *O. sativa* and a highly tolerant accession of *O. glaberrima*). Comparison of the *O. sativa* varieties shows that these five *indica* and *aus*/admixed varieties typically possess 3.1-3.5 million homozygous polymorphisms relative to the Nipponbare reference genome (Table 3.2); other *indica* and *aus* varieties might be similar.

This amount of polymorphism is obviously far higher than is required for typical QTL mapping approaches. However, it overcomes a key limitation of association mapping; due to the high degree of recombination that has occurred during the many generations that separate most varieties, linkage between markers and traits must be far tighter to produce significant association. Thus, association mapping requires a far higher marker density than QTL mapping, which has traditionally limited its usefulness in plant breeding. The impending availability of these high-density genotyping data, particularly if available directly from the parental lines of a QTL population, makes it a relatively simple process to extract sufficient polymorphic sites to allow the construction of a high-density map using SNP detection platforms such as custom oligonucleotide arrays from Affymetrix or Illumina. Such SNP information has already been used in association mapping studies in rice, to investigate agronomic characters and aluminum tolerance (Famoso et al. 2011; Zhao et al. 2011). In many cases, the high SNP density and extensive linkage decay present in natural populations allowed direct identification of candidate genes; in some cases, actual candidate mutations can be identified.

Likewise, the availability of large sets of SNPs genotyped across diverse germplasm provides a valuable resource for selecting targeted

Table 3.2. Coverage and polymorphism statistics for four *O. sativa* accessions resequenced on the Illumina GAIIx and HiSeq2000 platforms

Accession	Species group	Coverage (approximate)	# homozygous polymorphic sites	% polymorphic sites
FL478	*indica*	25×	3,214,904	0.86
Capsule	*indica*/admix	35×	3,207,701	0.86
Pokkali	*indica*/admix	55×	3,507,395	0.94
Hasawi	*aus*/admix	55×	3,178,538	0.85

subsets of SNP markers for breeding and genetics applications. For example, the SNP data from the 44,100 SNP genotyping array have been used to select informative sets of 384 SNP markers for running on an Illumina BeadXpress Reader targeted for specific germplasm groups for genetic diversity analysis, QTL mapping, and background selection (Thomson et al. 2012). In addition, more specific sets of SNPs for certain traits or chromosome regions can be selected from the SNP discovery pools and run on more flexible genotyping platforms such as the Fluidigm EP1 system. This SNP platform is currently being tested at IRRI to assist in fine-mapping the salinity QTL on the long arm of chromosome 1, in addition to targeting gene-specific SNP markers at loci of interest.

Previous association studies have relied on the 44,100 SNP genotyping array (Zhao et al. 2011), which is convenient but, due to sampling bias, may not provide optimum resolution for all genotypes. Researchers are also starting to use sequencing results directly to produce ultra-high-density datasets for association mapping. For example, two recent papers have used low-coverage sequencing data (approximately 0.5 – 1×), coupled with an imputation algorithm that makes use of haplotype information, to fill in sites with missing data, to conduct proof-of-principle association mapping studies on yield and agronomic trait data in rice (Huang et al. 2010, 2012). As mentioned above, even modest-coverage data are now relatively cheap, so producing low-coverage data like these for multiple accessions via multiplexing of the sequencing libraries should soon be within reach of most research groups, which should facilitate the rapid generation of high-density data for nearly every genetic research question. This approach is further facilitated by a genotyping by sequencing (GBS) approach employing a restriction digestion before sequencing, allowing even greater multiplexing and a lower cost per sample (Elshire et al. 2011). The costs for GBS are now approaching an amount at which it will soon be feasible to sequence entire mapping populations faster and more cheaply than running SSR markers, ushering in a new era of low-cost, high-resolution genetic mapping. One current limitation is the higher informatics capacity needed to analyse large amounts of sequence data, which requires improved analysis pipelines and computing facilities made available in a user-friendly environment for routine use by molecular breeders and geneticists.

A further consideration is that, unlike QTL mapping, the marker datasets used in association mapping are not limited to a single population or a single trait. Once the marker (e.g., SNP) information is available, the accessions used for association mapping can be phenotyped for multiple traits. In addition, if a particular phenotype is rare, it is relatively inexpensive to genotype a few extra accessions relevant to that trait and add these data to the association mapping dataset. Therefore, as the data from these large sequencing projects become available, the identification of loci controlling a trait will no longer be limited by choice of parents, and determination of a genomic location will no longer be limited by low marker density, low recombination rates, and difficulties of correspondence between genetic maps. Rather, limitations will continue to be the accuracy and level of detail available in the phenotyping data, and considerable research focus might be expected to shift from genotyping efforts on multiple populations to phenotyping and validation.

Identification of Candidate Genes

The utility and importance of cloning and identifying the genes responsible for QTLs have already been mentioned briefly. Progress in this area is hampered by the laborious nature and high cost of both fine-mapping and cloning/sequencing of a large number of genes. It is immediately obvious that association mapping stands to provide significant advantages in terms of reducing the size of QTL intervals and has the added benefit that multiple donors will be

identified. The identification of the actual mutation, and therefore gene, responsible for a QTL requires a level of marker density well beyond that typically employed by even association mapping procedures, but is still nevertheless well within reach of the data produced by, for example, the Illumina system. Combined with appropriate data mining and candidate gene identification, it even becomes possible to efficiently process quite large QTL intervals.

As a case in point, it has been mentioned that resequencing data from five *O. sativa* genotypes have been obtained in an attempt to identify the causal mutations underlying QTLs from various populations derived from these genotypes. One significant QTL identified from a number of mapping populations is located on the long arm of chromosome 3 (Figure 3.1b). This has been identified in populations derived from the crosses Capsule (tolerant) × BRRI dhan29 (sensitive), FL478 × Azucena, FL478 × IR29, and Pokkali × IR29. In all cases, the tolerant parent is likely to be the donor of the tolerant allele, so that FL478, Pokkali, and Capsule are likely donors, and IR29, Azucena, and BRRI dhan29 are recipients. The QTL was also identified in a population derived from a Nipponbare × Kasalath cross, in which Nipponbare was the donor (Takehisa et al. 2004). Traits mapped to this location are quite diverse, but typically include parameters related to biomass accumulation and injury scores, together with some effects on Na^+ and/or K^+ contents. Comparison of QTL positions mapped onto the Nipponbare reference genome shows that the QTL interval spans from approximately 26 Mb down to the end of the chromosome, with a peak between 29 and 31 Mb depending on the population (it is also possible that two QTLs exist in coupling, with ranges of 25 - 30 Mb and 29 Mb to the end, and peaks of 27 Mb and 33 Mb, respectively). Thus, the inclusive QTL interval spans around 10 Mb of the reference genome. Within this region, 1729 nonredundant gene models are found within either (or both) of the MSU v. 6.1 and RAP 4 annotations, of which 424 (∼25%) can be classified as potential candidate genes for the QTL (our unpublished data).

Clearly, it is not practical to examine all of these candidates via traditional cloning techniques. Cloning of the most likely candidates (based on annotated function) revealed those that showed little or no polymorphism between IR29 and/or Azucena and the tolerant donors. Thus, a very long list of secondary candidates remained. Using expression data obtained from Affymetrix arrays (Walia et al. 2005, 2007; Cotsaftis et al. 2011), it was possible to narrow this list to just 30 to 40 plausible candidates, but this number is still large to validate via traditional approaches. However, combined with polymorphism data derived from whole-genome resequencing, only three showed patterns of polymorphism fully consistent with the pattern of presumed donor/recipient combinations, located at 27, 29, and 35 Mb (Figure 3.1b). These positions clearly coincide well with the noted QTL peaks, although this does not help in determining whether one or two QTLs are present. But this number of candidate genes is quite feasible for traditional cloning and transgenic validation approaches. Thus, next-generation sequencing data make it possible to search through even quite large and complex QTL regions in an efficient manner and successfully identify good candidate genes for further validation.

As an adjunct to the identification of candidate genes, a very large number of SNP polymorphisms were identified within the QTL region. Once these SNPs and positions are known, it becomes possible to filter these and select a set of high-quality, robust SNPs for further fine-mapping of the QTL. As an example, SNP data can be post-processed in MS Access to (A) retrieve SNPs within genic (or coding) regions, (B) filter for SNPs that are present in donor lines but NOT present in any recipient line, and (C) retrieve SNPs with few or no polymorphisms in surrounding areas (e.g., a 500-bp window centered on the SNP site) and that are therefore likely to produce good markers. Filtering for SNPs that are present in most or all donor lines but not

present in any recipient lines increases the probability that each SNP will be useful in multiple mapping populations, particularly if data from diverse recipient (and donor) lines are available. Once this procedure is in place, it is a simple matter to inspect the QTL region and choose between 10 and 20 robust SNPs for marker design, with a view to producing a fine linkage map of the QTL region. These SNP markers can then be genotyped on larger populations using low-cost SNP genotyping platforms for fine-mapping and subsequent deployment of these markers in a breeding program for marker-assisted selection.

Conclusions

As these examples illustrate, the availability of cheap high-throughput sequencing technologies, combined with some lateral thinking in analysis techniques, stands to revolutionize the way the genetics of quantitative traits are analyzed, such as the case with salt tolerance. These technologies should provide a level of precision that has never yet been matched in terms of marker density. As data from significant numbers of genotypes become available, association mapping will finally free QTL discovery from the expense and restrictive nature of mapping populations, and resequencing information should allow the rapid identification of genes underlying association or traditional QTLs. The low cost of these technologies should allow even modest-sized research groups to make significant advances. Thus, the genetics of quantitative traits may be expected to mature rapidly, and the emphasis will shift back to deriving accurate and detailed phenotype information, both as information to feed into association mapping and to validate identified candidate genes. The applications of NGS (next-generation sequencing) will be of particular importance for complex traits for which QTLs of relatively large effects could be identified, and where several donors of novel QTLs are available, as is the case with salinity tolerance in rice. Resequencing tolerant donors and few sensitive lines used in mapping populations will considerably speed the process of QTL cloning and marker development as outlined above.

References

Akbar, M., Yabuno, T., and Nakao, S. 1972. Breeding for saline-resistant varieties of rice I. Variability for salt tolerance among some rice varieties. *Japanese Journal of Breeding* 22(5):277–284.

Alam, R., Rahman, M. S., Seraj, Z. I., Thomson, M. J., Ismail, A. M., Tumimbang-Raiz, E., and Gregorio, G. B. 2011. Investigation of seedling-stage salinity tolerance QTLs using backcross lines derived from *Oryza sativa* L. Pokkali. *Plant Breeding* 130(4):430–437.

Alkan, C., Sajjadian, S., and Eichler, E. E. 2011. Limitations of next-generation genome sequence assembly. *Nature Methods* 8(1):61–65.

An, P., Inanaga, S., Shimizu, H., El-Sidding, K., Li, X. J., Zheng, Y. R., Hibino, T., Morita, S., and Sugimoto, Y. 2004. Ameliorating effect of calcium on primary root elongation of soybean under sodium stress. *Biologia* 59(13):129–135.

Badawi, G. H., Yamauchi, Y., Shimada, E., Sasaki, R., Kawano, N., Tanaka, K., and Tanaka, K. 2004. Enhanced tolerance to salt stress and water deficit by overexpressing superoxide dismutase in tobacco (*Nicotiana tabacum*) chloroplasts. *Plant Science* 166(4):919–928.

Bahaji, A., Mateu, I., Sanz, A., and Cornejo, M. J. 2002. Common and distinctive responses of rice seedlings to saline- and osmotically-generated stress. *Plant Growth Regulation* 38(1):83–94.

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science* 48:1649–1664.

Bernstein, N., Shoresh, M., Xu, Y., and Huang, B. R. 2010. Involvement of the plant antioxidative response in the differential growth sensitivity to salinity of leaves vs roots during cell development. *Free Radical Biology and Medicine* 49(7):1161–1171.

Bonilla, P., Dvorak, J., Mackill, D., Deal, K., and Gregorio, G. 2002. RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (*Oryza sativa* L.) using recombinant inbred lines. *The Philippine Agricultural Scientist* 85(1):68–76.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 18:630–634.

Brumos, J., Colmenero-Flores, J. M., Conesa, A., Izquierdo, P., Sanchez, G., Iglesias, D. J., Lopez-Climent, M. F., Gomez-Cadenas, A., and Talon, M. 2009.

Membrane transporters and carbon metabolism implicated in chloride homeostasis differentiate salt stress responses in tolerant and sensitive Citrus rootstocks. *Functional and Integrative Genomics* 9(3):293–309.

Byrt, C. S., Platten, J. D., Spielmeyer, W., James, R. A., Lagudah, E. S., Dennis, E. S., Tester, M., and Munns, R. 2007. HKT1;5-like cation transporters linked to Na^+ exclusion loci in wheat, *Nax2* and *Kna1*. *Plant Physiology* 143(4):1918–1928.

Castillo, E., Tuong, T. P., Ismail, A., and Inubushi, K. 2007. Rice response to salinity: comparative effects of osmotic and ionic stresses. *Plant Production Science* 10:159–170.

Collard, B. C. Y., and Mackill, D. J. 2008. Marker assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of Biological Sciences* 363 (1491):557–572.

Cotsaftis, O., Plett, D., Johnson, A. A. T., Walia, H., Wilson, C., Ismail, A. M., Close, T. J., Tester, M., and Baumann, U. 2011. Root-specific transcript profiling of contrasting rice genotypes in response to salinity stress. *Molecular Plant* 4(1):25–41.

Diedhiou, C. and Golldack, D. 2006. Salt-dependent regulation of chloride channel transcripts in rice. *Plant Science* 170(4):793–800.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5):e19379.

Famoso, A. N., Zhao, K., Clark, R. T., Tung, C.-W., Wright, M. H., Bustamante, C., Kochian, L. V., and McCouch, S. R. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genetics* 7(8):e1002221.

Flores, N. R. L., Ismail, A. M., and Dionisio-Sese, M. L. 2005. Gas exchange and leaf water status of contrasting rice (*Oryza sativa* L.) genotypes as affected by salt stress. *Philippine Agricultural Scientist* 88(1):40–48.

Flowers, T., Galal, H., and Bromham, L. 2010. Evolution of halophytes: multiple origins of salt tolerance in land plants. *Functional Plant Biology* 37(7):604–612.

Fukuda, A., Nakamura, A., Hara, N., Toki, S., and Tanaka, Y. 2011. Molecular and functional analyses of rice NHX-type Na^+/H^+ antiporter genes. *Planta* 233(1):175–188.

Fukuda, A., Nakamura, A., Tagiri, A., Tanaka, H., Miyao, A., Hirochika, H., and Tanaka, Y. 2004. Function, intracellular localization and the importance in salt tolerance of a vacuolar Na^+/H^+ antiporter from rice. *Plant and Cell Physiology* 45(2):146–159.

Garthwaite, A. J., von Bothmer, R., and Colmer, T. D. 2005. Salt tolerance in wild *Hordeum* species is associated with restricted entry of Na^+ and Cl^- into the shoots. *Journal of Experimental Botany* 56(419):2365–2378.

Ghosh, N., Adak, M. K., Ghosh, P. D., Gupta, S., Sen Gupta, D. N., and Mandal, C. 2011. Differential responses of two rice varieties to salt stress. *Plant Biotechnology Reports* 5(1):89–103.

Gong, J., Zheng, X., Du, B., Qian, Q., Chen, S., Zhu, L., and He, P. 2001. Comparative study of QTLs for agronomic traits of rice (*Oriza sativa* L.) between salt stress and nonstress environment. *Science in China Series C-Life Science* 44(1):73–82.

Hou, X., Xie, K., Yao, J., Qi, Z., and Xiong, L. 2009. A homolog of human ski-interacting protein in rice positively regulates cell viability and stress tolerance. *Proceedings of the National Academy of Sciences of the United States of America* 106(15):6410–6415.

Huang, S., Spielmeyer, W., Lagudah, E. S., James, R. A., Platten, J. D., Dennis, E. S., and Munns, R. 2006. A sodium transporter (*HKT7*) is a candidate for *Nax1*, a gene for salt tolerance in durum wheat. *Plant Physiology* 142(4):1718–1727.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E. S., Qian, Q., Zhang, Q.-F., Li, J., and Han, B. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* 42(11):961–967.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., Fan, D., Lu, Y., Weng, Q., Liu, K., Zhou, T., Jing, Y., Si, L., Dong, G., Huang, T., Lu, T., Feng, Q., Qian, Q., Li, J., and Han, B. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44(1): 32–39.

Ismail, A., Heuer, S., Thomson, M., and Wissuwa, M. 2007. Genetic and genomic approaches to develop rice germplasm for problem soils. *Plant Molecular Biology* 65(4):547–570.

Ismail, A., and Thomson, M. 2011. "Root Genomics". In *Molecular breeding of rice for problem soils*, ed. A.C. de Oliveira and R.K. Varshney, pp. 289–312. New York: Springer.

Khan, M. S. A., Hamid, A., Salahuddin, A. B. M., Quasem, A., and Karim, M. A. 1997. Effect of sodium chloride on growth, photosynthesis and mineral ions accumulation of different types of rice (*Oryza sativa* L.). *Journal of Agronomy and Crop Science* 179(3):149–161.

Lee, S. Y., Ahn, J. H., Cha, Y. S., Yun, D. W., Lee, M. C., Ko, J. C., Lee, K. S., and Eun, M. Y. 2007. Mapping QTLs related to salinity tolerance of rice at the young seedling stage. *Plant Breeding* 126:43–46.

Lin, H. X., Zhu, M. Z., Yano, M., Gao, J. P., Liang, Z. W., Su, W. A., Hu, X. H., Ren, Z. H., and Chao, D. Y. 2004. QTLs for Na^+ and K^+ uptake of the shoots and roots controlling rice salt tolerance. *Theoretical and Applied Genetics* 108(2):253–260.

Lu, Y. Z., Feng, Z., Bian, L. Y., Xie, H., and Liang, J. S. 2011. miR398 regulation in rice of the responses to

abiotic and biotic stresses depends on CSD1 and CSD2 expression. *Functional Plant Biology* 38(1):44–53.

Mackill, D. J., Ismail, A. M., Singh, U. S., Labios, R. V., and Paris, T. R. 2012. Development and rapid adoption of submergence-tolerant (Sub1) rice varieties. *Advances in Agronomy* 115:303–356.

McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., Zeller, G., Clark, R. M., Hoen, D. R., Bureau, T. E., Stokowski, R., Ballinger, D. G., Frazer, K. A., Cox, D. R., Padhukasahasram, B., Bustamante, C. D., Weigel, D., Mackill, D. J., Bruskiewich, R. M., Rätsch, G., Buell, C. R., Leung, H., and Leach, J. E. 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences of the United States of America* 106(30):12273–12278.

Moradi, F., and Ismail, A. M. 2007. Responses of photosynthesis, chlorophyll fluorescence and ROS-scavenging systems to salt stress during seedling and reproductive stages in rice. *Annals of Botany* 99(6):1161–1173.

Moradi, F., Ismail, A. M., Gregorio, G., and Egdane, J. 2003. Salinity tolerance of rice during reproductive development and association with tolerance at seedling stage. *Indian Journal of Plant Physiology* 8:105–116.

Møller, I. S., Gilliham, M., Jha, D., Mayo, G. M., Roy, S. J., Coates, J. C., Haseloff, J., and Tester, M. 2009. Shoot Na^+ exclusion and increased salinity tolerance engineered by cell type-specific alteration of Na^+ transport in *Arabidopsis*. *Plant Cell* 21(7):2163–2178.

Munns, R., Schachtman, D. P., and Condon, A. G. 1995. The significance of a two-phase growth response to salinity in wheat and barley. *Australian Journal of Plant Physiology* 22(4):561–569.

Nemati, I., Moradi, F., Gholizadeh, S., Esmaeili, M. A., and Bihamta, M. R. 2011. The effect of salinity stress on ions and soluble sugars distribution in leaves, leaf sheaths and roots of rice (*Oryza sativa* L.) seedlings. *Plant Soil and Environment* 57(1):26–33.

Netondo, G. W., Onyango, J. C., and Beck, E. 2004. Sorghum and salinity: I. Response of growth, water relations, and ion accumulation to NaCl salinity. *Crop Science* 44(3):797–805.

Peng, S., and Ismail, A. 2004. Physiology and biotechnology integration for plant breeding. In *Physiological basis of yield and environmental adaptation in rice*, ed. H. T. Nguyen and A. Blum, pp. 83–140. New York: Marcel Dekker Inc.

Platten, J., Cotsaftis, O., Berthomieu, P., Bohnert, H., Davenport, R., Fairbairn, D., Horie, T., Leigh, R., Lin, H.-X., Luan, S., Maser, P., Pantoja, O., Rodríguez-Navarro, A., Schachtman, D., Schroeder, J., Sentenac, H., Uozumi, N., Very, A., Zhu, J.-K., Dennis, E., and Tester, M. 2006. Nomenclature for *HKT* transporters, key determinants of plant salinity tolerance. *Trends in Plant Science* 11(8):372–374.

Prasad, S. R., Bagali, P. G., Hittalmani, S., and Shashidhar, H. E. 2000. Molecular mapping of quantitative trait loci associated with seedling tolerance to salt stress in rice (*Oryza sativa* L.). *Current Science* 78:162–164.

Ren, Z.-H., Gao, J.-P., Li, L.-G., Cai, X.-L., Huang, W., Chao, D.-Y., Zhu, M.-Z., Wang, Z.-Y., Luan, S., and Lin, H.-X. 2005. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nature Genetics* 37(10):1141–1146.

Rodriguez, A. A., Cordoba, A. R., Ortega, L., and Taleisnik, E. 2004. Decreased reactive oxygen species concentration in the elongation zone contributes to the reduction in maize leaf growth under salinity. *Journal of Experimental Botany* 55:1383–1390.

Sabouri, H., Rezai, A., Moumeni, A., Kavousi, A., Katouzi, M., and Sabouri, A. 2009. QTLs mapping of physiological traits related to salt tolerance in young rice seedlings. *Biologia Plantarum* 53(4):657–662.

Septiningsih, E. M., Pamplona, A. M., Sanchez, D. L., Neeraja, C. N., Vergara, G. V., Heuer, S., Ismail, A. M., and Mackill, D. J. 2009. Development of submergence tolerant rice cultivars: the Sub1 locus and beyond. *Annals of Botany* 103:151–160.

Shabala, S., Demidchik, V., Shabala, L., Cuin, T. A., Smith, S. J., Miller, A. J., Davies, J. M., and Newman, I. A. 2006. Extracellular Ca^{2+} ameliorates NaCl-induced K^+ loss from *Arabidopsis* root and leaf cells by controlling plasma membrane K^+-permeable channels. *Plant Physiology* 141(4):1653–1665.

Shi, H., Ishitani, M., Kim, C., and Zhu, J.-K. 2000. The *Arabidopsis thaliana* salt tolerance gene *SOS1* encodes a putative Na^+/H^+ antiporter. *Proceedings of the National Academy of Sciences of the United States of America* 97(12):6896–6901.

Shi, H., Quintero, F. J., Pardo, J. M., and Zhu, J.-K. 2002. The putative plasma membrane Na^+/H^+ antiporter *SOS1* controls long-distance Na^+ transport in plants. *Plant Cell* 14(2):465–477.

Sunkar, R., Kapoor, A., and Zhu, J.-K. 2006. Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell* 18(8):2051–2065.

Takehisa, H., Shimodate, T., Fukuta, Y., Ueda, T., Yano, M., Yamaya, T., Kameya, T., and Sato, T. 2004. Identification of quantitative trait loci for plant growth of rice in paddy field flooded with salt water. *Field Crops Research* 89(1):85–95.

Taleisnik, E., Rodriguez, A. A., Bustos, D., Erdei, L., Ortega, L., and Senn, M. E. 2009. Leaf expansion in grasses under salt stress. *Journal of Plant Physiology* 166(11):1123–1140.

Tavakkoli, E., Fatehi, F., Coventry, S., Rengasamy, P., and McDonald, G. K. 2011. Additive effects of Na^+ and Cl^- ions on barley growth under salinity stress. *Journal of Experimental Botany* 62(6):2189–2203.

Tavakkoli, E., Rengasamy, P., and McDonald, G. K. 2010. High concentrations of Na^+ and Cl^- ions in soil solution have simultaneous detrimental effects on growth of

faba bean under salinity stress. *Journal of Experimental Botany* 61(15):4449–4459.

Teakle, N. L., Flowers, T. J., Real, D., and Colmer, T. D. 2007. *Lotus tenuis* tolerates the interactive effects of salinity and waterlogging by 'excluding' Na^+ and Cl^- from the xylem. *Journal of Experimental Botany* 58(8):2169–2180.

Thomson, M. J., de Ocampo, M., Egdane, J., Rahman, M. A., Sajise, A. G., Adorada, D. L., Tumimbang-Raiz, E., Blumwald, E., Seraj, Z. I., Singh, R. K., Gregorio, G. B., and Ismail, A. M. 2010. Characterizing the *Saltol* quantitative trait locus for salinity tolerance in rice. *Rice* 3(2-3):148–160.

Thomson, M. J., Zhao, K., Wright, M., McNally, K. L., Rey, J., Tung, C. W., Reynolds, A., Scheffler, B., Eizenga, G., McClung, A., Kim, H., Ismail, A. M., de Ocampo, M., Mojica, C., Reveche, M. Y., Dilla, C. J., Mauleon, R., Leung, H., Bustamante, C., and McCouch, S. R. 2012. High-throughput SNP genotyping for breeding applications in rice using the BeadXpress platform. Mol. Breed. 29:875–886.

Tobe, K., Zhang, L. P., and Omasa, K. 2003. Alleviatory effects of calcium on the toxicity of sodium, potassium and magnesium chlorides to seed germination in three non-halophytes. *Seed Science Research* 13(1): 47–54.

Walia, H., Wilson, C., Condamine, P., Liu, X., Ismail, A. M., Zeng, L., Wanamaker, S. I., Mandal, J., Xu, J., Cui, X., and Close, T. J. 2005. Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage. *Plant Physiology* 139(2):822–835.

Walia, H., Wilson, C., Zeng, L., Ismail, A., Condamine, P., and Close, T. 2007. Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage. *Plant Molecular Biology* 63(5):609–623.

Yamane, K., Mitsuya, S., Kawasaki, M., Taniguchi, M., and Miyake, H. 2009. Antioxidant capacity and damages caused by salinity stress in apical and basal regions of rice leaf. *Plant Production Science* 12(3):319–326.

Yeo, A. R., Caporn, S. J. M., and Flowers, T. J. 1985. The effect of salinity upon photosynthesis in rice (*Oryza sativa* L.): gas exchange by individual leaves in relation to their salt content. *Journal of Experimental Botany* 36(8):1240–1248.

Young, N.D., and Tanksley, S. D. 1989. RFLP analysis of the size of chromosomal segments retained around the tm-2 locus of tomato during backcross breeding. *Theoretical and Applied Genetics* 77:353–359.

Yeo, A. R., Yeo, M. E., Flowers, S. A., and Flowers, T. J. 1990. Screening of rice (*Oryza sativa* L.) genotypes for physiological characters contributing to salinity resistance, and their relationship to overall performance. *Theoretical and Applied Genetics* 79(3):377–384.

Zang, J., Sun, Y., Wang, Y., Yang, J., Li, F., Zhou, Y., Zhu, L., Jessica, R., Mohammadhosein, F., Xu, J., and Li, Z. 2008. Dissection of genetic overlap of salt tolerance QTLs at the seedling and tillering stages using backcross introgression lines in rice. *Science in China Series C: Life Sciences* 51(7):583–591.

Zhang, Z-H., Liu, Q., Song, H-X., Rong, X-M., and Ismail, A. M. 2011. Responses of contrasting rice (*Oryza sativa*) genotypes to salt stress as affected by nutrient concentrations. *Agricultural Sciences in China* 10:195–206.

Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., Norton, G. J., Islam, M. R., Reynolds, A., Mezey, J., McClung, A. M., Bustamante, C. D., and McCouch, S. R. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications* 2:467.

Zhu, J.-K. 2003. Regulation of ion homeostasis under salt stress. *Current Opinion in Plant Biology* 6(5):441–445.

Zhu, J.-K., Liu, J., and Xiong, L. 1998. Genetic analysis of salt tolerance in *Arabidopsis*: evidence for a critical role of potassium nutrition. *Plant Cell* 10(7):1181–1192.

Chapter 4

Marker-Assisted Introgression of Major QTLs for Grain Yield Under Drought in Rice

Arvind Kumar, Shalabh Dixit, and Amelia Henry

Abstract

Developing rice for drought tolerance has been one of the greatest challenges in the field of rice breeding. Several selection criteria based on morpho-physiological traits have been used in the past to identify drought-tolerant rice genotypes, quantitative trait loci (QTL), and genes. However, the success of these methods, in terms of adopting them into conventional and molecular breeding programs, has been limited. Recent studies have shown the effectiveness of screening rice genotypes for grain yield under drought stress for the development of drought-tolerant varieties. It has been observed that screening rice lines in the field under both managed drought stress and non-stress conditions provides an opportunity for combining high-yield potential with drought tolerance. The success of these screening strategies has led to the development and release of several drought-tolerant lines with high yield under non-stress conditions. Genotyping strategies such as bulk segregant analysis (BSA) and selective genotyping (SG) have been used to handle large populations derived from multiple donors and recipients. Several large-effect QTLs for increased grain yield under drought stress with consistent performance across multiple genetic backgrounds have been detected through the combination of screening for grain yield under drought and these genotyping strategies. The drought-yield QTLs have shown promising results in terms of their effect under a wide range of environments and are being successfully utilized in marker-assisted selection (MAS) programs. Physiological analysis of near-isogenic lines with these QTLs has revealed novel drought-related traits such as increased lateral roots and decreased xylem vessel diameter. Detection and pyramiding of such large-effect QTLs with genes for tolerance to other biotic stresses in the background of widely adapted cultivars can be an efficient strategy for increasing productivity of rice in rain-fed areas.

Introduction

Drought Tolerance: Complex or Simple?

Drought tolerance is defined as the ability of a plant to live, grow, and reproduce satisfactorily with limited water supply or under periodic conditions of water deficit (Turner 1979). Crop plants do not need to simply survive under drought, but should also possess the ability to

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

produce a harvestable yield. Drought tolerance is a quantitative trait long considered a complex trait because of its phenotypic and genetic control by multiple components (McWilliam 1989). The mechanisms behind drought tolerance in the context of grain yield have been considered complex by many physiologists as well as breeders, since it is combinations of different plant traits that are thought to determine the overall ability of plants to produce good yield under drought. Recently, drought tolerance has been described to be as simple a trait as any quantitative trait – provided the severity and intensities of drought stress are clearly classified and the ecosystem is well defined. The label of 'complex' has been attributed to the large number (1000+) of genes differentially regulated under dehydration stress in gene expression studies (Blum 2011). When drought tolerance is measured in terms of yield under drought stress, grain yield is found to be associated with many large- and small-effect QTLs, leading to the conclusion that drought tolerance is complex (Blum 2011). On the other hand, there are very few reports of successful improvement to grain yield under drought by improving a single physiological trait. Therefore, although the concept of drought tolerance may be considered either simple or complex, the prospects of improving drought tolerance have clearly been regarded as challenging.

Rice and Drought

Rice is the staple food for more than half of the world's population and is grown in more than 110 countries around the world. However, more than 90% of global rice is grown and consumed in Asia (Asia Rice Foundation 2005). In 2011, estimates showed a production of around 690 million tons of rice from a planted area of about 160 million hectares (FAO 2011). About 40% of the cultivated rice area is classified as 'rain-fed,' that is, totally dependent on rainfall as a source of water supply, with no or very little irrigation infrastructure available. Around 23 million ha of rice area in Asia is considered to be drought-prone (Pandey and Bhandari 2008) because of insufficient or sporadic rainfall. Decline in total rainfall as well as uneven distribution of rainfall during the growing season reduce rice yields in rain-fed areas. Furthermore, the diverse topography of the rain-fed ecosystem that ranges from the upper topo-sequence to lower topo-sequence, and includes upland and lowland ecosystems, exposing plants to variable severity of drought stress. Improving rice yields under drought must therefore overcome a range of different types of drought stress in terms of both severity and timing.

The Current Drought-Tolerance Improvement Strategy at the International Rice Research Institute (IRRI)

Direct Selection for Grain Yield is Achieved with Proper Drought Treatments

Many of the initial efforts to improve grain yield under drought focused on improving secondary traits such as root architecture, leaf water potential, panicle water potential, osmotic adjustment, and relative water content (Fukai et al. 1999; Price and Courtois 1999; Jongdee et al. 2002; Pantuwan et al. 2002; Toorchi et al. 2003). Earlier attempts to improve grain yield through improvement of physiological traits were based on the findings of several studies in which low heritability for grain yield under drought stress was reported (Rosielle and Hamblin 1981; Blum 1988; Edmeades et al. 1989). With the increased ability to appropriately classify drought stress, the ability to do better standardized phenotyping, and the availability of advanced statistical designs, many recent studies (Atlin and Lafitte 2002; Bernier et al. 2007; Kumar et al. 2008; Vikram et al. 2011) have shown that broad-sense heritability (H) for grain yield can be as high as that of its component traits, which are often not highly correlated with grain yield. Recently, in many of the mapping studies following direct selection for grain yield, the heritability of grain yield under drought was comparable with the heritability of grain yield under irrigated

situations under both upland (Bernier et al. 2007) and lowland conditions (Venuprasad et al. 2009; Vikram et al. 2011). In rice, mapping and breeding studies are increasingly using direct selection for grain yield as a criterion for identifying major QTLs for grain yield under drought and to improve grain yield under drought through conventional breeding approaches (Verulkar et al. 2010; Mandal et al. 2010) or in marker-assisted breeding (IRRI unpublished) targeted to improving the grain yield of popular high-yielding but drought-susceptible varieties IR64, Swarna, and MTU1010. In Brazil and Thailand also, direct selection for yield has been successfully applied in breeding for drought tolerance in upland and rain-fed lowland conditions (Ouk et al. 2006).

Identification of Large-Effect QTLs for Grain Yield under Drought

The basic requirements of QTL mapping for high grain yield under drought are the same as in any other mapping study. However, the complexity of genetic control for grain yield under drought demands highly controlled field experiments with uniform stress treatments. Similarly, highly controlled non-stress control experiments are needed for the assessment of yield potential, plant type, grain type, flowering, and so forth. This kind of experimentation helps in the identification of suitable lines from the mapping population combining high yield, drought tolerance, and good plant type for marker-assisted backcross programs.

Development of Mapping Populations: Identification of Donors and Recipient Variety

The objective of any QTL identification program should be the ultimate use of the identified QTLs for improving grain yield under drought through marker-assisted introgression of the identified QTLs into an identified improved variety. However, most QTL identification studies on grain yield as well as secondary traits have focused more on developing mapping populations in very low yielding, highly drought susceptible backgrounds that enable the study to show high values of QTL effects under drought. For example, some studies targeting the identification of QTLs in upland drought-prone ecosystems have used lowland varieties as recipients and screened the population in upland drought situations. The low adaptability of the lowland recipient variety resulted in large effects of the QTLs under upland conditions. However, a desirable QTL allele discovered in non-elite genetic material and showing a large effect may not offer any improvement in the target genetic background because the allele may already be ubiquitous in current varieties (Collins et al. 2008). The lack of repeatability of QTL effects across different populations (QTL × genetic background) and across environments (QTL × environmental interaction) has been another factor limiting the use of QTLs in molecular breeding (Price et al. 2002; Courtois et al. 2003, Lafitte et al. 2004; Bernier et al. 2008), and this necessitates that donor and recipient varieties be selected with appropriate consideration. The recipient variety for QTL studies should be an improved, high-yielding, and popular variety in the drought-prone environment. The selection of the drought-tolerant donor should depend upon the duration of the recipient variety. Most of the drought-tolerant donors are of early duration and are likely to reduce the duration of the recipient variety. Due consideration for growth duration, in addition to drought tolerance, as well as resistance to insects and disease, should be given when selecting a donor. Overall, complementation of the recipient and donor for different abiotic, biotic, and grain quality traits will enable the improvement of lines for more than one trait while improving yield under drought stress.

A wide range of population structures can be used for QTL mapping, the minimal requirement being the establishment of linkage disequilibrium between defined genotypes (Peterson 2002). Plants that tolerate inbreeding can accommodate especially great latitude for different population structures (Peterson 2002). a feature that makes self-pollinated species such as rice highly suitable for QTL mapping.

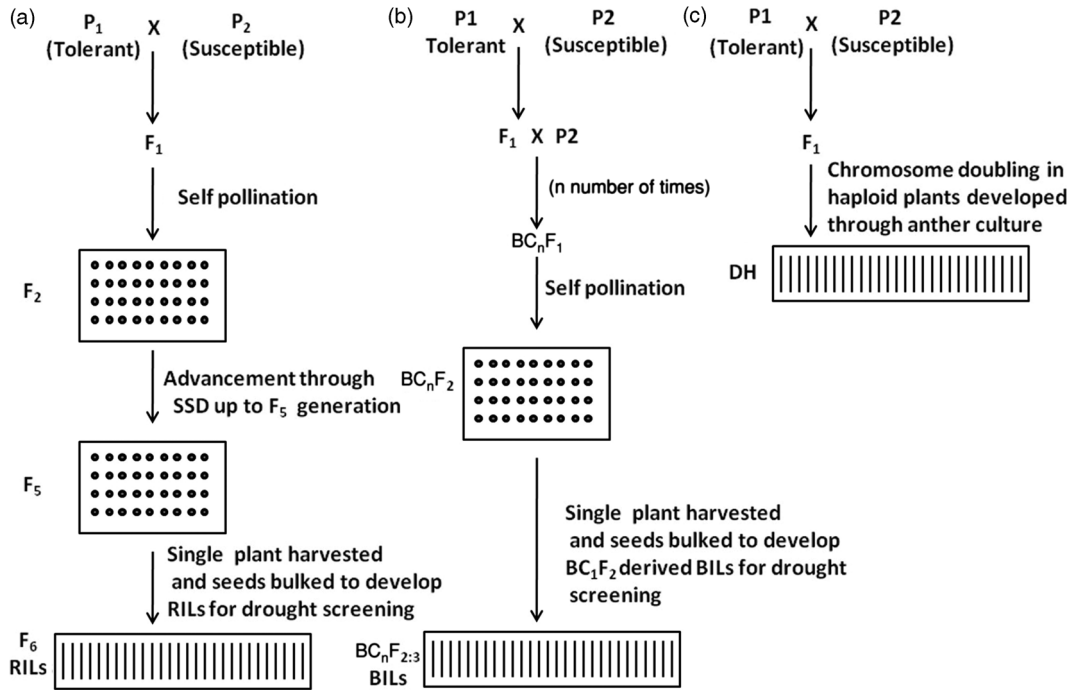


Fig. 4.1. Development strategies for different types of mapping populations in self-pollinated species (a) Recombinant inbred lines (RILs), (b) Backcross inbred lines (BILs), (c) Doubled haploids (DH).

Moreover, the relative ease of maintaining the purity of the lines, resulting from a high percentage of self-pollination in the subsequent generations after crossing, also makes population development easier. The complexity of genetic control of grain yield under drought in rice requires the development of large populations for the mapping of QTLs. A population size of more than 300 lines can be considered suitable for the mapping of these QTLs. A variety of populations can be used for mapping QTLs for high grain yield under drought. Some of them are described in the sections below.

Recombinant inbred lines (RILs) are developed by crossing two parents that contrast for the trait of interest, followed by subsequent selfing and advancement through the single seed descent (SSD) method to achieve nearly homozygous lines (Figure 4.1a). Backcross inbred lines (BILs) are another form of mapping populations, which are widely used for QTL mapping. F_1s developed by crossing two parents contrasting for the target trait are backcrossed (*n* number of times) to one of the parents to develop a BC_nF_1 population (Figure 4.1b). Most often, the parent used for backcrossing is the susceptible parent into which the QTL has to be finally introgressed. These $BCnF_1$ plants are then selfed through the SSD method to a BC_nF_3 generation and then bulked to develop $BC_nF_{3:4}$ lines that can be used for drought screening and the identification of QTLs. These populations are especially suitable for introgression of exotic germplasm from wild species to domestic varieties (Peterson 2002). Although most of the donors of drought QTLs are traditional varieties or landraces possessing valuable genes for drought tolerance, their yield potential is usually low under irrigated conditions. BIL populations are highly suitable for such cases. Apart from this, the relatively higher percentage of recipient genome in the background leads to a more convenient transfer of QTLs to the recipient. Advanced backcross populations are also suitable for fine-mapping purposes when large near-isogenic line (NIL) populations are needed with different segments of the

target QTL and similar genetic backgrounds. An advanced backcross QTL (AB-QTL) approach provides the opportunity to simultaneously identify and introgress QTLs in the recurrent parent, and saves time involved in varietal development (Tanksley et al. 1996). In doubled haploid (DH) lines, an attempt is made to combine the advantages of homozygosity with the speed at which an early-generation population can be made (Peterson 2002). DH populations can be produced by regenerating plants via the induction of chromosome doubling from pollen grains (Figure 4.1c); however, the production of DH populations is possible only in species that are amenable to tissue culture (Collard et al. 2005). DH populations have been used previously for the identification of QTLs for reproductive- and seedling-stage drought tolerance in rice (Lanceras et al. 2004; Xu et al. 2011).

Marker Genotyping of Mapping Populations

Although uniform drought phenotyping is the most important requirement of any QTL identification program (the IRRI drought-screening phenotyping protocol has been described previously, e.g., Venuprasad et al. 2008; Serraj et al. 2011), the appropriate use of molecular tools has become another important aspect of QTL mapping. The challenge of identifying a gene or QTL within a plant genome is like finding the proverbial needle in a haystack (Collard et al. 2005). However, genetic markers can be used to develop a systematic linkage map of the chromosome, dividing the chromosome into smaller sections in which it is easier to search for a putative QTL. Rice microsatellite (SSR) markers provide a suitable base for constructing a genetic linkage map. The ease of polymerase chain reaction (PCR) amplification and electrophoresis makes SSR markers suitable for quick and easy genotyping of large mapping populations. Moreover, the abundance of these markers across the rice genome and their codominant nature make them suitable for genotyping a variety of mapping populations with elaborate coverage of the genome. Earlier, whole-genome scanning (WGS) was used as a genotyping strategy in many studies (Figure 4.2a). WGS, although costly, provides an opportunity for identifying major- as well as minor-effect QTLs and at the same time identifying the interaction between different loci. Bulk segregant analysis (BSA) is a DNA pooling technique that was first proposed by Michelmore and colleagues (1991), in which markers linked to disease resistance genes were identified. It involves pooling of the DNA of phenotypic extremes used for developing high and low bulks that are genotyped along with the parents with all polymorphic markers (Figure 4.2b). The markers with bulk bands corresponding clearly to the parents are considered to be candidates, and a full population is genotyped with these markers for the identification of QTLs. Apart from the fact that this strategy is highly cost-effective and timesaving, it also eliminates the possibility of identifying any small-effect QTLs. The only drawback of this strategy is that it concentrates on just one segment of the genome and hence the amount of information it provides is limited. This technique has been used to identify large-effect QTLs for grain yield under drought (Venuprasad et al. 2009; Vikram et al. 2011). Venuprasad and colleagues (2009) used a subset of 4% of the total lines to constitute the bulks, which led to the identification of $qDTY_{2.1}$ and $qDTY_{3.1}$, two large-effect QTLs for grain yield under lowland drought. Another trait-based genotypic analysis called "selective genotyping" (SG) was suggested by Lebowitz and colleagues (1987). In this approach, 10-15% of the lines from the phenotypic extremes are selected for genotyping (Figure 4.2c). This leads to the generation of genotypic data for a subset of the whole population, which in turn can be used for developing a linkage map and in mapping and interaction studies. The markers found to be significant in the subset of the population can be used to genotype the whole population for a more precise estimation of the QTL effect. This strategy is cost-effective, timesaving, and provides information similar to that provided by WGS. One of the largest-effect QTLs known for grain yield under drought, $qDTY_{12.1,}$

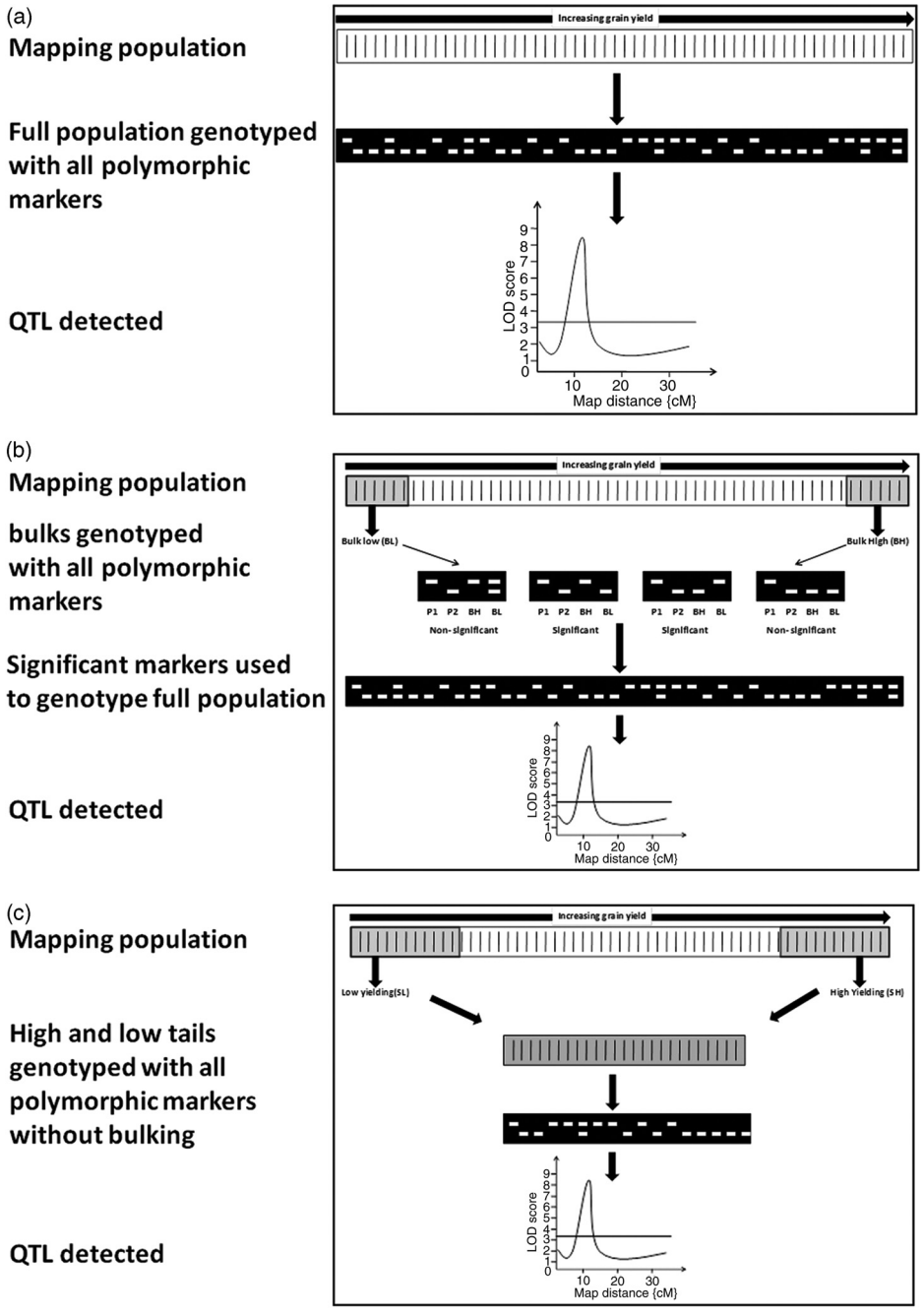


Fig. 4.2. Genotyping strategies for mapping QTLs for high grain yield under drought. (a) Whole-genome scan, (b) Bulk segregant analysis, (c) Selective genotyping.

was identified through this strategy, in which 12% of the phenotypic extremes were used for SG (Bernier et al. 2007). Navabi and colleagues (2009) were able to detect this QTL by genotyping as few as 20 selected lines (4.5% of the population). The success of both BSA and SG depends on the precision of phenotyping and the identification strategy of phenotypic extremes. These methods can reliably detect large-effect QTLs with minimum genotyping and thus allow screening of larger numbers of mapping populations, identifying useful QTLs that are effective across genetic backgrounds, or multiple QTLs from different donors that are effective in the same genetic background.

Major Rice QTLs Reported for Grain Yield under Drought

The mapping strategy described above has enabled the identification of a number of large-effect QTLs affecting grain yield under drought in both upland and lowland ecosystems. Table 4.1 presents a summary of such QTLs reported in rice. The first reported large-effect QTL for grain yield under drought was *qDTY*$_{12.1}$ (Bernier et al. 2007). This QTL was identified in a population of 436 random F_3-derived lines from a cross between the upland rice cultivars Vandana and Way Rarem. Located between RM28048 and RM28166, this QTL explained an R^2 of 33% under severe upland drought conditions. Apart from this, the locus showed its effect on many traits that affect grain yield under drought, such as days to flowering (DTF), plant height, biomass, harvest index (HI), drought-response index (DRI), and panicle number m^{-2}.

Two large-effect QTLs affecting grain yield under lowland drought, *qDTY*$_{2.1}$ and *qDTY*$_{3.1}$, were identified in a BIL population derived from a cross of the high-yielding lowland rice variety Swarna and the upland rice variety Apo. Both QTLs showed a very high effect under severe lowland drought (R^2 values 16.3% and 30.7%). The effect of both these QTLs was also seen on other traits such as DTF and plant height. BSA was successfully used for the first time to identify such large-effect QTLs for grain yield under drought (Venuprasad et al. 2009). Another large-effect QTL for grain yield under favorable aerobic and irrigated lowland conditions, *qDTY*$_{6.1}$, was identified in this population (Venuprasad et al. 2012a). This QTL explained an R^2 value under upland and lowland non-stress conditions of up to 66% and 39%, respectively.

A series of experiments also was also begun on F_3-derived populations developed from the cross of drought-tolerant donor N22 with high-yielding mega-varieties Swarna, IR64, and MTU1010, that resulted in the identification of *qDTY*$_{1.1}$, a large-effect QTL having an effect on grain yield under severe lowland drought across these three populations. This QTL showed an R^2 value of 13.4%, 16.9%, and 12.6% across two seasons of screening under severe lowland drought in N22/Swarna, N22/IR64, and N22/MTU1010 populations, respectively (Vikram et al. 2011). QTLs for grain yield under drought at this locus have also been reported in other populations derived from crosses of CT9993-5-10-1-M/IR62266-42-6-2 and Apo/IR64 (Kumar et al. 2007; Venuprasad et al. 2012b).

QTL x Environment and QTL x Genotype Interactions

For marker-assisted selection (MAS) to be worthwhile, it is important that the identified QTL show large and consistent effects under varying environmental conditions and across a wide range of genetic backgrounds (Bernier et al. 2009; Vikram et al. 2011). The high specificity of the QTLs toward environmental conditions has been one reason that, out of a large number of QTLs identified for drought tolerance, only a few could be used in MAS. This situation becomes even more challenging if QTLs are to be identified for complex traits such as grain yield under drought. A variety of factors apart from drought, such as soil, nutrients, solar radiation, biotic stresses, and so forth, affect

Table 4.1. Large effect QTLs reported for high rice yield under severe reproductive stage drought in upland and lowland ecosystems.

						Significance			R^2 (%)		
QTL	Population	Ecosystem/Stress	Chr.	Marker/Interval	Position (cM)	LOD	F value	P value	Phenotypic	Genetic	Reported by
qDTY$_{1.1}$	N22/Swarna	Lowland/Stress	1	RM11943- RM12091			40.8		13.4		Vikram et al., 2011
	N22/IR64	Lowland/Stress	1	RM11943- RM12091			57.6		16.9		Vikram et al., 2011
	N22/MTU1010	Lowland/Stress	1	RM11943- RM12091			40.0		12.6		Vikram et al., 2011
	Apo/IR64	Upland/Stress	1	RM472	162.9	9.26				54.0	Venuprasad et al., 2012b
	CT9993-5-10-1-M/ IR62266-42-6-2	Lowland/Stress	1	EM11_11- RG109	206.5	4.4				32.0	Kumar et al., 2007
qDTY$_{2.1}$	Apo/2∗Swarna	Lowland/Stress	2	RM324				0.01		16.4	Venuprasad et al., 2009
qDTY$_{3.1}$	Apo/2∗Swarna	Lowland/Stress	3	RM416				0.01		30.7	Venuprasad et al., 2009
qDTY$_{3.2}$	N22/Swarna	Lowland/Stress	3	RM60- RM22			15.1		18.5		Vikram et al., 2011
qDTY$_{6.1}$	Apo/2∗Swarna	Upland/Non- stress	6	RM510	11.3			<0.0001		66.0	Vikram et al., 2009
	Apo/2∗Swarna	Lowland /Non stress	6	RM510	11.3			<0.0001		39.0	Venuprasad et al., 2012a
	Vandana/IR72	Upland/Stress	6	RM19367	9.2			<0.0001		40.0	Venuprasad et al., 2012a
	Apo/IR72	Upland/Stress	6	RM510	11.3			<0.0001		63.0	Venuprasad et al., 2012a
qDTY$_{10.1}$	N22/MTU1010	Lowland/Stress	10	RM216-RM304			24.2		5.0		Vikram et al., 2011
qDTY$_{12.1}$	Vandana/Way Rarem	Upland/Stress	12	RM28048-RM28166	49.0	34.0			33.0	51.0	Bernier et al., 2007

yield per se under field conditions. It is therefore very important that the identified QTL be robust enough to show an effect under a variety of environmental conditions as well as under varying drought intensities. One way to overcome this challenge could be to screen the mapping population for high grain yield in the target environment under naturally occurring drought stress and to use these data for QTL identification. However, performing such large-scale experiments in the target environment could be expensive, and the chances of failure can be high because of the chance of rainfall during the critical stress period. Another alternative could be to identify QTLs under managed stress conditions and test a random subset of the mapping population under naturally occurring drought in multi-location trials. This strategy enables widespread testing of the QTL under varying environmental conditions and water regimes. Bernier and colleagues (2009) tested a random set of RILs in 21 experiments conducted at IRRI and in eastern India. These experiments confirmed the effect of $qDTY_{12.1}$ under varied situations and further showed that the relative effect of the QTL increased with increasing severity of stress. Under well-watered conditions, $qDTY_{12.1}$ showed no effect but in the most severe stress treatments, it showed an additive effect of more than 40% of the mean trial yield. The second important aspect of drought-yield QTLs has been their effect across elite genetic backgrounds. Among a large number of QTLs identified for drought tolerance in rice, only a few could show their effect across genetic backgrounds. Genotyping strategies such as BSA also make it possible to screen a large number of mapping populations simultaneously for the presence of a QTL affecting grain yield in more than one background. The QTL $qDTY_{1.1}$ – contributed by the donor N22 and functional across the backgrounds of the mega-varieties MTU1010, IR64, and Swarna (Vikram et al. 2011) – was identified using BSA for two of the three populations, while WGS was carried out for the third population. Swamy and colleagues (2011) identified 14 meta-QTLs among 53 yield QTLs reported in 15 studies. This study also tested a panel of random drought-tolerant donors for the identified drought yield QTLs and reported the presence of $qDTY_{12.1}$ in 85% of the lines, followed by $qDTY_{4.1}$ in 79%, and $qDTY_{1.1}$ in 64% of the lines.

Effect of Drought Yield QTLs on Multiple Yield-Related Traits under Drought

It is interesting to see that, within or near the regions showing an effect on grain yield under drought, regions affecting plant height, biomass, HI, and DTF under drought were also identified. In some cases, the same regions are likely to show an effect on yield under drought and DTF or plant height. This may be either a result of a close linkage between regions affecting two traits or a pleiotropic effect of the QTLs for grain yield under drought on DTF and plant height. For example, $qDTY_{12.1}$, a large-effect QTL for grain yield under upland drought, showed its effect on almost all traits, such as DTF, plant height, biomass, HI, DRI, and panicle number m^{-2}. Similarly, $qDTY_{3.1}$ showed an effect on DTF along with grain yield under severe lowland drought (Venuprasad et al. 2009). Another recently reported QTL, $qDTY_{1.1}$, was related to DTF, plant height, HI, and biomass across N22/Swarna and N22/MTU1010 populations, and was related to DTF, plant height, and biomass in N22/IR64 populations. Swamy and colleagues (2011) compared the synteny of a meta-QTL coinciding with this region with other crops. It was observed that this region was syntenic to a region on chromosome 3 in maize near marker *msu2*, on chromosome 4B in wheat near marker *Rht b1*, and in barley on chromosome 6H near marker *Bmac0316*. In the same study, another meta-QTL coinciding with $qDTY_{3.1}$ was also found in maize on chromosome 1 near marker *Umac107a*. The study also reported the association of all these markers to grain yield under drought in their respective crops (Swamy

et al. 2011). These studies not only suggest the presence of conserved genes across drought yield QTLs but also the presence of conserved regions affecting yield or stress tolerance across different species of grass families.

Candidate Gene Content and Comparative Genomics of Drought Yield QTLs

In the study of 53 yield QTLs reported on rice, Swamy and colleagues (2011) identified 14 meta-QTLs and compiled a list of candidate genes in these regions using annotated gene information available in rice databases. Most of the genes present in these regions were genes for hypothetical and expressed proteins, pseudo genes, genes for signal transduction, and transposable elements. However, a large number of annotated genes and gene families that were common across these meta-QTLs were also identified. Stress-inducible genes, growth- and development-related genes, and sugar transport-related genes were frequently observed in these regions (Swamy et al. 2011). Most of the meta-QTLs reported in this study either coincide with or are located very near the large-effect drought yield QTLs identified in various populations, and hence these common gene families could be candidates for high yield under drought. The study also reported the presence of leucine rich receptor (LRR) kinase, leucine zipper, cell-division controlling proteins, sugar transport protein-like genes, no apical meristem (NAM), pentatricopeptide repeat proteins, cytokinin oxidase, F-box proteins, AP2 domain-containing proteins, and zinc-finger transcription factors in 6 of the 14 identified meta-QTLs. Although most of these genes are already known to affect yield, yield-related traits, or stress responsiveness in a variety of crops, they are likely affecting drought response in these QTL regions (Swamy et al. 2011). Similarly, differential expression of 4,5-DOPA dioxygenase extradiol, glycosyltransferase, amino acid transporters, MADS-box (MCM-1, Agamous, Deficiens, SRF) family genes, and serine/threonine protein kinase within the $qDTY_{1.1}$ region between N22 and IR64 have recently been reported (Lenka et al. 2011; Vikram et al. 2011).

Physiology Studies to Characterize the Mechanisms by Which Major-Effect QTLs Confer Improved Yield under Drought

For physiological characterization, field and greenhouse studies are conducted using +QTL and −QTL lines, as well as the donor and recipient parents. The number of entries used in physiology experiments is small: typically three +QTL lines and three −QTL lines. Field or rainout shelter experiments are planted in plots of three or four 3-m rows, with four replicates per genotype. The strategy is to start with high-throughput/broad-scale measurements, such as canopy temperature, normalized difference vegetation index (NDVI), biomass, and harvest index. At IRRI, a semiautomated instrument rack that rolls along the rails of the rainout shelter allows for frequent monitoring of canopy temperature and NDVI (Figure 4.3). The initial results from these high-throughput and broad-scale measurements guide the direction of subsequent, more detailed measurements. For example, differences in canopy temperature under drought are verified with stomatal conductance (porometer) measurements. Then, since transpiration results from water uptake by roots,

Fig. 4.3. A semi-automated instrument rack for monitoring canopy temperature and NDVI of +QTL and –QTL lines under drought in the rainout shelter at IRRI. For a color version of this figure, please refer to the color plate.

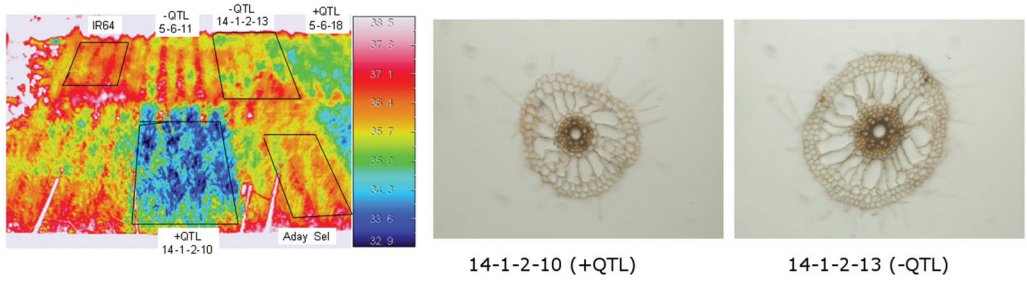


Fig. 4.4. The IR64 x Aday Sel NILs showed large differences in canopy temperature under severe drought and smaller root and xylem vessel diameters. For a color version of this figure, please refer to the color plate.

soil is sampled to measure root length density by depth and root length within diameter classes. If no differences in canopy temperature under drought are observed in root architecture between +QTL lines and −QTL lines, root anatomy and function (e.g., hydraulic conductance) are then investigated. If the predominant differences between +QTL and −QTL lines under drought are NDVI and biomass, subsequent measurements focus on biomass partitioning and carbohydrate remobilization during grain filling.

Thus far, physiological mechanisms behind +QTL lines from two separate populations have been characterized at IRRI. Interestingly, the traits identified are different from the most prevalent traits that have been used for selection for drought tolerance; deep root growth and large-diameter roots. The QTL $qDTY_{12.1}$ was identified from a Vandana x Way Rarem cross for upland systems (Bernier et al. 2007), and it shows an increasing effect with increasing severity of drought stress. To compare +QTL and −QTL lines, a lysimeter study was conducted to monitor water uptake by pot weight (Bernier et al. 2009). Lines with $qDTY_{12.1}$ showed a small but significant (7%) increase in water uptake. This QTL was subsequently characterized in rainout shelter conditions, where +QTL lines showed significantly lower canopy temperature under drought, which was associated with a greater proportion of fine (lateral) roots at shallow soil depths. The increased proportion of lateral roots may allow +QTL lines to perform better under drought through better access to increasingly shrinking columns of water in the soil, and perhaps improved nutrient uptake.

The QTLs $qDTY_{2.2}$ and $qDTY_{4.1}$ were identified from an IR64 x Aday Sel population to confer greater yield under drought (Swamy et al. 2013). The +QTL lines from this population also show an increased effect when drought stress becomes more severe. Large differences in canopy temperature and stomatal conductance were observed between +QTL and −QTL lines under severe drought stress, but not under moderate or mild drought stress, and no differences in root architecture were observed between +QTL and −QTL lines (Swamy et al. 2013). A greenhouse study was then conducted to examine root hydraulic conductance and anatomy, from which +QTL lines stood out as having significantly lower root hydraulic conductance and smaller xylem and apical root diameter (Figure 4.4). It is uncertain exactly how such altered root anatomy contributes to yield under drought. More measurements on these lines, including xylem vessel cavitation and nighttime water flux from roots, are necessary.

Perspectives

Novel Marker-Assisted Breeding Approaches

The availability of large-effect QTLs affecting grain yield under drought has opened the way for the large-scale development of drought-tolerant versions of susceptible high-yielding varieties through marker-assisted breeding. Novel approaches to marker-assisted breeding

have helped to develop products through the pyramiding of large-effect QTLs for grain yield under drought. Efforts are also being made to pyramid these QTLs along with QTLs affecting other biotic and abiotic stresses, for the development of NILs with multiple stress tolerance.

Marker-Assisted Backcrossing

The identification and transfer of large-effect QTLs through marker-assisted backcrossing (MAB) is suggested as a rapid approach for developing tolerant versions of popular rice varieties (Bernier et al. 2007). The use of molecular markers permits the genetic dissection of the progeny in each generation and increases the speed of the selection process, thus increasing genetic gain per unit time (Tanksley et al. 1989; Hospital 2003). It also provides an efficient foreground selection for the target locus, background recovery for the recurrent parent genome, minimization of linkage drag surrounding the locus being introgressed, and rapid breeding of new genotypes with favorable traits (Neeraja et al. 2007). It has also been reported that screening the populations with a few well-placed markers (two to four markers on a chromosome of 100 cM) provides adequate genome coverage as observed in previous backcross programs (Visscher et al. 1996: Servin and Hospital 2002). Such targeted introgression leads to the development of uniform lines that are similar to the recipient parent, with improved drought tolerance. In the case of drought, it is important that the lines developed at each stage of the backcross program be tested under drought. This not only provides additional data bearing on the effect of QTLs in the backcross generation but also allows the identification of the best NILs for further backcrossing. With the few high-yielding varieties such as IR64, IR36, MTU1010, Swarna, BR11, and TDK 1 being highly popular among farmers in drought-prone rain-fed areas, MAB introgression of the major drought-yield QTLs appears to be a preferred strategy for improving such popular varieties. MAB also provides better opportunities for keeping grain quality traits of the varieties intact, which is essential for retaining favor with farmers. NILs developed with such high background recovery can be disseminated and adopted quickly because of their similarity to the popular recipient parent. Introgression of the large-effect QTL *qDTY*$_{12.1}$ into recipient parent Vandana was undertaken at the IRRI. Large BC_2F_2 and BC_3F_2 populations segregating for the QTL were screened with foreground and background markers, and NILs with high background recovery were identified. The highest yielding NILs showed a more than 4-fold average advantage over the recipient Vandana under six upland drought experiments with varying degrees of stress. Under severe drought stress, *qDTY*$_{12.1}$-introgressed Vandana lines showed a yield advantage of 0.5 t ha^{-1}.

Pyramiding of Drought QTLs

Each of the individual drought-yield QTLs shows a yield advantage of 300-500 kg ha^{-1} under moderate to severe drought conditions. This yield advantage may look significant from an academic point of view, but it is not high enough to attract the attention of farmers and motivate them to replace their currently cultivated varieties with newly developed varieties with a single drought yield QTL. In order to provide farmers with economic yield advantages of 1.0 t ha^{-1} or more, it is necessary for two to three or more drought yield QTLs to be pyramided together. It is suggested that pyramiding of more than one QTL from the same or different donors working in the same genetic background could be a good strategy for increasing yield under drought to the maximum possible extent. Recently experiments have proved that a combination of two to three drought-yield QTLs has given a higher advantage than a single QTL in IR64 (data not presented). IR64 lines with two and three pyramided QTLs showed a yield advantage of 1.2–1.5 t ha^{-1} over IR64, under moderate to severe drought conditions, while maintaining similar yield potential

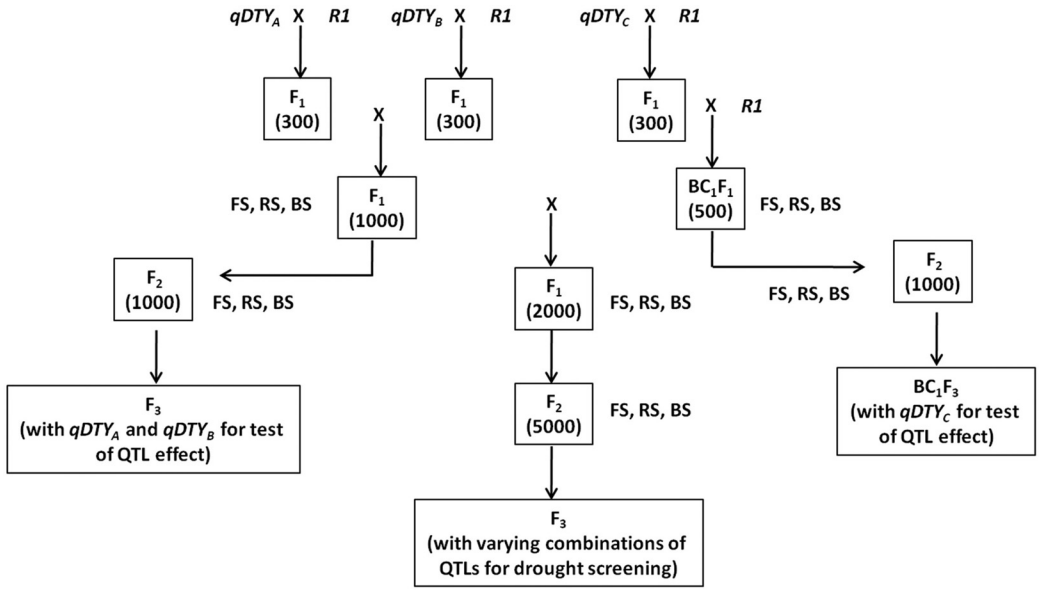


Fig. 4.5. Marker-assisted backcrossing strategy for pyramiding three large-effect QTLs for drought tolerance in rice. FS- foreground selection, RS- recombinant selection, BS- background selection.

under normal irrigated conditions (IRRI, unpublished). Because of this initial success, pyramiding of three large-effect QTLs (*qDTY*$_{1.1}$, *qDTY*$_{2.1}$ and *qDTY*$_{3.1}$) in the popular variety Swarna, and pyramiding *qDTY*$_{1.1}$, *qDTY*$_{1.2}$, *qDTY*$_{2.2}$, and *qDTY*$_{12.1}$ in IR64 is under way. Lines with a QTL from the original mapping population are intercrossed to bring the QTLs together, followed by a series of backcrosses to recover the recipient parent background (Figure 4.5). Foreground and background selections are made in each generation of the backcross program. A large F_1 population is developed in each backcross generation, which also provides an opportunity to self and test the effect of QTLs and to identify any possible undesirable effects in early backcross generations.

Combining Drought QTLs with QTLs Affecting Other Biotic/Abiotic Stresses

Recent trends in climate change have increased the possibility of biotic and abiotic stresses in rice-growing areas (Wassmann et al. 2009). It is commonly observed that during the cropping season uneven rainfall patterns can cause flash floods, while in another stage of growth, crops can be affected by drought. These occurrences not only lead to direct yield losses but also increase the incidence of insects and disease. For example, the incidence of rice blast is seen to increase with drought. It therefore becomes important that genes/QTLs conferring tolerance for a wide range of biotic and abiotic stresses be combined with drought-tolerant QTLs. Efforts are now being made to combine *qDTY*$_{1.1}$, *qDTY*$_{2.1}$, and *qDTY*$_{3.1}$ with *SUB1* in the popular variety Swarna and in *qDTY*$_{1.1}$, *qDTY*$_{1.2}$, *qDTY*$_{2.2}$, and *qDTY*$_{12.1}$ in IR64. Similarly, efforts are being made to combine these QTLs with genes such as *Bph20* and *Bph21* for brown planthopper; *xa5, Xa13,* and *Xa21* for bacterial leaf blight; and *Pi1* and *Pi2* for blast tolerance (Figure 4.6).

Marker-Assisted Recurrent Selection and Genome-Wide Selection

Molecular markers have been used in breeding programs by means of identifying markers that

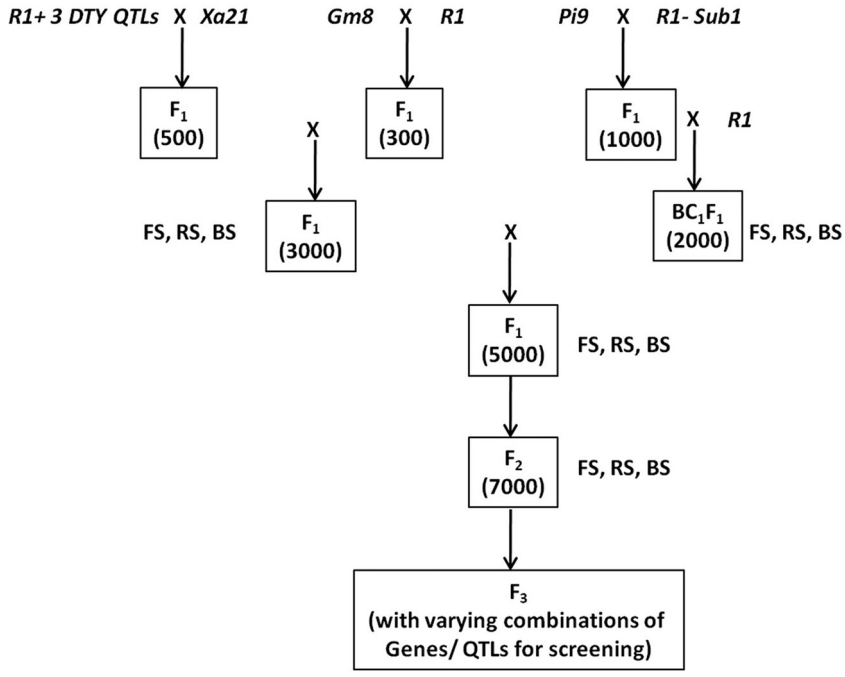


Fig. 4.6. Marker-assisted backcrossing strategy for large-scale pyramiding of drought yield QTLs with genes/QTLs for tolerance of biotic stresses such as bacterial leaf blight (*Xa21*), gall midge (*Gm8*), blast (*Pi9*), and submergence (*SUB1*). FS- foreground selection, RS- recombinant selection, BS- background selection.

are linked to QTLs affecting the trait of interest and using them to trace the presence of QTLs in a different generation of a backcross program leading to NILs with QTLs affecting the trait of interest. Recent advances in high-throughput marker technology have led to a significant increase in the number of markers available in rice. Single nucleotide polymorphisms (SNPs) in particular provide for a very elaborate coverage of the rice genome. Moreover, high-throughput facilities for SNP genotyping have markedly increased the efficiency and reduced the cost of genotyping. These facilities can now be exploited for rapid population improvement programs for a diverse set of traits. Two approaches that enable the use of markers for population improvement are marker-assisted recurrent selection (MARS) and genomic selection (GS). MARS refers to the improvement of an F_2 population by one cycle of marker-assisted selection (i.e., based on phenotypic data and marker scores), followed by three cycles of marker-based selection that is based on marker scores only (Bernardo and Charcosset 2006). The marker scores are usually determined from 20 to 35 markers that show significant association with one or more traits of interest in a multiple-regression model (Edwards and Johnson 1994; Koebner 2003). In contrast, GS refers to marker-based selection without significance testing and without identifying a subset of markers associated with the trait (Meuwissen et al. 2001). In this technique, the lines are genotyped with markers spread evenly across the genome. The effects on the quantitative trait (i.e., breeding values) of all markers are fitted as random effects in a linear model. Trait values are then predicted as the sum of an individual's breeding values across all marker loci, and selection is subsequently based on these genome-wide predictions (Bernardo and Yu 2007). A high

correlation between predicted and true breeding values has been reported (Meuwissen et al. 2001). Both these techniques can play an important role in the rapid development of breeding populations rich in alleles for a wide range of biotic and abiotic stresses. Such populations also form a strong base for a study of the interaction of alleles across the genome for complex quantitative traits such as grain yield under drought.

Collaborative Strategies of Breeding and Physiology for Improvement of Drought Tolerance in Rice

Successful breeding for drought tolerance requires knowing the precise mechanisms of drought tolerance operational in the donor and the combining ability of the donor. Some donors, such as Moroberekan, do not possess good combining ability and have not been used extensively in breeding programs or for mapping studies. On the other hand, donors such as N22 and Aday Sel have been used extensively in mapping and candidate gene studies and have been the source of major QTLs for grain yield under drought. A clear-cut understanding of the physiological mechanisms of drought tolerance of different QTLs provides opportunities for combining QTLs that contribute to the increase in grain yield under drought through different mechanisms that either add to or complement effects of each trait. For example, a QTL that enhances water uptake, if combined with QTLs that increase yield through efficient translocation of resources from shoots to panicles during grain filling or QTLs related to transpiration efficiency, will result in the development of an efficient drought-tolerant plant type. QTL combinations pyramiding complementary traits (such as a QTL conferring improved nutrient uptake with a QTL conferring improved water uptake) may confer a synergistic interaction resulting in yield advantages beyond the sum of the effects of the two individual QTLs. Further, QTLs with effects at different stages of rice growth – seedling, vegetative, or reproductive stage – would provide the opportunity to combine such QTLs and develop rice lines tolerant of drought at all stages.

The most direct role of physiology in breeding for drought tolerance in rice has traditionally been to recommend traits for selection (Courtois and Lafitte 1999). While the identified traits such as root growth at depth were determined based on many empirical studies relating genetic variation for root growth at depth with physiological response to drought, the approach of selecting for specific traits has relied on several assumptions. The trait selection approach depends on scientists' interpretation of plant function and drought response, the prediction that a similar response will be observed when the underlying QTL is introgressed into a susceptible variety, and the assumption that the trait and resulting improvement in drought response (i.e., leaf water status) will be tied to improved yield under drought. The limitation of this strategy is that if there is some aspect of the response of the trait/QTL to environmental or genetic factors that is not fully understood, the trait/QTL might not be effective when introgressed into drought-susceptible backgrounds.

However, recent advances in plant physiology, particularly in root biology, have identified several traits showing direct correlations between their expression and yield via nutrient and water uptake (Lynch 2007), making them strong candidates for selection for drought improvement. One recent case of a successful introgression of a root QTL for improved yield under drought in rice is that of the QTL *Dro1* (Uga et al. 2011), which causes deeper root growth angle, resulting in greater yield under drought. Another case is rice ARB lines, which were developed with successive selections for increased maximum root depth (Shashidhar 2008) and have shown promising drought responses in terms of yield (Verulkar et al. 2010). Therefore, although direct selection for yield is currently the most effective approach for improving rice yield under drought, selection for other traits may also become more effective

as our understanding of plant response to stress increases.

Identifying drought tolerance traits using major-effect drought yield QTL lines presents a unique opportunity to explore the traits that are affecting plant response to drought, with confidence that the differences in traits between +QTL and −QTL lines have a high probability of being responsible for differences in yield. Working with material that is known to have improved yield under drought and being able to compare genetically similar germplasm (+QTL and −QTL lines) facilitates more rapid pinpointing of the traits behind the QTLs. Furthermore, this approach allows us to be open-minded in our understanding of the mechanisms behind drought tolerance; selecting for yield under drought, as opposed to selecting for specific traits, is confirming that many mechanisms – some unexpected – can contribute to grain yield under drought.

Summary

Recent progress in improving rice yield under drought through direct selection for yield is benefiting from the efficiency and high-throughput capacity of molecular strategies. Advancement of generations while simultaneously applying drought stress and the use of large (>300 lines) populations have been key components of this progress. Major-effect drought-yield QTLs that are effective in multiple backgrounds across a wide range of environments have been identified. Some of these QTLs are also seen to be contributed from multiple drought-tolerance donor genotypes. The drought response of physiological traits associated with the drought-yield QTLs has agreed with the yield response to increasing drought stress severity and has pointed to some novel drought traits. Marker-assisted backcrossing and the development of BIL and NIL populations of the drought yield QTLs in multiple susceptible backgrounds is an efficient way of improving the drought tolerance of local varieties whose grain and plant types are popular in respective drought-prone rice-growing regions. Pyramiding of multiple DTY QTLs, together with pathogen-resistance genes, is the next step for meeting the goal of appreciable (>1 t ha^{-1}) yield advantages under drought stress that could be adopted by farmers.

References

Asia Rice Foundation. 2005. http://www.asiarice.org/sections/learnrice/riceislife.html.

Atlin G.N., Lafitte H.R., 2002. Marker-assisted breeding versus direct selection for drought tolerance in rice. In: Saxena N.P., O'Toole J.C. (Eds.), *Field Screening for Drought Tolerance in Crop Plants with Emphasis on Rice. Proceedings of an International Workshop on Field Screening for Drought Tolerance in Rice*, 11–14 Dec 2000, ICRISAT, Patancheru, India, p. 208.

Bernardo R., Charcosset A. 2006. Usefulness of gene information in marker-assisted recurrent selection: A simulation appraisal. *Crop Science* 46:614–621.

Bernardo R., Yu J. 2007. Prospects of genome wide selection for quantitative traits in maize. *Crop Science* 47:1082–1090.

Bernier J., Atlin G.N., Serraj R., Kumar A., Spaner D. 2008. Breeding upland rice for drought resistance. *Journal of the Science of Food and Agriculture* 88:927–939.

Bernier J., Kumar A., Venuprasad R., Spaner D., Atlin G.N. 2007. A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Science* 47:507–516.

Bernier J., Serraj R., Kumar A., Venuprasad R., Impa S., Gowda R.P.V., Oane R., Spaner D., Atlin G. 2009. The large-effect drought-resistance QTL qtl12.1 increases water uptake in upland rice. *Field Crops Research* 110:139–146.

Blum A. 1988. *Plant Breeding for Stress Environments*. CRC Press, Boca Raton, FL.

Blum A. 2011. Drought resistance – is it really a complex trait? *Functional Plant Biology* 38:753–757.

Collard B.C.Y., Jahufer M.Z.Z., Brouwer J.B., Pang E.C.K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142:169–196.

Collins N.C., Tardieu F., Tuberosa R. 2008. Quantitative trait loci and crop performance under abiotic stress: Where do we stand? *Plant Physiology* 147:469–486.

Courtois B., Shen L., Petalcorin, Carandang S., Mauleon R., Li Z.K. 2003. Locating QTLs controlling constitutive root traits in the rice population IAC 165 x Co39. *Euphytica* 134:335–345.

Courtois B., Lafitte R. 1999. Improving rice for drought-prone upland environments. In: Ito O., O'Toole J., Hardy B. (Eds.) *Genetic Improvement of Rice for Water-limited*

Environments. International Rice Research Institute, Los Baños, Philippines.

Edmeades G.O., Bolanos J., Lafitte H.R., Rajaram S., Pfeiffer W., Fischer R.A. 1989.Traditional approaches to breeding for drought resistance in cereals. In: Baker F.W.G. (Ed.), *Drought Resistance in Cereals*. CAB International, Wallingford, Oxon, UK, pp. 27–52.

Edwards M., Johnson L. 1994. RFLPs for rapid recurrent selection. In: Analysis of Molecular Marker Data. *Joint Plant Breeding Symposia Series*. ASA, Madison, WI, pp. 33–40

FAO. 2011. FAOstat. www.faostat.fao.org

Fukai S., Pantuwan G., Jongdee B., Cooper M. 1999. Screening for drought resistance in rainfed lowland rice. *Field Crops Research* 64:61–74.

Hospital F. 2003. Marker-assisted breeding. In: Newbury H.J. (Ed.) *Plant Molecular Breeding*, Blackwell Publishing, Oxford, pp. 30–59.

Jongdee B., Fukai S., Cooper M. 2002. Leaf water potential and osmotic adjustment as physiological traits to improve drought tolerance in rice. *Field Crops Research* 76:153–163.

Koebner R. 2003. MAS in cereals: Green for maize, amber for rice, still red for wheat and barley. In: *Marker Assisted Selection: A Fast Track to Increase Genetic Gain in Plant and Animal Breeding*? Turin, Italy. 17–18 Oct. 2003. Available at http://www.fao.org/biotech/docs/Koebner.pdf (verified 15 Feb. 2007). FAO, Rome.

Kumar A., Bernier J., Verulkar S., Lafitte H.R., Atlin G.N. 2008. Breeding for drought tolerance: Direct selection for yield, response to selection and use of drought-tolerant donors in upland and lowland-adapted populations. *Field Crops Research* 107:221–231.

Kumar R., Venuprasad R., Atlin G.N. 2007. Genetic analysis of rainfed lowland rice drought tolerance under naturally-occurring stress in eastern India: Heritability and QTL effects. *Field Crops Research* 103:42–52.

Lafitte H.R., Price A.H., Courtois B. 2004. Yield response to water deficit in an upland rice mapping population: Associations among traits and genetic markers. *Theoretical and Applied Genetics* 109:1237–1246.

Lanceras J.C., Pantuwan G., Jongdee B., Toojinda T. 2004. Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiology* 135:384–399.

Lebowitz R.J., Soller M., Beckmann J.S. 1987. Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theoretical and Applied Genetics* 73:556–562.

Lenka S.K., Katiyar A., Chinnusamy V., Bansal K.C. 2011. Comparative analysis of drought-responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. *Plant Biotech Journal* 9:315–327.

Lynch J.P. 2007. Roots of the second green revolution. *Australian Journal of Botany* 55:493–512.

Mandal N.P., Sinha P.K., Variar M., Shukla V.D., Perraju P., Mehta A., Pathak A.R., Dwivedi J.L., Rathi S.P.S., Bhandarkar S., Singh B.N., Singh D.N., Panda S., Mishra N.C., Singh Y.V., Pandya R., Singh M.K., Sanger R.B.S., Bhatt J.C., Sharma R.K., Raman A., Kumar A., Atlin G.N. 2010. Implications of genotype x input interactions in breeding superior genotypes for favorable and unfavorable rainfed upland environments. *Field Crops Research* 118:135–144.

McWilliam J.R. 1989. The dimensions of drought. In: *Drought Resistance in Cereals* Baker F.W.G. (Ed.), CAB International, Wallingford, Oxon, UK, pp. 1–11

Meuwissen T.H.E., Hayes B.J., Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.

Michelmore R.W., Paran I., Kesseli R.V. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences USA* 88:9828–9832.

Navabi A., Mather D.E., Bernier J., Spaner D.M., Atlin G.N. 2009. QTL detection with bidirectional and unidirectional selective genotyping: Marker-based and trait-based analyses. *Theoretical and Applied Genetics* 118:347–358.

Neeraja C.N., Maghirang-Rodriguez R., Pamplona A., Heuer S., Collard B.C.Y., Septiningsih E.M., Vergara G., Sanchez D., Xu K., Ismail A.M., Mackill D.J. 2007. A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theoretical and Applied Genetics* 115:767–776.

Ouk M., Basnayake J., Tsubo M., Fukai S., Fischer K.S., Cooper M., Nesbitt H. 2006. Use of drought response index for identification of drought tolerant genotypes in rainfed lowland rice. *Field Crops Research* 99: 48–58.

Pandey S., Bhandari H. 2008. Drought: Economic costs and research implications. In: Serraj R., Bennett J., Hardy B. (Eds.) *Drought Frontiers in Rice: Crop Improvement for Increased Rainfed Production*. Singapore: World Scientific Publishing and Los Baños (Philippines): International Rice Research Institute.

Pantuwan G., Fukai S., Cooper M., Rajatasereekul S., O'Toole J.C. 2002. Yield response of rice (*Oryza sativa* L.) to drought under rainfed lowlands. 3. Plant factors contributing to drought resistance. *Field Crops Research* 73:181–200.

Peterson A.H. 2002. What has QTL mapping taught us about plant domestication? *New Phytologist* 154:591–608.

Price A.H., Cairns J.E., Horton P., Jones R.G.W., Griffiths H. 2002. Linking drought-resistance mechanisms to drought avoidance in upland rice during a QTL approach: Progress and new opportunities to integrate stomatal and mesophyll responses. *Journal of Experimental Botany* 53:989–1004.

Price A., Courtois B. 1999. Mapping QTLs associated with drought resistance in rice: Progress, problems, and prospects. *Plant Growth Regulation* 29:123–133.

Rosielle A.A., Hamblin J. 1981. Theoretical aspects of selection for yield in stress and non-stress environments. *Crop Science* 21:943–946.

Serraj R., McNally K.L., Slamet-Loedin I., Kohli A., Haefele S.M., Atlin G., Kumar A. 2011. Drought resistance improvement in rice: An integrated genetic and resource management strategy. *Plant Production Science* 14:1–14.

Servin B., Hospital F. 2002. Optimal positioning of markers to control genetic background in marker-assisted backcrossing. *Jounal of Heredity* 93:214–217.

Shahshidhar, H.E. 2008. Aerobic rice – an efficient water management strategy for rice production. In: Aswathanarayana U. (Ed.) Food and Water Security. Taylor and Francis, London, U.K.

Swamy B.P.M., Ahmed H.U., Henry A., Mauleon R., Dixit S., Vikram P., Ram T., Verulkar S.B., Perraju P., Mandal N.P., Variar M., Robin S., Chandrababu R., Singh O.N., Dwivedi J.L., Das S.P., Mishra K.K., Yadaw R.B., Aditya T.L., Karmakar B., Satoh K., Moumeni A., Kikuchi S., Leung H., Kumar A. 2013. Genetic, physiological, and gene expression analyses reveal that multiple QTL enhance yield of rice mega-variety IR64 under drought. *PLoS One* 8(5): e62795.doi:10.1371/journal.pone.0062795.

Swamy B.P.M., Vikram P., Dixit S., Ahmed H.U., Kumar A. 2011. Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. *BMC Genomics* 12:319.

Tanksley S.D., Grandillo S., Fulton T.M., Zamir D., Eshed Y., Petiard V., Lopez J., Beck-Bunn T. 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and Applied Genetics* 92:213–224.

Tanksley S.D., Young N.D., Paterson A.H., Bonierbale M.W. 1989. RFLP mapping in plant breeding: New tools for an old science. *Biotechnology* 7:257–264.

Toorchi M., Shashidhar H.E., Gireesha T.M., Hittalmani S. 2003. Performance of backcross involving trangressant doubled haploid lines in rice under contrasting moisture regimes: Yield components and marker heterozygosity. *Crop Science* 43:1448–1456.

Turner N.C. 1979. Drought resistance and adaptation to water deficits in crop plants. In: Mussell H., Staples R.C. (Eds.) *Stress Physiology in Crop Plants*. Wiley, New York, pp. 181–194.

Uga Y., Okuno K., Yano M. 2011. *Dro1*, a major QTL involved in deep rooting of rice under upland field conditions. *Journal of Experimental Botany* 62:2485–2494.

Venuprasad R., Bool M.E., Quiatchon L., Atlin G.N. 2012a. A QTL for rice grain yield in aerobic environments with large effects in three genetic backgrounds. *Theoretical and Applied Genetics* 124:323–332.

Venuprasad R., Bool M.E., Quiatchon L., Sta Cruz M.T., Amante M., Atlin G.N. 2012b. A large effect QTL for rice grain yield under upland drought stress on chromosome 1. *Molecular Breeding* 30:535–547.

Venuprasad R., Dalid C.O., Del Valle M., Zhao D., Espiritu M., Sta Cruz M.T., Amante M., Kumar A., Atlin G.N. 2009. Identification and characterization of large-effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. *Theoretical and Applied Genetics* 120:177–190.

Venuprasad R., Impa S.M., Gowda R.P.V., Atlin G.N., Serraj R. 2011. Rice near-isogenic-lines (NILs) contrasting for grain yield under lowland drought stress. *Field Crops Research* 123:38–46.

Venuprasad R., Sta Cruz M.T., Amante M., Magbanua R., Kumar A., Atlin G.N. 2008. Response to two cycles of divergent selection for grain yield under drought stress in four rice breeding populations. *Field Crops Research* 107:232–244.

Verulkar S.B., Mandal N.P., Dwivedi J.L., Singh B.N., Sinha P.K., Dongre P., Singh O.N., Bose L.K., Swain P., Robin S., Chandrababu R., Senthil S., Jain A., Shashidhar H.E., Hittalmani S., Vera Cruz C., Paris T., Raman A., Haefele S., Serraj R., Atlin G., Kumar A. 2010. Breeding resilient and productive genotypes adapted to drought prone rainfed ecosystems of India. *Field Crops Research* 117:197–208.

Vikram P., Mallikarjuna Swamy B.P., Dixit S., Ahmed H.U., Sta Cruz M.T., Singh A.K., Kumar A. 2011. *qDTY1.1*, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. *BMC Genetics* 12:89.

Visscher P.M., Haley C.S., Thompson R. 1996. Marker-assisted introgression in backcross breeding programs. *Genetics* 144:1923–1932.

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

Xu Q., Yuan X.P., Yu H.Y., Wang Y.P., Tang S.X., Wei X.H. 2011. Mapping QTLs for drought tolerance at seedling stage in rice using doubled haploid population. *Rice Science* 18:23–28.

Chapter 5

Molecular Breeding for Phosphorus-efficient Rice

Sigrid Heuer, J.H. Chin, R. Gamuyao, S.M. Haefele, and M. Wissuwa

Abstract

Rice is the main staple food for more than half of the world's population and the main source of calories in most Asian and many African countries. Since many rice-dependent countries are poor, it is critically important to keep rice prices low and increase productivity to provide sufficient food for a growing population. However, a sustainable increase in rice production is possible only if nutrients removed with the harvest are replaced by application of either mineral fertilizers or manure. Since fertilizer costs are rising, depletion of soil nutrients is an increasing problem, especially in the developing world where most farmers do not have the resources to purchase sufficient fertilizer or do not have access to fertilizer. In addition, the majority of rainfed rice in Asia is produced on poor quality and problem soils that are often low in nutrients or have properties, such as low pH or high aluminum and iron, that render phosphorus (P) unavailable to plants. On those soils, very high fertilizer doses have to be applied to provide sufficient plant-available P. Given that currently known rock phosphate reserves, the source of P fertilizer, are limited it can be expected that P deficiency will aggravate and will increasingly limit productivity, especially in poor countries. One way to address this problem is to develop crops that are more efficient in acquiring P from the soil and applied fertilizer, or crops with higher internal P-use efficiency, that is, with higher biomass production per unit P.

In this paper, we provide a brief comprehensive overview on P-related aspects of rice production and highlight the potential of molecular breeding approaches to improve P-efficiency. As an example, we describe the major quantitative trait locus *Phosphorus uptake 1* (*Pup1*), which confers tolerance of P deficiency.

Introduction

Phosphorus (P) is an essential macro-element for all living cells and therefore crucially important for crop production. It has been estimated that insufficient plant-available P constrains plant growth on more than 5.7 billion ha of land worldwide (Batjes 1997). According to a more recent study, more than 50% of global agricultural land is deficient in P (Lynch 2011). A study on P

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

balances, based on figures of global P fertilizer/manure application and removal of P with crop harvest, suggests that negative P balances occur on 29% of the world's crop land with the most severe deficits in Eastern Europe, Africa, and Southeast Asia (MacDonald et al. 2011). A global assessment of soils with P-retention properties further showed that vast areas in South America, West and Central Africa, and South-East Asia have low plant available P as a result of unfavorable soil properties, such as low pH or high concentrations of aluminum and iron (Batjes, 2011). On those soils, P would be sufficient to support high crop yields for probable decades if it were not "locked up" in soil-P pools of very low plant availability. Even where plant-available soil P was sufficient initially, the continuous removal (mining) of soil P in harvested grain will eventually induce P deficiency unless P mining is balanced by sufficient P fertilizer application (Rose et al. 2010). The extent to which neutral P balances can be maintained in high-input agriculture or achieved in less intensive systems will, to a large extent, depend on the cost of P fertilizers. Given that high-quality phosphate rock, the source of P fertilizers, is a non-renewable and increasingly limited resource, and that production and transportation costs strongly depend on the price of oil, it is foreseeable that P-fertilizer prices will continue to increase (Cordell et al. 2009; Van Kauwenbergh 2010; Vance and Chiou 2011). Such price increases will place a considerable burden on many of the resource-poor farmers in developing countries. It has therefore been suggested to increase efforts to develop crop cultivars with enhanced P efficiency (Ismail et al. 2007; Wissuwa et al. 2009; Richardson et al. 2011).

Plant Responses to P Deficiency

Plants have developed adaptive mechanisms to access P beyond the rhizosphere when bioavailable P is limited. P-starved plants mobilize the Al-, Fe-, or Ca-bound P through secretion of organic acids, acid phosphatases, and ribonucleases, resulting in enhanced P uptake (Lambers et al. 2006; MacIntosh 2011). Modifications in root architecture and morphology facilitate exploration beyond the root-soil interface to forage for P. Root structural changes include increased root hair growth and length, bias toward lateral root growth over primary root growth, and, in the case of lupin plants, formation of cluster roots (Péret et al. 2011; Sato and Miura 2011). To further increase the soil area foraged and P acquisition, plants form symbiotic associations with mycorrhizal fungi in the root cortex, which is referred to as arbuscular-mycorrhizal (AM) symbiosis (for review, see Bucher 2007; Sawers et al. 2008).

Within the plant, internal P is remobilized from RNA molecules by intracellular ribonucleases (MacIntosh 2011) and from phospholipids in membranes that are then replaced with galacto- and sulfolipids (Nussaume et al. 2011). Under P-deficient conditions, intracellular ATP/ADP and P_i become low, whereas the level of pyrophosphate (PP_i) remains unchanged (Plaxton and Podestá 2006). Hence, stressed plants use bypass enzymes involved in PP_i-dependent metabolic reactions (Plaxton and Tran 2011).

P is taken up by plants as inorganic $H_2PO_4^-$ and HPO_4^{2-} phosphate ions wherein the former form is prevalent in soils with pH 4.5-5.0 (Raghothama 1999). P uptake through the root requires P transporters located in the plasma membrane. In the rice reference genome, thirteen P transporters (PT) have been identified, of which OsPT2 has been classified as a low-affinity transporter, whereas OsPT1 and OsPT6 are high-affinity transporters (Seo et al. 2008; Ai et al. 2009). Analysis of transgenic plants showed that OsPT8 regulates P homeostasis as overexpression, and silencing of this gene resulted in the expected increase and decrease, respectively, of P uptake, but plants showed aberrant P accumulation and plant growth (Jia et al. 2011). Expression of the P transporter OsPT11 is closely associated with AM symbiosis as its expression is specific to roots colonized with mycorrhiza (Paszkowski

et al. 2002). These functional analyses of rice P transporters clearly show the significant role of P transporters in P uptake and homeostasis.

P homeostasis and the associated complex regulatory network has begun to be understood in *Arabidopsis* and, for some of the identified genes, rice orthologs are known. One of the important genes for P homeostasis, the *Arabidopsis PHO2* gene or the rice ortholog *Leaf Tip Necrosis 1* (*LTN1*), encodes a member of the E2 ubiquitin conjugase family (Kraft et al. 2005; Aung et al. 2006; Bari et al. 2006; Hu et al. 2011). Mutation of *PHO2/LTN1* leads to the overaccumulation of P in leaves, a result of the deregulation of P transfer from roots to shoots via the phloem (Delhazie and Randall 1995; Dong et al. 1998; Hu et al. 2011). The expression of both genes is regulated by microRNA *miR399*/*OsmiR399* as shown by the down-regulation of *PHO2*/*LTN1* transcripts (Fujii et al. 2005; Bari et al. 2006; Chiou et al. 2006; Hu et al. 2011), and *miR399* overexpression plants phenocopy the *pho2* mutant (Aung et al. 2006). Regulation of *PHO2* by *miR399* is part of the systemic control in response to P starvation and maintenance of P homeostasis, as demonstrated by the long-distance signaling of mobile *miR399* from shoots to roots through the phloem (Lin et al. 2008; Lin and Chiou 2008; Pant et al. 2008). In turn, *miR399* activity is inhibited by *AtIPS1* and *At4,* which are non-protein-coding genes highly induced under P deficiency (Bari et al. 2006; Aung et al. 2006). Interestingly, it has been shown that *AtIPS1* contains a short sequence complementary to *miR399.* However, the presence of a mismatch at the miRNA cleavage site prevents cleavage of *AtIPS* and "sequesters" *miR399* (Franco-Zorrilla et al. 2007). Hence, *miR399* is a negative regulator of *PHO2* while *AtIPS1* acts as a positive regulator, thereby fine-tuning transfer of P from roots to shoots.

The orthologous genes *AtPHR1* and *OsPHR2* encode MYB-type transcription factors, and their overexpression resulted in increased P content, while the *phr1* mutant displayed altered P allocation between shoot and root (Rubio et al. 2001; Nilsson et al. 2007; Zhou et al. 2008). As part of the regulation of P-starvation responses, *AtPHR1* is post-translationally modified by *SIZ1*, a SUMO E3 ligase (Miura et al. 2005). It was further shown that, in *phr1* and *siz1* mutants, expression of four SPX-domain proteins was reduced (Duan et al. 2008). Moreover, *AtSPX1* positively regulates a set of *P-starvation-induced* (PSI) genes, while *AtSPX3* modulates expression of *AtSPX1* and PSI genes through negative feedback mechanisms (Duan et al. 2008). In rice, it was shown that *OsPHR2* expression was highly induced in *OsSPX1* RNAi plants as well as in *pho2* mutants and that the displayed phenotype was similar to that of *OsPHR2* overexpressing plants, suggesting that *OsSPX1* contributes to P homeostasis as a negative regulator (Wang et al. 2009).

Increased understanding of the components involved in plant P homeostasis has begun to reveal a regulatory network that appears to be more complex than previously thought. Undoubtedly the information has broadened our scientific understanding of P-response processes and highlights the importance and dynamic of P in plant growth and development. Unfortunately, the identified regulatory factors and pathways have not yet been systematically assessed for a possible role in stress tolerance, which might employ mechanisms distinct from P-starvation responses. In fact, a recent study in *Arabidopsis* showed suppression of P-starvation responses in tolerant genotypes (Rouached et al. 2011) and, similarly, P-response genes were not differentially expressed in tolerant rice (see below).

Phosphorus in Rice Cropping Systems

Every year about 2.1 million t of fertilizer phosphorus (P) is applied to rice at a cost exceeding US$11 billion (Rose et al. 2010). Within Asia, negative P balances have been reported mainly for Southeast Asian countries such as Burma, Cambodia, Indonesia, the Philippines, and

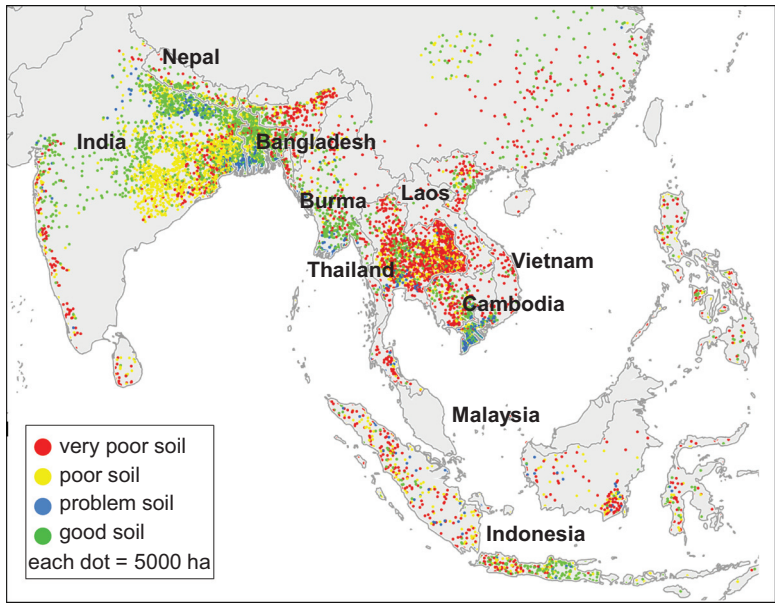


Fig. 5.1. Soil quality in rain-fed lowland rice systems. Soils in rain-fed (intermediate and shallow) rice systems in Asia are often constrained by abiotic stresses and nutrient deficiency (Haefele and Hijmans 2007; Haefele and Hijmans 2009). Note: Rain-fed upland rice areas are not included in this map. For a color version of this figure, please refer to the color plate.

Thailand, but also for parts of India (MacDonald et al. 2011; Rose et al. 2010). Soils with high P retention potential are widespread in Southeast Asia, as well as in Africa and South America (Batjes et al. 2011). Unless high doses of P fertilizer are applied or other measures, such as liming on acidic soils, are taken to reduce P retention, plant-available P is low on such soils, leading to P deficiency and reduced yield. In addition to P fixation, some of the major rice-growing regions coincide with areas of poor and problem soils (Figure 5.1; Table 5.1) where P deficiency is widespread but may not be the only factor limiting plant growth (Haefele and Hijmans 2007 and 2009; Kirk 2004).

Obviously, the degree of P deficit (or surplus) depends on the amount of P fertilizer applied per field; however, several additional factors related to soil properties and agronomic practices, most importantly water management, strongly affect plant P status. Generally, P deficiency is most common in rain-fed rice systems (Table 5.1) and least common in irrigated rice systems. The factors that affect P availability across the main rice systems can be summarized as follows:

I. Irrigated rice systems and favorable rain-fed lowlands:
   - Soil water saturation offers optimal conditions for P movement (diffusion) in soil and therefore typically provides more P for uptake by the root (Huguenin-Elie et al. 2009);
   - Alterations in soil chemistry after flooding enhance P solubility as a result of reductive processes and rhizosphere acidification (Huguenin-Elie et al. 2009; Kirk 2004);
   - A reliable water supply enhances not only yield but also yield predictability; this in turn leads farmers to increase investments in agrochemicals, such as P fertilizer. Irrigated and favorable lowland fields are therefore more likely to be well fertilized;
   - Higher yields, on the other hand, result in higher rates of P removal with harvested

Table 5.1. Soil quality of rice soils in rainfed lowlands of Asia

	Total area	Very poor soils		Poor soils		Problem soils		Good soils	
Country	'000 ha	'000 ha	% of total	'000 ha	% of total	'000 ha	% of total	'000 ha	% of total
India	16,139	2,187	14	6,541	41	1,218	8	6,193	38
Thailand	8,202	5,195	63	1,658	20	396	5	920	11
Bangladesh	5,093	539	11	558	11	627	12	3,388	67
Indonesia	4,006	1,604	40	1,023	26	275	7	1,108	28
Vietnam	2,911	764	26	327	11	869	30	945	32
Myanmar	2,411	577	24	234	10	354	15	1,246	52
China	1,746	716	41	262	15	45	3	716	41
Cambodia	1,573	748	48	192	12	146	9	483	31
Philippines	1,323	559	42	340	26	0	0	423	32
Nepal	760	209	27	253	33	0	0	299	39
Laos	438	342	78	31	7	18	4	48	11
South Korea	279	120	43	42	15	0	0	117	42
Sri Lanka	218	95	44	79	37	12	6	30	14
North Korea	115	3	3	23	20	2	2	86	75
Buthan	17	8	46	6	37	0	0	3	17
Timor-Leste	7	2	23	0	4	0	1	5	72
total	**45237**	**13,668**		**11,569**		**3,962**		**16,010**	

Data based on Haefele and Hijmans 2009.

grain. Maintenance P application rates are expected to be higher in irrigated rice, a measure aimed at balancing this depletion.

II. Rain-fed upland rice. In addition to the factors listed above, several upland-specific factors affect P availability, which consequently tends to be at the opposite (low) end of the P availability spectrum:
   - Typically, soils in uplands are older and more weathered than lowland soils, especially compared with alluvial soils in the vicinity of rivers. As a result, the soil concentration of P and other nutrients can be low;
   - Rates of P fixation by soil constituents can be extremely high as a result of weathering. With the tendency for P to be immobilized rapidly, only a small fraction (< 10-20%) of fertilizer P may be available to a crop in the year of P-fertilizer application;
   - The lack of water control in rain-fed upland systems increases the risk of droughts – drying of the topsoil where most of the P is concentrated thus temporarily reduces P availability further;
   - Since weeds cannot be controlled by flooding the field, weeds are a major competitor for resources in rain-fed systems, especially during crop establishment.

Breeding Targets Related to P Efficiency in Rice

Selection of modern rice cultivars has typically been performed under high-nutrient conditions with a primary focus on yield. Although some of these modern varieties may perform well under moderate P deficiency, it has been pointed out that this is most likely a result of their superior harvest index (HI), which allows them to turn a higher proportion of their total biomass into grain yield, and not due to specific tolerance mechanisms that increase P uptake or total biomass (Wissuwa et al. 2009). Selection for yield conducted in target environments, such as generally poor or P-deficient soils, may therefore offer advantages. However, conclusive data sets comparing the effect of selection under different P levels on subsequent yield performance of rice in different environments do not exist. Yet

the presence of large genotypic variation for P-deficiency tolerance in rice (Fageria et al. 1988; Wissuwa and Ae 2001a) suggests that breeding for P-deficiency tolerance in rice is feasible. That most of the tolerant genotypes identified were landraces or of traditional plant type further suggests that such selection has been practiced locally in the pre–green revolution era (Wissuwa et al. 2009). The breeding target that we follow in our own work is to combine the superior tolerance to P deficiency of older cultivars and landraces with the responsiveness to P fertilizer typically found in modern high yielding varieties. This responsiveness would mainly be a result of the higher yield potential of modern varieties, whereas better P acquisition or higher internal P utilization efficiency are expected to drive the higher tolerance of P deficiency in older cultivars.

Tolerance of P deficiency is a complex trait with a multitude of plant physiological and morphological adaptations that are of potential importance for increased P uptake or internal P-use efficiency (Lambers et al. 2006; Rose and Wissuwa 2012). Furthermore, P deficiency itself is not a very well-defined condition, as it may refer to a situation in which no P fertilizer is applied on poor soils with average yields of 1-2 t ha^{-1}, or to a situation in which additional P fertilizer application may further increase already high rice yields of 5-6 t ha^{-1}. In each case, breeding targets and traits for selection would likely differ and it may therefore be of value to briefly review the potential benefits of targeting P uptake or internal P use:

(i) Benefits of improving P uptake (P-acquisition efficiency, PAE) have typically been associated with tolerance of more severe P-deficiency as encountered in highly P-fixing soils where less than 1% of total soil P may be plant-available. P solubilization due to plant-induced rhizosphere modifications would enhance P availability. However, the same processes should allow crops to access a bigger portion of the fixed fertilizer P that has accumulated in fertilized soils over the years, and may therefore be a crucial factor allowing a reduction of maintenance-fertilizer rates without sacrificing yield.

(ii) Improving internal P-use efficiency (PUE) through breeding has been suggested but never really attempted, despite the obvious advantages of reducing the amount of plant P needed for producing equal biomass or grain yield (Rose and Wissuwa 2012). The benefits of enhancing PUE would theoretically improve yield across environments, irrespective of whether P was applied or not. However, little is known about whether different PUE mechanisms operate across the spectrum of P deficiency.

(iii) Recently, selection for reduced grain-P concentration has been suggested as an alternative PUE mechanism beneficial under P deficiency, and as a general way to reduce removal of P from fields at harvest (Rose et al. 2011). Since export of P from fields with harvested grain is the main driver of the P cycle in agriculture, cultivars with lower grain-P concentrations would reduce the need for continuously high P-fertilizer application in high-input systems, or would reduce P mining in low-input systems.

Given that only a fraction of total soil P is readily plant-available in any agricultural soil, it is not surprising that most approaches to improve crop yields under P deficiency have targeted P-acquisition efficiency. In following sections we describe the most comprehensive approach undertaken so far to develop rice with improved tolerance of P deficiency, and we describe in detail the identification and characterization of *Phosphorus uptake 1* (*Pup1*), a major quantitative trait locus (QTL) for P uptake in rice.

The *Pup1* QTL and Its Application in Molecular Breeding

Based on a field screening of thirty diverse rice varieties under P-deficient rain-fed conditions, the *aus*-type rice variety Kasalath was

identified as P-deficiency tolerant, based on its superior P uptake from a highly P-fixing soil (Figure 5.2a). Although Dular, likewise an *aus*-type variety, was more tolerant, Kasalath was chosen since bacterial artificial chromosome (BAC) libraries were available, which later facilitated sequencing of the *Pup1* locus (see below). The *japonica*-type variety Nipponbare was used as the intolerant parent for the development of a mapping population, and the *Pup1* region on Chr. 12 was shown to significantly enhance P uptake and plant performance under the applied screening conditions (Figure 5.2b, c; Wissuwa and Ae 2001b; Wissuwa et al. 2002). An intermediate-effect QTL at the tip of Chr. 6 was additionally identified, and it was shown that this region contains a cluster of P-response genes (Heuer et al. 2009), including *OsPTF1*, a transcription factor gene that was shown to enhance plant performance under P deficiency (Yi et al. 2005). In contrast, none of the currently known P-response genes was located within or near the *Pup1* region, based on an analysis of the Nipponbare reference genome (Heuer et al. 2009). In agreement with this, physiological studies of Nipponbare-*Pup1* near-isogenic lines (NILs) failed to reveal commonly known P-starvation responses, such as exudation of organic acids or elongation of

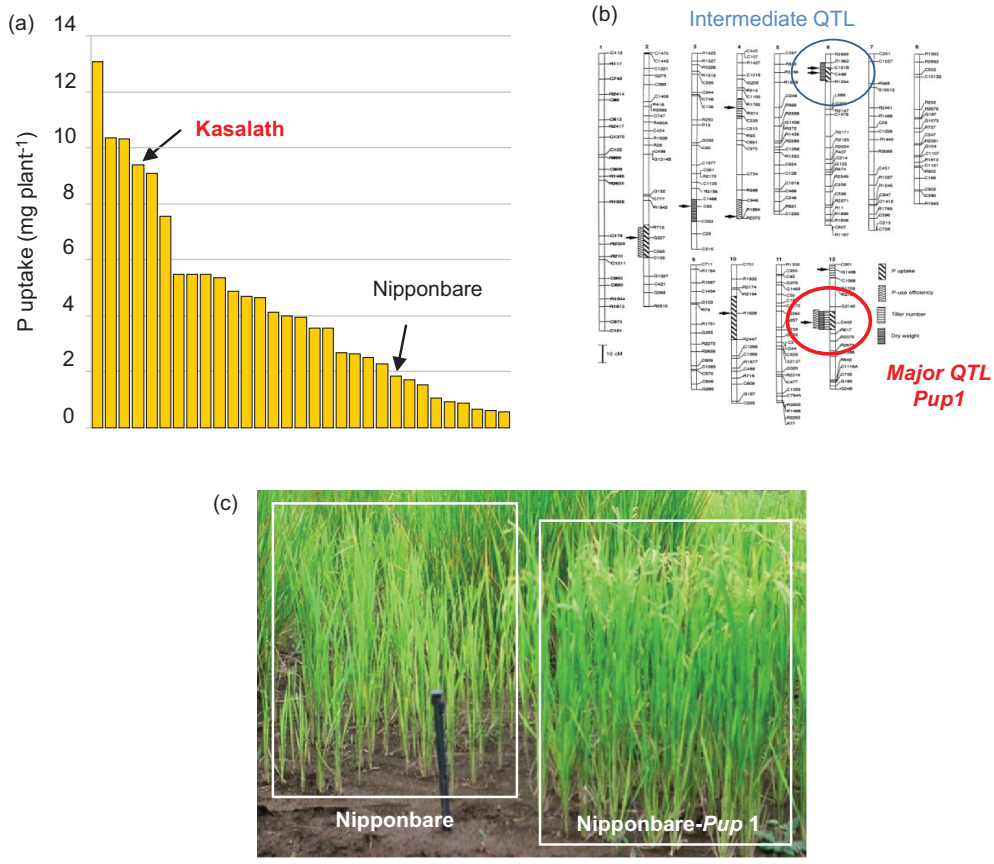


Fig. 5.2. Mapping of the *Pup1* major QTL. A screening of 30 diverse rice accessions under P-deficient field conditions was conducted to identify genotypes with high P-uptake ability (**a**). The data were derived from Wissuwa and Ae (2001a). QTL mapping was conducted using a Kasalath x Nipponbare mapping population and *Pup1* was identified as a large-effect QTL on Chr. 12 (**b**). An intermediate-effect QTL is located on Chr. 6 (Wissuwa et al. 1998; see text for details). Nipponbare and a tolerant near-isogenic line with the *Pup1* QTL were grown in a P-deficient field in Tsukuba, Japan (**c**). For a color version of this figure, please refer to the color plate.

root hairs (Wissuwa et al. unpublished). Likewise, a gene-expression analysis of roots using Agilent arrays did not reveal differential expression of P-starvation genes in the Nipponbare-*Pup1* NILs (Pariasca-Tanaka et al. 2009).

Since the function of *Pup1* could not be deciphered, the genomic region was sequenced using Kasalath BAC clones to identify the major determinant of tolerance and gain insight into the tolerance mechanisms. The sequence of the Kasalath *Pup1* and flanking regions is available under gene bank accession number AB458444. Comparative sequence analyses between the Kasalath *Pup1* locus (Kas-Pup1) and the syntenic region in the Nipponbare (Nip-Pup1) reference genome revealed major differences in size and gene content (Heuer et al. 2009). The main structural difference between the two *Pup1* genomic regions is a large (~90 kb) insertion-deletion (INDEL) that is absent in Nipponbare (Figure 5.3a). Based on an in silico analysis, sixty-eight gene models were predicted, including more than 54% transposable elements (TEs; Heuer et al. 2009). Because many predicted genes showed partial sequence similarity to TEs and functional genes, extensive sequence analyses were required to validate the gene models and to short-list a set of candidate genes. To determine whether the short-listed genes were expressed, gene expression analyses were conducted using Nipponbare NILs with and without the *Pup1* locus grown under P-fertilized (+P) and P-deficient (–P) conditions. Based on this comprehensive gene assessment, five genes were eventually short-listed, including a fatty acid alpha dioxygenase (*OsPupK04-1*), a hypothetical protein located in reverse orientation within an intron of this gene (*OsPupK05-1*), a dirigent gene (*OsPupK20-2*), and a hypothetical protein (*OsPupK29-1*), as well as a protein

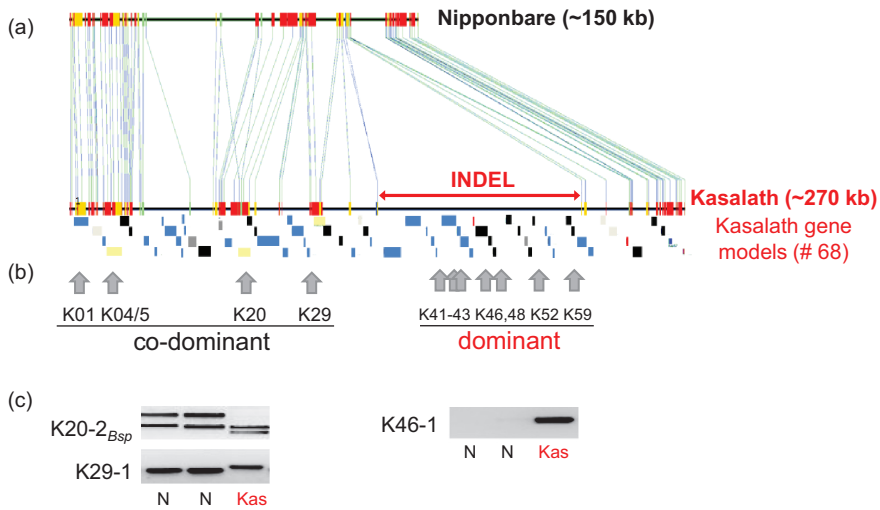


Fig. 5.3. Genomic sequence of the *Pup1* region and gene-specific markers. The *Pup1* genomic region derived from sequencing of Kasalath BAC clones was aligned with the corresponding region in the Nipponbare reference genome (**a**). Some regions show partial sequence similarity (indicated by vertical lines). A large insertion/deletion (INDEL) specific to Kasalath *Pup1* is indicated. Sixty-eight Kasalath gene models (indicated by different-size blocks) were predicted in silico and validated gene models were targeted for the design of allele-specific codominant and dominant markers (**b**). Three ideal *Pup1* markers were identified and are recommended for breeding applications (**c**). Details and references are given in the text. For a color version of this figure, please refer to the color plate.

kinase (*OsPupK46-2*) (Chin et al. 2011). The predicted gene models of the dirigent gene and the protein kinase were subsequently revised based on detailed sequence comparisons with other members of the dirigent and protein kinase gene families, as well as by cDNA sequencing. The revised gene models are available under accession number BAK26565 and BAK26566, respectively. *OsPupK46-2* encodes a functional Ser/Thr protein kinase and was recently renamed *PHOSPHORUS STARVATION TOLERANCE 1* (*OsPSTOL1*; Gamuyao et al. 2012) after it was shown that this gene is the major determinant of tolerance. Analyses of transgenic plants with constitutive *OsPSTOL1* expression (35S::*OsPSTOL1*) showed that *OsPSTOL1* acts as an enhancer of crown root growth at an early developmental stage. This increases the root surface area and enables plants to forage a larger soil area and to take up more P and other nutrients (Gamuyao et al. 2012). It was further shown that the *OsPSTOL1* promoter was active in the coleoptilar node, specifically in the parenchymatic cell layer and in primordia of crown roots, which constitute the post-germination root system in rice. Gene expression profiling using an Affymetrix array further showed that root growth-related and stress-responsive genes were differentially expressed in 35S::*OsPSTOL1* overexpression plants (Gamuyao et al. 2012), whereas P-starvation genes were not differentially expressed in agreement with earlier data derived from Nipponbare-*Pup1* NILs (Pariasca-Tanaka et al. 2009).

Importantly, *OsPSTOL1* is located within the Kasalath-specific INDEL region and is therefore absent from the Nipponbare reference genome. The same was found for the major QTL *SUBMERGENCE 1* (*SUB1*), which confers tolerance of complete submergence for up to two weeks (for review, see Septiningsih et al. 2013). In Nipponbare, the *SUB1* locus contains two ethylene-responsive transcription factor genes (*SUB1B* and *SUB1C*), whereas an additional gene (*SUB1A*) is present in the tolerant *SUB1* locus, which was shown to be the major determinant of tolerance (Xu et al. 2006). The *SUB1* and *Pup1* QTLs thereby exemplify possible limitations of reverse genetic approaches, since they show that tolerant genotypes might possess novel genes and/or employ mechanisms and pathways that are distinct from stress responses in the model variety Nipponbare, which is intolerant.

Pup1-specific Molecular Markers

The identification of major tolerance genes in QTLs is of value for breeding since highly specific markers can be developed. However, knowing the specific gene is not a prerequisite for breeding as long as closely flanking markers are available. More important for breeding applications is that the effect of a given QTL is determined in different environments and genetic backgrounds. Since this requires several generations of crossings and screening in at least two sites and seasons, considerable resources are needed. Most published QTLs are therefore not advanced to this level and this is the main reason why, despite the large number of reported QTLs, only a very few are actively used in breeding programs (Xu and Crouch 2008). However, some major QTLs are already used for molecular breeding of tolerant rice varieties. This includes the *SUB1* QTL (Xu et al. 2006; Septiningsih et al. 2009), the salinity tolerance QTLs *SalTol* and *qSKC1* (Thomson et al. 2010; Ren et al. 2005), as well as QTLs for anaerobic germination (Angaji et al. 2010) and drought tolerance (Kumar et al. 2007; Bernier et al. 2009a; Bernier et al. 2009b; Vikram et al. 2011). Submergence-tolerant rice varieties with the *SUB1* QTL have been quickly adopted by farmers and already show a large impact in flood-prone environments in Asia (Mackill et al. 2012; Manzanilla et al. 2011).

The Sub1 rice varieties have been developed by a marker-assisted backcrossing (MABC) approach, which facilitates targeted and precise introgression of a target QTL. In contrast to other approaches, MABC uses locally adapted and widely grown rice varieties as the QTL

recipient, and subsequently restores the genetic background of the respective local variety by repeated backcrossing and marker selection. Generally, three types of molecular markers are needed for this approach: (1) QTL-specific foreground markers, (2) flanking markers to select for plants with a crossover at the 5′ and 3′ border of the QTL, and (3) background markers to select for plants with maximum recovery of the recipient parent genome (Collard and Mackill 2008). The outcome of this selection is a modified rice variety that is indistinguishable from the original recipient parent with respect to plant type and grain quality, except that it is tolerant. The MABC approach has the advantage that crop duration and desirable traits, such as superior grain quality and disease resistance, remain unchanged and farmers therefore do not need to adjust crop management practices.

Applying this technology for the breeding of P-deficiency-tolerant rice, molecular markers for the *Pup1* QTL were developed based on a comparative sequence analysis between the *Pup1* locus in Kasalath and Nipponbare (Figure 5.3b, c). For regions within the *Pup1* locus that are partially conserved between the two loci, codominant markers were developed based on sequence polymorphisms mainly targeting small (<50 bp) INDELs and single nucleotide polymorphism (SNPs). In contrast, markers targeting the Kas-*Pup1*-specific INDEL region are naturally dominant since this region is absent from Nip-*Pup1* (Chin et al. 2010; Chin et al. 2011). The developed codominant and dominant markers were subsequently tested in a wide range of diverse rice accessions, which revealed that some markers that are polymorphic between Nip-*Pup1* and Kas-*Pup1* were monomorphic in other genotypes (data not shown). Such markers are not generally suitable for breeding applications, but may be used in specific crosses if the parental lines are polymorphic. Among the various markers tested, dominant markers targeting the INDEL (Figure 5.3a, b) were generally the most robust and best associated with *Pup1* across diverse rice genotypes (Chin et al. 2011). However, since dominant markers do not distinguish between homozygous and heterozygous plants, additional genotyping with codominant markers is required. Based on these data and a genotype x marker association analysis, three *Pup1* markers (two codominant and one dominant; Figure 5.3c) have been identified that are ideal foreground markers to select for *Pup1* introgression in MABC-derived progenies (Chin et al. 2011). These three markers also provide sufficient information on the *Pup1* haplotype across diverse rice genotypes (Chin et al., 2011). The primer sequences and additional information on the markers are provided in Table 5.2.

Evaluation of *Pup1* in Different Genetic Backgrounds and Environments

As mentioned above, only QTLs that show a significant effect in different genetic backgrounds and environments are of interest to breeders. For the evaluation of *Pup1*, it was necessary to transfer *Pup1* into *indica*-type rice varieties, since the initially developed Nipponbare-*Pup1* NILs flower early under tropical short-day conditions and therefore were not suitable for field experiments in most of the targeted Asian countries. For this purpose, two *indica*-type irrigated rice varieties (IR64 and IR74) and three Indonesian upland varieties (Situ Bagendit, Dodokan, Batur) were initially selected as *Pup1* recipient parents.

For the development of IR64-*Pup1* and IR74-*Pup1*, two continuous generations of segregating populations were genotyped to select plants with a homozygous *Pup1* locus (Figure 5.4). The use of markers located within (K46-2) and outside (K20-2, K29-1) the INDEL (Figure 5.3b, c) ensured that no recombination occurred within the *Pup1* region. In addition to foreground selection, background markers were used to select for plants with minimum remaining donor segments in combination with visual selection of plants most similar to the recipient parent (Figure 5.4). For background selection, a set of SNP (single nucleotide polymorphism) markers was

Table 5.2. *Pup1*-marker sequences

Marker name	*Pup1* gene model	Marker type	Physical location (bp)		Band size (bp) K/N	Primer sequence (5′−3′)
			Kasalath AB458444.1	Nipponbare chromosome 12		
K1	*OsPupK01-1*	co-dominant	96175−96299	15315156−15315277	125 / 122	F: AGTCTGGATGGACAACTCTGCCTG R: TGCTAGCTCATTGCCGTTACGTCG
K5	*OsPupK05-1*	co-dominant	116430−116701	15336063−15336342	272 / 280	F: ATTCAGACATCGACGGCGAC R: TCCTCGTAAACATGGCTTGC
K20−1	*OsPupK20-2*	co-dominant	169881−170120	15410254−15410496	240 / 243	F: TCAGGTGATGGGAATCATTG R: TGTTCCAACCAAACAACCTG
K20−1$_{MSE}$	*OsPupK20-2*	CAPS (*Mse1*)			201 / 243	
K20−2	*OsPupK20-2*	co-dominant	169290−170607	15409652−15410981	982 / 995	F: TCAAAAATTTCTTCAGGTATGTACTCC R: TTGGGTGATCAGCTTTCAGA
K20−2$_{BSP}$	*OsPupK20-2*	CAPS (*Bsp1286I*)			K: 231+349+402 N: 413+582	
K29−1	*OsPupK29-1*	co-dominant	205067−205287	15431572−15431786	212 / 206	F: ATGGCCAACGGGGTAGAG R: GTCCAGGTAACCACGAGGAA
K29−2	*OsPupK29-1*	co-dominant	204398−204616	15430672−15430883	291 / 212	F: CCCGTCTGCGTTCTACCTTA R: CTCCCGTCAAGCACAAATCT
K29−3	*OsPupK29-1*	co-dominant	202698−202933	15419578−15419825	236 / 248	F: TTCGTCCAGATGCTGCTATG R: TCTTCGGTGTAATTGGCACA
K41	*OsPupK41-1*	dominant	262050−262431	absent	382 / –	F: TGATGAATCCATAGGACAGCGT R: TCAGGTGGTGCTTCGTTGGTA
K42	*OsPupK42-1*	dominant	267154−268071	absent	918 / –	F: CCCGAGAGTTCATCAGAAGGA R: AGTGAGTGGCGTTTGCGAT
K43	*OsPupK43-1*	dominant	268590−269501	absent	912 / –	F: AGGAGGATGAGCCTGAAGAGA R: TCGCACTAACAGCAGCAGATT
K45	*OsPupK45-1*	dominant	274072−274344	absent	276 / –	F: GCGGAAGAAGAGGATAACGA R: TCTAGGCTTCGTTTGGCAAG
K46−1	*OsPupK46-2*	dominant	275710−276232	absent	523 / –	F: TGAGATAGCCGTCAAGATGCT R: AAGGACCACCATTCCATAGC
K46−2	*OsPupK46-2*	dominant	276371−276597	absent	227 / –	F: AGGAAGATGGTTGTCGTTGG R: TTCACACCAAACAGTGTTGTC
K48	*OsPupK48-1*	dominant	282795−283640	absent	847 / –	F: CAGCATTCAGCAAGACAACAG R: ATCCGTGTGGAGCAACTCATC
K52	*OsPupK52-1*	dominant	300870−301374	absent	505 / –	F: ACCGTTCCCAACAGATTCCAT R: CCCGTAATAGCAACAACCCAA
K59	*OsPupK59-1*	dominant	324843−325392	absent	550 / –	F: GGACACGGATTCAAGGAGGA R: TGCTTTCCATTTGCGGCTC

F: forward primer; R: reverse primer.

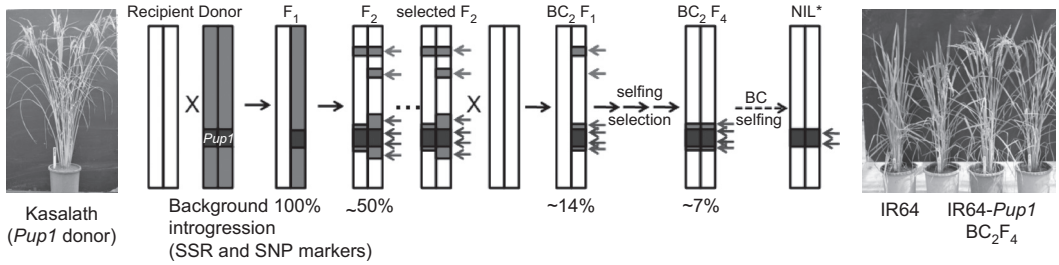


Fig. 5.4. Marker-assisted breeding of Pup1 varieties. The tolerant rice variety Kasalath (photo at the left) or Nipponbare-*Pup1* lines were used as donors for the *Pup1* QTL located on chromosome 12 (shown as red bars; *Pup1* QTL indicated in blue). After crossing with an intolerant recipient variety without *Pup1* (Chr. 12 indicated as white bars), F_1 progenies and the following backcross (BC) and selfing generations are selected using *Pup1* foreground markers (blue arrows), flanking markers (purple arrows), and background markers (red arrows). Representative IR64-*Pup1* plants at the BC_2F_4 generation are shown in the photo to the right. Remaining donor introgression in the different generations was determined using SNP markers and indicated as percentage (all chromosomes). *The introgressions in the schemes do represent the actual data; NILs without any remaining background introgressions are hypothetical and probably not attainable. For a color version of this figure, please refer to the color plate.

used (Thomson et al. 2011). In the case of IR64-*Pup1*, the SNP data showed about 14% remaining donor segments in selected BC_2F_1 plants and about 7% in BC_2F_4 (Figure 5.4). To further restore the genome of the recipient parent, continuous selection under stress and control conditions and additional backcrosses are required. However, with increasing marker density, it can be expected that small remaining donor introgressions will be detected even in later generations.

The availability of molecular markers revealed that, of the varieties that were selected as *Pup1*-recipient parents, only IR64, IR74, and Situ Bagendit did not naturally possess *Pup1*. In contrast, the tolerant allele for *OsPSTOL1* was detected in Batur and Dodokan (Figure 5.5b; Chin et al. 2011). In agreement with this, data derived from a field experiment in Indonesia showed that a beneficial effect of *Pup1* under low-P conditions was indeed observed only in Situ Bagendit-*Pup1* breeding lines (Chin et al. 2011). The natural presence of the *OsPSTOL1* gene in Batur and Dodokan, which are superior Indonesian upland varieties, is not surprising since earlier data had already shown a high degree of *Pup1* conservation in drought-tolerant rice (Heuer et al. 2009; Chin et al. 2010; Chin et al. 2011). In contrast, the *Pup1* locus is absent from most irrigated rice varieties and it was unclear whether *Pup1* would show any effect in irrigated rice systems. It is therefore encouraging that data derived from P-deficient irrigated field experiments showed enhanced early vegetative growth of IR74-*Pup1* breeding lines (Figure 5.5c), as well as enhanced grain yield compared with IR74 and sister lines without *Pup1* (Figure 5.5c, d; Chin et al. 2011). However, under these conditions, *Pup1* does not appear to have an effect in the genetic background of IR64 (Figure 5.4d) and additional field screenings under rain-fed/drought conditions are now ongoing to further evaluate IR64-*Pup1* and to draw final conclusions. Overall, the field data derived from Situ Bagendit-*Pup1* and IR74-*Pup1* suggest that *Pup1* has the potential to confer yield advantages of 10–50% under low-P conditions, depending on the genotype and experimental conditions (Chin et al. 2011; Heuer et al. unpublished).

Phenotyping for Low-P Tolerance

Screening for tolerance of P deficiency under field conditions is constrained by the fact that the distribution of P is typically not uniform across field plots. In addition, P deficiency often occurs in conjunction with other stresses, such as Fe or Al toxicity or drought, and, since these

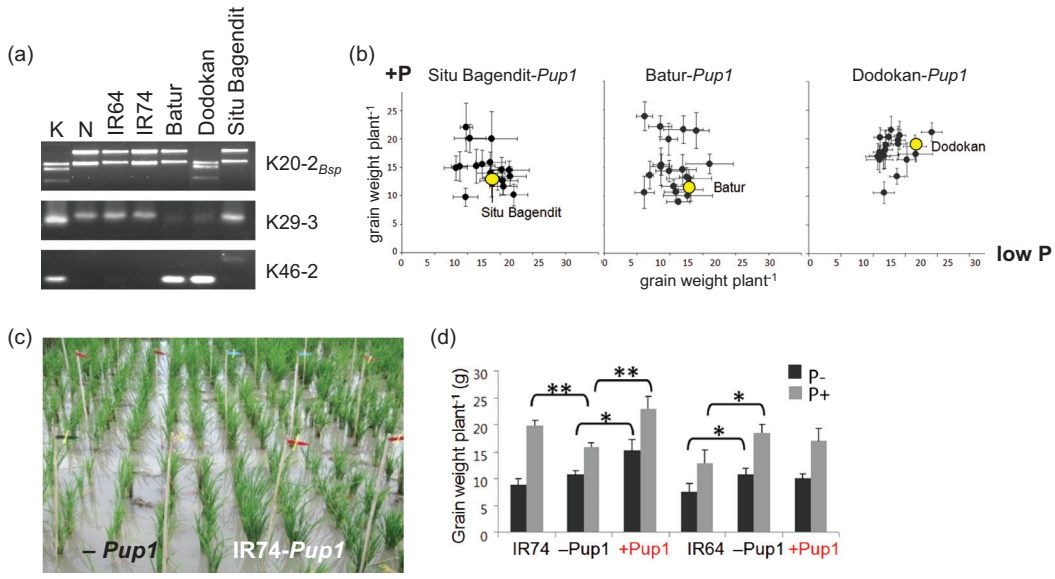


Fig. 5.5. Pup1 breeding lines. The recipient parents selected for *Pup1* introgression were genotyped with three *Pup1* markers (**a**). IR64, IR74, and Situ Bagendit have intolerant Nipponbare-type (N) alleles for all three markers. In contrast, Dodokan and Batur have some Kasalath-type (K) alleles, including marker K46-2, which targets *OsPSTOL1*. The Indonesian Pup1 breeding lines (BC_2F_4) were evaluated in an upland field experiment in West Java (Indonesia) under high- and low-P conditions (**b**). The IR64-*Pup1* and IR74-*Pup1* breeding lines were tested under irrigated field conditions in P-deficient soils showing vigorous growth (**c**) and yield advantage (**d**) of IR74-*Pup1* lines. For a color version of this figure, please refer to the color plate.

accompanying stresses may be equally "patchy," a common problem of field screenings is the high variability and low reproducibility of data. An additional concern is that accompanying stresses, such as Al toxicity, directly affect root growth with subsequent indirect effects on P uptake. In such cases, observed genotypic differences in response to low P may be incorrectly attributed to differences in tolerance, when the driving factor in fact was differences in tolerance of Al toxicity (or other abiotic stresses). It is therefore important to carefully select and analyze soils intended for P screenings.

To avoid complications associated with screening in soil, most studies on P deficiency have been conducted in hydroponic culture solution at P concentrations ranging from 3 μM (Wissuwa et al. 2005) to 15 μM (Ni et al. 1998). This rather wide range of P concentrations highlights the fact that P concentrations alone may be slightly misleading; the total amount of P available per plant over a period of time should be the most important criterion, and that in turn will depend on plant age, duration of the experiment, interval of renewing nutrient solutions, total volume, and number of co-cultivated plants sharing the same solution. Thus, P treatments in nutrient solution experiments should ideally be defined by total nutrient per plant and unit time (e.g., μg P $plant^{-1}$ day^{-1}).

Whether the ease of screening and the high repeatability of nutrient solution experiments outweigh the main drawback, namely, the uncertainty about whether results obtained can be transferred to field situations, is a matter of debate. In the field, tolerance of P deficiency will depend on multiple tolerance mechanisms that can be classified into three broad categories (Ismail et al. 2007): (1) root interception of P as affected by root growth, anatomy, and architecture; (2) P-acquisition efficiency due to rhizosphere modifications by the plant root that

would enhance soil-P availability; and (3) plant-internal P-use efficiency, which allows more efficient genotypes to produce more biomass per unit P taken up. Screening in nutrient solution may detect genotypic differences in mechanisms related to (1) and (3), depending on the design of the experiment, and might be essential for screenings for internal P-use efficiency. For the latter, it needs to be assured that an equal amount of P is provided to each plant, which requires that genotypes be grown individually in separate containers supplied with the same amount of P (Rose et al. 2011). If distinct genotypes share the same container, they will compete for nutrients and this typically favors genotypes with a larger root system (Yi et al. 2005; Wissuwa et al. 2006), presumably because the number of P uptake sites (P transporters) is proportional to the root surface area. Internal efficiency would potentially play an additional role and may even be driving root growth. Thus, screening of different genotypes in the same container will select for a combination of traits related to root growth and internal P-use efficiency.

Comparative experiments in nutrient solution and soil or field are rare and it is therefore difficult to come to a general conclusion. Yi and colleagues (2005) showed that overexpressing the transcription factor gene *OsPTF1* increased performance under P deficiency by about 30%, and that this positive effect was similar in pots filled with soil and in nutrient solution. On the other hand, Wissuwa and colleagues (2005) did not detect significant growth differences in solution between Nipponbare-*Pup1* sister lines with and without the *Pup1* QTL, whereas that same pair differed by more than 2-fold when grown in soil (Wissuwa 2005). On the other hand, analyses of IR64-*Pup1* and IR74-*Pup1* NILs in nutrient solution revealed differences in root surface area in young seedlings compared with non-*Pup1* sister lines (Gamuyao et al. 2012). Even if nutrient solution experiments offer some initial advantages, it will be crucial to validate and confirm genotypic differences and QTL effects in field experiments to assure that the observed effects are relevant for breeding applications. Furthermore, it is critically important to carefully consider target environments, particularly with regard to soil type and water management, since the effect of certain tolerance mechanisms (see above) might be soil-specific and affected by the water regime during the cropping cycle.

Outlook and Perspectives

So far, *Pup1* is the only available major QTL for tolerance of P deficiency that is used for molecular breeding of tolerant rice varieties. That *Pup1* has been identified in a field screening assured its practical relevance, which was an important driver for the continuous effort to elucidate its function and establish a molecular breeding platform. The amount of resources and time required to develop and test markers and to evaluate QTLs in different genetic backgrounds and environments is considerable and can be pursued only for large-effect QTLs with potentially high impact. However, with the experiences gained and the progress in molecular technologies, it will become possible to identify and validate QTLs more efficiently and faster.

Given that other rice genotypes have been found with a higher P deficiency tolerance than Kasalath (Wissuwa and Ae 2001a; Wissuwa unpublished), additional major QTLs can likely be identified if suitable screening conditions are applied. This could follow the traditional approach using bi-parental mapping populations as employed successfully in the case of *Pup1*. In this case, one parent should have exceptionally high tolerance of P deficiency in order to increase the likelihood of identifying strong (and possibly rare) loci/alleles since not all QTLs identified justify the effort needed to initiate a MABC breeding program. Increasingly, mapping of novel alleles will rely on genome-wide association studies (GWAS) that typically use several hundred gene-bank accessions, therefore capturing a much larger portion of the genetic variation. New technologies, such as high-density SNP markers and genotyping by

sequencing (Yu et al. 2011; Thomson et al. 2011; Elshire et al. 2011), are fast evolving and already facilitate rapid and high-throughput genotyping.

Field screening experiments conducted with bi-parental or association mapping populations should ideally be conducted in target environments. If the genetic material screened is of rather different plant type, as one typically observes in association panels used in GWAS, screening for grain yield alone would be potentially misleading, since traditional rice genotypes may be very tolerant but typically have low yield potential and low harvest index. It may therefore be more advisable to screen for component traits that contribute to P efficiency, such as P-acquisition efficiency (PAE) or internal P-utilization efficiency (PUE). In contrast to evaluations of PAE that ultimately need to be done in soil, Rose and colleagues (2011) established that screening under such realistic soil/field conditions would make the detection of genotypic differences in PUE nearly impossible because any difference in P uptake between genotypes would confound measurements of PUE. It is therefore crucial to evaluate PUE in a setup that guarantees that genotypes have equal P content, and the simple method developed by Rose and colleagues (2011) was designed to achieve that. One additional P-efficiency trait that has been advocated only recently is the development of cultivars with reduced grain-P loading (Rose and Wissuwa 2012). Currently, about 75% of the P taken up by the rice crop is exported from the field with the harvested grain (Rose et al. 2010). Lowering grain-P concentrations to a level that does not affect seedling vigor would therefore offer opportunities to enhance long-term sustainability of low-input systems by reducing P mining, and sustainability of high-input systems by reducing requirements for the application of maintenance-P fertilizer.

The availability of large-effect genes/QTLs and molecular markers facilitates pyramiding of desirable alleles through MABC. Pyramiding of submergence (*SUB1*) and salinity (*SalTol*) tolerance is already at an advanced stage and pyramiding of *Pup1* with major drought QTLs is in preparation (IRRI unpublished data). For pyramiding, individual QTLs are first introgressed into a widely grown and locally well-adapted rice variety (mega-variety; e.g., IR64) and the genetic makeup of the recipient parent is then restored through MABC, that is, a series of two to three backcrosses and marker-assisted selection of progenies with the QTL and minimal presence of markers for background donor segments. In a second step, the individual introgression lines (i.e., IR64-SUB1, IR64-SalTol, and IR64-Pup1) are crossed and progenies selected using QTL-specific foreground markers. Since the genetic background of the individual QTL lines is nearly identical, background genotyping can be reduced to a minimum.

This principle, so far followed to combine tolerance of multiple stresses, would be equally suited for pyramiding tolerance components for a given trait such as P efficiency. As outlined above, the limiting factor is now a lack of validated, large-effect QTLs/genes and efforts should thus focus on identifying those. This search for novel P-efficiency alleles would have to be broadened beyond the current focus on mechanisms improving P uptake, to include QTLs/genes controlling internal P-utilization efficiency and grain-P content. We believe this to be crucially important because improvements in either alone would not assure long-term sustainability of cropping systems with regard to P.

In fact, the very recent paradigm shift in dissecting natural variation in crops such as rice away from bi-parental populations with their limited variability to SNP-based or even sequence-based association mapping of panels consisting of gene bank accessions for the first time makes it possible to tap the variation present in gene pools hidden in gene-bank collections. Advances in genotyping technologies are beginning to remove most limitations with regard to genotypic characterizations, for both exploratory and breeding purposes. Decreasing costs per marker or data point and the possibility of outsourcing marker analysis to service labs

furthermore means that MABC strategies could be employed even in less developed regions of the world. With molecular tools routinely applied in breeding programs, the actual challenge is to carefully choose relevant traits and to develop phenotyping protocols that are a realistic reflection of the field situation.

References

Ai P, Sun S, Zhao J, Fan X, Xin W, Guo Q, Yu L, Shen Q, Wu P, Miller AJ, Xu G. 2009. Two rice phosphate transporters, OsPht1;2 and OsPht1;6, have different functions and kinetic properties in uptake and translocation. *Plant J* 57:798–809.

Angaji SA, Septiningsih EM, Mackill DJ, Ismail AM. 2010. QTLs associated with tolerance to anaerobic conditions during germination in rice (*Oryza sativa* L.). *Euphytica* 172:159–168.

Aung K, Lin SI, Wu CC, Huang YT, Su CL, Chiou TJ. 2006. *pho2*, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. *Plant Physiol* 141:1000–1011.

Bari R, Datt Pant B, Stitt M, Scheible WR. 2006. PHO2, microRNA399 and PHR1 define a phosphate–signaling pathway in plants. *Plant Physiol* 141:988–999.

Batjes NH. 1997. A world data set of derived soil properties by FAO-UNESCO soil unit for global modeling. *Soil Use Manage* 13:9–16.

Batjes NH. 2011. Global distribution of soil phosphorus retention potential. Wageningen, ISRIC - World Soil Information (with dataset), ISRIC Report 2011/06, 42 pp.

Bernier J, Kumar A, Venuprasad R, Spaner D, Verulkar S, Mandal NP, Sinha PK, Peeraju P, Dongre PR, Mahto RN, Atlin G. 2009a. Characterization of the effect of a QTL for drought resistance in rice, *qtl12.1*, over a range of environments in the Philippines and eastern India. *Euphytica* 166:207–217.

Bernier J, Serraj R, Kumar A, Venuprasad R, Impa S, Gowda RPV, Oane R, Spaner D, Atlin G. 2009b. The large-effect drought-resistance QTL *qtl12.1* increases water uptake in upland rice. *Field Crops Res* 110:139–146.

Bucher M. 2007. Functional biology of plant phosphate uptake at root and mycorrhiza interfaces. *New Phytol* 173:11–26.

Chin JH, Gamuyao R, Dalid C, Bustamam M, Prasetiyono J, Moeljopawiro S, Wissuwa M, Heuer S. 2011. Developing rice with high yield under phosphorus deficiency: *Pup1* sequence to application. *Plant Physiol* 156:1202–1216.

Chin JH, Lu X, Haefele SM, Gamuyao R, Ismail A, Wissuwa M, Heuer S. 2010. Development and application of gene-based markers for the major rice QTL *Phosphorus uptake 1*. *Theor Appl Genet* 120:1073–1086.

Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL. 2006. Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. *Plant Cell* 18:412–421.

Collard BCY, Mackill DJ. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Phil Trans R Soc B* 363:557–572.

Cordell D, Drangert JO, White S. 2009. The story of phosphorus: Global food security and food for thought. *Global Environ Change* 19:292–305.

Delhaize E, Randall PJ. 1995. Characterization of a phosphate-accumulator mutant of *Arabidopsis thaliana*. *Plant Physiol* 107:207–213.

Dong B, Rengel Z, Delhaize E. 1998. Uptake and translocation of phosphate by *pho2* mutant and wild-type seedlings of *Arabidopsis thaliana*. *Planta* 205:251–256.

Duan K, Yi K, Dang L, Huang H, Wu W, Wu P. 2008. Characterization of a sub–family of *Arabidopsis* genes with the SPX domain reveals their diverse functions in plant tolerance to phosphorus starvation. *Plant J* 54:965–975.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5): e19379. doi:10.1371/journal.pone.0019379.

Fageria NK, Wright RJ, Baligar VC. 1988. Rice cultivar evaluation for phosphorus use efficiency. *Plant Soil* 111:105–109.

Franco–Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio–Somoza I, Leyva A, Weigel D, García JA, Paz–Ares J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 39:1033–1037.

Fujii H, Chiou TJ, Lin SI, Aung K, Zhu JK. 2005. A miRNA involved in phosphate–starvation response in *Arabidopsis*. *Curr Biol* 5:2038–2043.

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S. 2012. The protein kinase pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* 488:535–541.

Haefele SM, Hijmans RJ. 2007. Soil quality in rice-based rainfed lowlands of Asia: Characterization and distribution. In Aggarwal PK, Ladha JK, Singh RK, Devakumar C, Hardy B, eds, *Science, technology, and trade for peace and prosperity: Proceedings of the 26th International Rice Research Conference, New Delhi, India*. International Rice Research Institute, Indian Council of Agricultural Research, and National Academy of Agricultural Sciences, Los Baños, Philippines, and New Delhi, India, pp. 297–308.

Haefele SM, Hijmans RJ. 2009. Soil quality in rainfed lowland rice. *Rice Today* 8 (1):30–31.

Heuer S, Lu X, Chin JH, Tanaka JP, Kanamori H, Matsumoto T, De Leon T, Ulat VJ, Ismail AM, Yano M, Wissuwa M. 2009. Comparative sequence analyses of the major

quantitative trait locus *phosphorus uptake 1* (*Pup1*) reveal a complex genetic structure. *Plant Biotech J* 7:456–457.

Hu B, Zhu C, Li F, Tang J, Wang Y, Lin A, Liu L, Che R and Chu C. 2011. LEAF TIP NECROSIS1 plays a pivotal role in the regulation of multiple phosphate starvation responses in rice. *Plant Physiol* 156:1101–1115.

Huguenin-Elie O, Kirk GJD, Frossard E. 2009. The effects of water regime on phosphorus responses of rainfed lowland rice cultivars. *Ann Bot* 103:211–220.

Ismail AM, Heuer S, Thomson MJ, Wissuwa M. 2007. Genetic and genomic approaches to develop rice germplasm for problem soils. *Plant Mol Biol* 65:547–570.

Jia H, Ren H, Gu M, Zhao J, Sun S, Zhang X, Chen J, Wu P, Xu G. 2011. The phosphate transporter gene *OsPht1;8* is involved in phosphate homeostasis in rice. *Plant Physiol* 156:1164–1175.

Kirk GJD. 2004. *The biogeochemistry of submerged soils*. Chichester: Wiley.

Kraft E, Stone SL, Ma L, Su N, Gao Y, Lau OS, Deng XW, Callis J. 2005. Genome analysis and functional characterization of the E2 and RING–type E3 ligase ubiquitination enzymes of *Arabidopsis*. *Plant Physiol* 139:1597–1611.

Kumar R, Venuprasad R, Atlin GN. 2007. Genetic analysis of rainfed lowland rice drought tolerance under naturally-occurring stress in eastern India: Heritability and QTL effects. *Field Crops Res* 103:42–52.

Lambers H, Shane MW, Cramer MD, Pearse SJ, Veneklaas EJ. 2006. Root structure and functioning for efficient acquisition of phosphorus: Matching morphological and physiological traits. *Ann Bot* 98:693–713.

Lin SI, Chiang SF, Lin WY, Chen JW, Tseng CY, Wu PC, Chiou TJ. 2008. Regulatory network of microRNA399 and PHO2 by systemic signaling. *Plant Physiol* 147:732–746.

Lin SI, Chiou TJ. 2008. Long–distance movement and differential targeting of microRNA399s. *Plant Signal Behav* 3:730–732.

Lynch JP. 2011. Root phenes for enhanced soil exploration and phosphorus acquisition: Tools for future crops. *Plant Physiol* 156:1041–1049.

MacDonald GK, Bennett EM, Potter PA, Ramankutty N. 2011. Agronomic phosphorus imbalances across the world's croplands. *Proc Nat Acad Sci USA* 108:3086–3091.

MacIntosh GC. 2011. RNase T2 family: Enzymatic properties, functional diversity, and evolution of ancient ribonucleases. In: Nicholson AWW, ed, *Ribonucleases*, Vol 26. Berlin and Heidelberg: Springer, pp. 89–114.

Mackill DJ, Ismail AM, Singh US, Labios RV, Paris TR. 2012. Development and rapid adoption of submergence-tolerant (Sub1) rice varieties. *Adv Agron* 11:299–352.

Manzanilla DO, Paris TR, Vergara GV, Ismail AM, Pandey S, Labios RV, Tatlonghari GT, Acda RD, Chi TTN, Duoangsila K, Siliphouthone I, Manikmas MOA, Mackill DJ. 2011. Submergence risks and farmers' preferences: Implications for breeding Sub1 rice in Southeast Asia. *Agric Syst* doi:10.1016/j.agsy.2010.12.005.

Miura K, Rus A, Sharkhuu A, Yokoi S, Karthikeyan AS, Raghothama KG, Baek D, Koo YD, Jin JB, Bressan RA, Yun DJ, Hasegawa PM. 2005. The *Arabidopsis* SUMO E3 ligase SIZ1 controls phosphate deficiency responses. *Proc Natl Acad Sci USA* 102:7760–7765.

Ni JJ, Wu P, Senadhira D, Huang N. 1998. Mapping QTLs for phosphorus deficiency tolerance in rice (*Oryza sativa* L.). *Theor Appl Genet* 97:1361–1369.

Nilsson L, Müller R, Nielsen TH. 2007. Increased expression of the MYB–related transcription factor, PHR1, leads to enhanced phosphate uptake in *Arabidopsis thaliana*. *Plant Cell Environ* 30:1499–1512.

Nussaume L, Maréchal E, Thibaud MC, Block MA. 2011. Plant plasma membrane and phosphate deprivation. In Murphy AS, Peer W, and Schulz B, eds, *The plant cell membrane*. Plant Cell Monographs, Volume 19/2011, 237–251, DOI: 10.1007/978-3-642-13431-9_10.

Pant BD, Buhtz A, Kehr J, Scheible WR. 2008. MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *Plant J* 53:731–738.

Pariasca-Tanaka J, Satoh K, Rose T, Mauleon R, Wissuwa M. 2009. Stress response versus stress tolerance: A transcriptome analysis of two rice lines contrasting in tolerance to phosphorus deficiency. *Rice* 2:167–185.

Paszkowski U, Kroken S, Roux C, Briggs SP. 2002. Rice phosphate transporters include an evolutionarily divergent gene specifically activated in arbuscular mycorrhizal symbiosis. *Proc Natl Acad Sci USA* 99:13324–13329.

Péret B, Clément M, Nussaume L, Desnos T. 2011. Root developmental adaptation to phosphate starvation: Better safe than sorry. *Trends Plant Sci* 16:442–450.

Plaxton W, Podestá F. 2006. The functional organization and control of plant respiration. *Crit Rev Plant Sci* 25:159–198.

Plaxton WC, Tran HT. 2011. Metabolic adaptations of phosphate–starved plants. *Plant Physiol* 156:1006–1015.

Raghothama KG. 1999. Phosphate acquisition. *Annu Rev Plant Physiol Plant Mol Biol* 50:665–693.

Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX. 2005. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nature Genet* 37:1141–1146.

Richardson AE, Lynch JP, Ryan PR, Delhaize E, Smith FA, Smith SE, Harvey PR, Ryan MH, Veneklaas EJ, Lambers H, Oberson A, Culvenor RA, Simpson RJ. 2011. Plant and microbial strategies to improve the phosphorus efficiency of agriculture. *Plant Soil* 349:121–156.

Rose TJ, Pariasca-Tanaka J, Rose MT, Fukuta Y, Wissuwa M. 2010. Genotypic variation in grain phosphorus concentration and opportunities to improve P-use efficiency in rice. *Field Crops Res* 119: 154–160.

Rose TJ, Rose MT, Pariasca-Tanaka J, Heuer S, Wissuwa M. 2011. The frustration with utilization: Why have improvements in internal phosphorus utilization

efficiency in crops remained so elusive? *Front Plant Nutri* 2: doi: 10.3389/fpls.2011.00073.

Rose TJ, Wissuwa M. 2012. Rethinking internal phosphorus utilization efficiency (PUE): A new approach is needed to improve PUE in grain crops. *Adv Agron* 116:183–215.

Rouached H, Stefanovic A, Secco D, Arpat AB, Gout E, Bligny R, Poirier Y. 2011. Uncoupling phosphate deficiency from its major effects on growth and transcriptome via *PHO1* expression in *Arabidopsis*. *Plant J* 65:557–570.

Rubio V, Linhares F, Solano R, Martín AC, Iglesias J, Leyva A, Paz–Ares J. 2001. A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes Dev* 15:2122–2133.

Sato A, Miura K. 2011. Root architecture remodeling induced by phosphate starvation. *Plant Signal Behav* 6:1122–1126.

Sawers RJ, Gutjahr C, Paszkowski U. 2008. Cereal mycorrhiza: An ancient symbiosis in modern agriculture. *Trends Plant Sci* 13:93–97.

Seo HM, Jung Y, Song S, Kim Y, Kwon T, Kim DH, Jeung SJ, Yi YB, Yi G, Nam MH, Nam J. 2008. Increased expression of OsPT1, a high–affinity phosphate transporter, enhances phosphate acquisition in rice. *Biotechnol Lett* 30:1833–1838.

Septiningsih EM, Collard BCY, Heuer S, Bailey-Serres J, Ismail AM, Mackill D. 2013. Applying genomics tools for breeding submergence tolerance in rice. In: Varshney RK, Tuberosa R, eds, *Translational genomics for crop breeding, vol. 2, Abiotic stress, yield, and quality*. Ames IA, Chichester, and Oxford: John Wiley & Sons Inc.

Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara GV, Heuer S, Ismail AM, Mackill DJ. 2009. Development of submergence-tolerant rice cultivars: The *Sub1* locus and beyond. *Ann Bot* 103:151–160.

Thomson MJ, Ocampo M, Egdane J, Rahman MA, Sajise AG, Adorada DL, Tumimbang-Raiz E, Blumwald E, Seraj ZI, Singh RK, Gregorio GB, Ismail AM. 2010. Characterizing the *Saltol* quantitative trait locus for salinity tolerance in rice. *Rice* 3:148–160.

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung CW, Reynolds A, Scheffler B, Eizenga G, McClung A, Kim H, Ismail AM, de Ocampo M, Mojica C, Reveche MY, Dilla-Ermita CJ, Mauleon R, Leung H, Bustamante C, McCouch SR. 2011. High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Mol Breed* DOI 10.1007/s11032-011-9663-x.

Van Kauwenbergh SJ. 2010. World phosphate rock reserves and resources. *International Fertilizer Development Center, IFDC Technical Bulletin No. 75*. Muscle Shoals AL, 58 pp.

Vance CP, Chiou TJ. 2011. Phosphorus focus editorial. *Plant Physiol* 156:987–988.

Vikram P, Swamy BPM, Dixit S, Ahmed HU, Sta Cruz MT, Singh AK, Kumar A. 2011. *qDTY1.1*, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. *BMC Genet* 12:89 doi:10.1186/1471-2156-12-89.

Wang C, Ying S, Huang H, Li K, Wu P, Shou H. 2009. Involvement of OsSPX1 in phosphate homeostasis in rice. *Plant J* 57:895–904.

Wissuwa M. 2005. Combining a modeling with a genetic approach in establishing associations between genetic and physiological effects in relation to phosphorus uptake. *Plant Soil* 269:57–68.

Wissuwa M, Ae N. 2001a. Further characterization of two QTLs that increase phosphorus uptake of rice (*Oryza sativa* L.) under phosphorus deficiency. *Plant Soil* 237:275–286.

Wissuwa M, Ae N. 2001b. Genotypic variation for tolerance to phosphorus deficiency in rice and the potential for its exploitation in rice improvement. *Plant Breed* 120: 43–48.

Wissuwa M, Gamat G, Ismail AM. 2005. Is root growth under phosphorus deficiency affected by source or sink limitations? *J Exp Bot* 56:1943–1950.

Wissuwa M, Ismail AM, Yanagihara S. 2006. Effects of zinc deficiency on rice growth and genetic factors contributing to tolerance. *Plant Physiol* 142:731–741.

Wissuwa M, Mazzola M, Picard C. 2009. Novel approaches in plant breeding for rhizosphere-related traits. *Plant Soil* 321:409–430.

Wissuwa M, Wegner J, Ae N, Yano M. 2002. Substitution mapping of *Pup1*: A major QTL increasing phosphorus uptake of rice from a phosphorus deficient soil. *Theor Appl Genet* 105:890–897.

Wissuwa M, Yano M, Ae N. 1998. Mapping of QTLs for phosphorus-deficiency tolerance in rice (*Oryza sativa* L.). *Theor Appl Genet* 97:777–783.

Xu Y, Crouch JH. 2008. Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci* 48:391–407.

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ. 2006. *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442:705–708.

Yi K, Wu Z, Zhou J, Du L, Guo L, Wu Y, Wu P. 2005. *OsPTF1*, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiol* 138:2087–2096.

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q. 2011. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS ONE* 6(3):e17595.

Zhou J, Jiao F, Wu Z, Li Y, Wang X, He X, Zhong W, Wu P. 2008. OsPHR2 is involved in phosphate-starvation signaling and excessive phosphate accumulation in shoots of plants. *Plant Physiol* 146:1673–1686.

Chapter 6

Aluminum Tolerance in Sorghum and Maize

Jurandir V. Magalhaes, Lyza G. Maron, Miguel A. Piñeros, Claudia T. Guimarães, and Leon V. Kochian

Abstract

The soils of the tropics and subtropics are highly weathered, leading to poor soil fertility and low soil pH. Root growth and function on these acid soils is impaired by aluminium (Al) toxicity, leading to a yield instability that jeopardizes food security worldwide. A wealth of physiological evidence exists for an Al-tolerance mechanism based on Al exclusion from the growing root tip. This is facilitated by the release of Al-binding organic acids such as malate and citrate, which keeps rhizotoxic Al away from sensitive sites in the root apex. More recently, Al-activated organic acid transporters in the ALMT and MATE (multidrug and toxic compound extrusion) protein families have been cloned and provide the molecular support for this Al-tolerance mechanism. Here a historical review of Al tolerance in maize and sorghum is presented, followed by an analysis of the more recent research on the molecular determinants of Al tolerance. We show that Al tolerance provided by MATE proteins spans the genetic divergence between sorghum and maize, and is a conserved physiological mechanism in both species. Some features of this mechanism are strikingly common in sorghum and maize, such as the close relationship between phenotypic variation and *MATE* gene expression. However, while the genetic basis for maize aluminium tolerance is quantitative, in sorghum, *SbMATE* underlies a major Al-tolerance locus. More subtle features of this Al-tolerance trait are now emerging, such as the importance of *trans*-acting factors in sorghum, whereas Al-tolerance gene expression in maize appears to be predominantly controlled in cis. Knowledge of the molecular basis of Al tolerance is now providing the framework to address pivotal historical questions in the field, such as the occurrence of genetic background effects for Al tolerance. We advocate the point of view that the answer to such questions will inevitably form the basis for modern molecular breeding strategies designed to explore in full the potential for genetic solutions to the Al-tolerance problem for crops grown on acid soils.

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

Introduction

Importance of Acid Soils in Limiting Worldwide Agriculture

Aluminum (Al) toxicity is the primary factor limiting crop production on acidic soils. At soil pH values below 5, the rhizotoxic Al species, Al^{3+}, is solubilized into the soil solution, inhibiting root growth and function and thus reducing crop yields. Acid soils limit agricultural productivity in many regions of the world. Approximately 30% of the world's total land area consists of acid soils, and it has been estimated that more than 50% of the world's potentially arable lands are acidic (von Uexküll and Mutert 1995; Wood et al. 2000). Significant portions of the land acreage used to produce important grain crops are also acidic, and maize is one of the most important grain crops grown on acid soils. Approximately 20% of all maize grown is found on acid soils. A large proportion of the acid soils occur in developing countries in the tropics and subtropics; it has been estimated that the humid tropics account for 60% of the acid soils in the world. Thus, acid soils limit maize yields in many developing countries where food production is critical. Furthermore, in developed countries such as the United States, high-input farming practices such as the extensive use of ammonia fertilizers are causing additional soil acidification of agricultural soils. While liming of acid soils can ameliorate soil acidity, this is neither an economic option for poor farmers nor an effective strategy for alleviating subsoil acidity.

Progress on Physiology of Crop Al Tolerance

Plants avoid the phytotoxic effects of Al^{3+} by employing physiological mechanisms aimed at excluding Al^{3+} from entering the root apex, which is the primary site of Al toxicity (Al exclusion), and/or by mechanisms that confer the ability to tolerate Al as it enters the plant symplasm (Al tolerance). Compelling evidence indicates that many plant species make use of a generalized exclusion mechanism based on the chelation of Al^{3+} by organic acid anions transported out of root apex cells using specialized plasma membrane-localized transporters. Pioneer work established that Al tolerance in wheat was correlated with a strong Al-activated exudation of malate (a dicarboxylic acid anion) (Delhaize et al. 1993a; Delhaize et al. 1993b). Since then, Al-tolerant genotypes from many plant species, including maize, have been shown to make use of this same mechanism of Al-exclusion, with the identity of the organic acid released being the main difference among plant species (see Table I in Kochian et al. 2004).

Progress on the Molecular Biology of Crop Al Tolerance

Considerable progress has been made over the past eight years on identifying and characterizing plant Al-tolerance genes. The first plant Al-tolerance gene to be identified, *TaALMT1*, was cloned from wheat. *TaALMT1* encodes a member of a novel family of organic acid transporters, the ALMT family, which when expressed heterologously confers Al-activated malate efflux and increased Al tolerance in plants (Delhaize et al. 2004; Sasaki et al. 2004). Electrophysiological studies established that TaALMT1, as well as TaALMT1 orthologs identified in Arabidopsis (*Arabidopsis thaliana)* (AtALMT1; Hoekenga et al. 2006) and rape (*Brassica napus*; Ligaba et al. 2006), mediate a selective efflux of malate that is greatly enhanced by high affinity direct binding of Al^{3+} to the transporter (Piñeros et al. 2008; Zhang et al. 2008). The biochemistry and genetics of the ALMT family, and their involvement in mediating Al^{3+} tolerance mechanisms based on organic anion efflux, have been reviewed (Delhaize et al. 2007).

More recently a second class of Al-tolerance genes was identified, which are transporters in the MATE family that mediate root citrate efflux and contribute to Al tolerance in a number of plant species. MATEs were first identified as

Al-tolerance genes from the fine mapping of Al-tolerance loci in sorghum (Magalhaes et al. 2007) and barley (Furukawa et al., 2007). Our group identified the sorghum MATE via the positional cloning of the major sorghum Al-tolerance locus, Alt_{SB}. This gene, *SbMATE*, encodes a plasma membrane citrate transporter responsible for the root citrate exudation in response to Al stress (Magalhaes et al. 2007). More recently we identified a maize ortholog, *ZmMATE1*, as a major Al-tolerance gene in maize (Maron et al. 2010). The identification of these genes in sorghum and maize, along with HvMATE (Furukawa et al. 2007), AtMATE (Hoekenga et al. 2006), OsFRDL4 (Yokosho et al. 2011), ScFRDL2 (Yokosho et al. 2010), and VuMATE1 (Yang et al. 2011), which are root citrate efflux transporters involved in barley, Arabidopsis, rice, rye, and rice bean Al tolerance, indicates a broad role for this subgroup of plant MATEs in crop Al tolerance.

The identification of these plant Al-tolerance genes has provided the materials for more efficient and effective molecular breeding of enhanced crop Al tolerance. In this chapter, we will describe the characterization of the major sorghum and maize Al-tolerance genes, *SbMATE* and *ZmMATE1*, and how we are using the information from this research to develop molecular breeding pipelines for improved sorghum and maize adaptation to acid soils.

Sorghum Al Tolerance

Historical Aspects of Aluminum Tolerance in Sorghum

As Al tolerance is not frequently found in wheat, Garvin and Carver (2002) have suggested that this trait represents a derived state rather than an inherent characteristic of the crop. This also appears to pertain to sorghum, since among world populations of sorghum screened up to 1993, only 5% of the lines showed appreciable tolerance to Al (Foy et al. 1993). Nonetheless, extensive genetic variability for Al tolerance exists and has long been utilized by breeders to develop sorghum cultivars adapted to acid soils. In contrast to barley and, to a lesser extent, wheat, inheritance of Al tolerance in sorghum has not been historically explained by a simple genetic model. A high degree of general combining ability (GCA) for Al tolerance has been reported (Boye-Goni et al. 1985; Borgonovi et al. 1987a; Gourley et al. 1990; Flores et al. 1991), which suggests that genes with additive effects could be controlling sorghum Al tolerance. However, specific combining ability (SCA) effects were also found to be significant, although less important than GCA effects. This fact, along with an early report of a bimodal frequency distribution in progeny derived from the cross of Al-tolerant and Al-sensitive sorghum cultivars (Furlani and Bastos 1990), strongly suggested the existence of dominant genes with major phenotypic effects in some of the sources of sorghum Al tolerance. Borgonovi et al. (1987b) summarized the findings of a number of field and hydroponic-based studies on Al tolerance and stated that, in general, sorghum Al tolerance is controlled by a few major genes with dominant effects, probably one partially dominant gene, and several minor genes with some additive effects. Maternal effects for Al tolerance have not been commonly observed in sorghum. Heritability associated with the trait was reported as high for Al tolerance assessments based on Al inhibition of root growth (Boye-Goni and Marcarian 1983; Borgonovi et al. 1987a), indicating that a large proportion of the variability may be explained by genetic factors (Borgonovi et al. 1997a). The observed high degree of broad-sense heritability (Furlani and Bastos 1990) indicates that assessment of Al tolerance, either for breeding or genetic mapping purposes, should be effective in the F_2 generation.

SC283 (IS7173), a sorghum cultivar belonging to the guinea race collected in Tanzania, is the most widely accepted Al-tolerance standard in this species. This cultivar has consistently exhibited exceptional values for Al tolerance in the United States (Duncan et al. 1983; Duncan 1988; Foy et al. 1993) and Brazil

(Borgonovi et al. 1987a; Borgonovi et al. 1987b; Furlani et al. 1987), and was consistently scored as the most tolerant genotype both in field (acid soil) trials and in hydroponic studies (Duncan et al. 1983). Furthermore, SC283 was found to be the most Al-tolerant cultivar in a screening of 391 sorghum genotypes evaluated by Furlani and colleagues (1987). Accordingly SC283 is an outstanding genotype for inheritance studies and for molecular mapping of sorghum Al tolerance. In a study by Furlani and Bastos (1990), the inheritance of Al tolerance in crosses involving SC283 and two Al-sensitive sorghum lines was consistent with the action of a single dominant gene. Accordingly, dominant gene action was clearly observed in this study, since F_1 families derived from any of the susceptible/tolerant (SC283) crosses were associated with phenotypic values several times larger than either of the Al-susceptible parents and closer to the Al-tolerant parent.

The *Alt$_{SB}$* Locus in Sorghum

Previous studies reporting the presence of single major Al-tolerance loci in the long arm of wheat chromosome 4D (*Alt$_{BH}$*, Riede and Anderson 1996) and barley chromosome 4H (Tang et al. 2000), linked to the RFLP marker locus *Xbcd1230*, prompted efforts to use comparative mapping between sorghum and species in the Triticeae tribe to quickly identify Al-tolerance loci in sorghum (Magalhaes et al. 2004). As reported by Furlani and Bastos (1990), bimodal frequency distributions for Al tolerance in a SC283-derived population was also observed when $F_{2:3}$ families derived from a BR007 (highly Al-sensitive line from the Embrapa Maize and Sorghum breeding program) x SC283 cross was evaluated for Al tolerance in nutrient solution set at pH 4.0 and containing 27 μM Al^{3+} activity. In the study by Magalhaes and colleagues, the segregation ratio for Al tolerance and sensitivity conformed to that expected for a single major Al-tolerance locus, which nonetheless behaved in a semidominant fashion. A comparative map was subsequently developed in sorghum, with markers located in the syntenic region where major Al-tolerance loci, which are likely orthologous (Tang et al. 2000), reside in the Triticeae group 4 chromosomes. Although significant macrosynteny was observed, the major Al-tolerance locus in *Sorghum bicolor*, designated *Alt$_{SB}$*, was found to reside elsewhere, mapping to the terminal region of sorghum chromosome 3.

Genetic Diversity for Al Tolerance in Sorghum

Breeding advances depend on the existence of genetic variation that can be identified and manipulated to generate improved cultivars. For example, considering known Al-tolerance donors in barley, there is little potential for Al-tolerance improvement based on nonallelic additive genes, a result of the presence of a single Al-tolerance locus with multiple alleles (Minella and Sorrells 1992). Evaluating the range of genetic diversity controlling Al tolerance in sorghum based on the *Alt$_{SB}$* locus was therefore the key motivation in the Caniato and colleagues (2007) study. This effort included a genetic characterization of *Alt$_{SB}$*-based Al tolerance in sorghum accessions of different origins, which was overlaid with a broader genetic diversity study using SSR (simple sequence repeat) markers to elucidate genetic relationships. Although a major Al-tolerance gene exists in sorghum, similar to what was found in barley, in sorghum there seems to be potential for Al-tolerance improvement by exploiting additive or codominant effects of distinct Al-tolerance loci. This potential was emphasized by the observation that some sources showing variable degrees of Al tolerance did not appear to rely primarily on *Alt$_{SB}$*. Furthermore, a highly Al-tolerant transgressive segregant was in fact detected. However, an allelic series at the *Alt$_{SB}$* locus was also observed using near-isogenic lines (NILs), where *Alt$_{SB}$* alleles from different donors had been introgressed by marker-assisted backcrossing (MABC) into an Al-sensitive line. Multiple

Alt_{SB} alleles were found to encode a wide range of Al-tolerance phenotypes. In summary, the Caniato and colleagues (2007) study indicated that both allelic and non-allelic heterogeneity are important factors for breeding for Al-tolerant sorghum. Although *Alt_{SB}* is likely to be an important player relative to other loci in providing Al tolerance, recombination-based breeding strategies thus emerge as a potentially useful approach to exploiting transgressive segregation in sorghum acid soil breeding programs.

Molecular and Physiological Basis of Al Tolerance Conferred by *Alt_{SB}*

A positional cloning strategy was applied to elucidate the molecular nature of the Al-tolerance gene underlying the *Alt_{SB}* locus in sorghum, which was accompanied by a characterization of the physiological mechanism controlled by the underlying gene (Magalhaes et al. 2007). Data obtained with Al-tolerant (ATF10B) and Al-sensitive (ATF8B) NILs derived from BR007×SC283 provided support for an Al-exclusion mechanism based on Al-activated citrate release from Al-tolerant ATF10B root apices, with a general exclusion zone extending at least 20 mm from the root apex (Magalhaes 2002). Interestingly, both Al tolerance and Al-activated citrate release were found to be Al inducible, significantly increasing over time of exposure to Al (4-6 days; Magalhaes et al. 2007).

High-resolution mapping using different mapping populations derived from BR007×SC283 led to the identification of a 24.6-Kb region that harbored only three open reading frames (ORFs; Magalhaes et al. 2007). Only one of these ORFs, with high sequence similarity to Arabidopsis and rice genes encoding multidrug and toxic compound extrusion (MATE) family proteins (Brown et al. 1999), was highly expressed specifically in root apices of the Al-tolerant NIL, ATF10B. This *Sorghum bicolor* homolog in the MATE family, designated *SbMATE*, is distinct from wheat *ALMT1*, thus confirming that Al tolerance in wheat and sorghum is conferred by non-orthologous loci. Quantitative RT-PCR showed that *SbMATE* is more highly expressed in the first centimeter of the root specifically in the Al-tolerant genotypes and *SbMATE* expression increases over time of exposure to Al^{3+}. This incremental increase in *SbMATE* expression in response to Al correlated closely with the observed increase in root citrate exudation and the Al-tolerance induction, both over time in Al, strongly suggesting that SbMATE is an organic acid transporter that confers Al tolerance via the *Alt_{SB}* locus. Cellular localization studies using SbMATE::GFP fusion proteins and electrophysiological analysis in *Xenopus laevis* oocytes indicated that SbMATE functions as a plasma membrane anion efflux transporter responsible for citrate release into the rhizosphere. Finally, genetic complementation experiments where *SbMATE* was expressed in a highly Al-sensitive Arabidopsis T-DNA knockout mutant in which an Arabidopsis homolog of the wheat *ALMT1* gene is disrupted, demonstrated that expression of *SbMATE* resulted in a significant increase in Al tolerance as well as Al-activated root citrate exudation. These findings indicate that MATE transporter functions in sorghum as an Al-activated citrate transporter that confers Al tolerance via the *Alt_{SB}* locus. Initially characterized as microbial drug transporters, MATE transporters are in fact polyspecific, and different plant MATEs have been shown to transport a range of organic substrates (discussed in Magalhaes 2010).

The Relationship between Population Structure and Al Tolerance in Sorghum

The cloning of *SbMATE* as the major gene responsible for the *Alt_{SB}* locus opened up new, gene-based avenues for molecular breeding strategies aimed at improving Al tolerance in sorghum. Al tolerance can be targeted traditionally by marker-assisted backcross programs to introgress *Alt_{SB}* into cultivars that are sensitive to Al toxicity. However, the identity of *SbMATE* now makes the development of functional

markers possible, and these markers are potentially useful for identifying Al-tolerant accessions in diverse germplasm collections. One clear gain in such a strategy is the identification of materials that may already be adapted to the target environmental conditions. However, a better understanding of the distribution of Al tolerance with regard to patterns of genetic diversity in sorghum are needed to direct such molecular breeding strategies. With this purpose, Caniato and colleagues (2011) assessed population structure and Al tolerance in a diverse cultivated sorghum collection and, as expected, observed a rather low frequency of Al tolerance in sorghum (approximately 5%). From this survey the highly Al-tolerant line, SC566, was identified as a useful Alt_{SB} donor (also, Magalhaes et al. 2004 showed the presence of a functional Alt_{SB} allele in SC566). Genetic analysis based on markers tightly linked to Alt_{SB} and *SbMATE* expression analysis confirmed an important role for Alt_{SB} in providing Al tolerance to most of the Al-tolerant accessions. The fact that the vast majority of the panel was composed of either Al-sensitive (80%) or intermediately tolerant (14%) accessions emphasizes the need for elucidating a possible relationship between genetic divergence and Al tolerance, as a guide for pre-breeding efforts aimed at the identification of novel sources of Al tolerance in sorghum. A population structure analysis revealed clusters that were consistent with both geographical and racial origins as previously described by Deu and colleagues (2006). Interestingly, Al tolerance was not randomly distributed across the species-diversity continuum, being more prevalent in certain genetically differentiated subgroups featuring specific racial and geographical origins. In general, subpopulations containing guinea types from western and southern Africa and, to a lesser extent, caudatum subpopulations are important repositories of Al tolerance in sorghum. These results indicate that efforts toward the identification of novel Al-tolerance sources in sorghum have to be undertaken in light of the species genetic diversity, and that marker-assisted introgression will be needed in the observed cases where Al tolerance does not overlap with population substructure.

The degree of dominance related to Al tolerance was assessed as the ratio between dominance (d) and additive (a) effects based on F_1 hybrids generated by crossing 17 accessions to one common Al-sensitive line, BR007. Although Al tolerance has been attributed to either dominant (Furlani and Bastos 1990) or partially dominant Al-tolerance genes (Magalhaes et al. 2004; Caniato et al. 2011), the results indicated additive gene action in 4 donors, whereas Al tolerance in 11 out of the 17 sorghum accessions was either a recessive or partially recessive trait. Only two accessions including SC283 showed a degree of dominance (d/a) exceeding 0.3 and strict complete dominance was never observed.

Implications for Molecular Breeding Strategies Aimed at Improving Al tolerance in Sorghum

Although Alt_{SB} explains a large portion of the phenotypic variation for Al tolerance in some crosses, the current data supports early reports indicating the presence of additional Al-tolerance genes in the sorghum genome. Furthermore, because highly Al-tolerant transgressive segregants have been detected in progeny derived from SC283, these other genes apparently act additively to Alt_{SB} and may thus be used in recombination-based schemes for improving sorghum Al tolerance. These additional genes are still unknown and may encode completely distinct Al-tolerance mechanisms to Al-activated organic acid release. Another possibility is that these genes may act epistatically to Alt_{SB} and enhance Al tolerance by the same Al-activated citrate release pathway controlled by SbMATE. If so, accessory genes may influence *SbMATE* expression or interact with the transporter protein, changing its permeability properties. With the goal of improving sorghum Al tolerance, accessory loci may offer an opportunity and

concomitantly present difficulties as the flip side of the Al-tolerance breeding equation. Because transgressive segregation can be exploited, elucidating the molecular nature of Alt_{SB}-interacting genes may bring breeding efforts to a new level, resulting in the development of highly Al-tolerant lines. However, genetic background effects can also reduce the efficiency of molecular breeding strategies based solely on Alt_{SB}, leading to incomplete transfer of Al tolerance from donors to the recipient lines. Efforts are now under way to identify and characterize additional Al-tolerance genes so that comprehensive molecular breeding strategies can be formulated and applied to breeding highly Al-tolerant sorghum cultivars.

Maize Al Tolerance

Physiological Mechanisms of Maize Al Tolerance

Root growth inhibition is one of the earliest targets and symptoms of Al toxicity, making relative root growth (i.e., the root growth ratio between +Al/-Al growth conditions) a common and suitable phenotypic criteria to assess Al tolerance. However, phenotypic comparisons should be made with caution, as the experimental conditions employed (e.g., composition of the nutrient solutions and Al^{3+} activities) vary significantly among different studies. Factors such as amelioration of Al toxicity in high-ionic-strength nutrient solutions (Magnavaca et al. 1987) should be taken into account when comparing the Al tolerance of a given genotype relative to that reported in studies where tolerance has been assessed in less physiological, simple-salt solutions. Even so, a general pattern has emerged correlating the degree of maize Al tolerance with lower levels of Al accumulation in the root tips, a result that strongly supports the assumption that Al tolerance in maize is being mediated by an Al-exclusion mechanism. This hypothesized Al-tolerance mechanism is also supported by the observation of rapid Al-activated root citrate exudation, with exudation rates usually being higher in the Al-resistant maize genotypes studied (SA3: Pellet et al. 1995; IAC-TAIUBA: Jorge and Arruda 1997; ATP-Y: Kollmeier et al. 2001; CMS36: Cateto Colômbia and Cateto 100-6:Piñeros et al. 2002; Mariano and Keltjens 2003; Piñeros et al. 2005; Piñeros et al. 2008b). However, in contrast to the strong correlation between Al tolerance and Al-activated root malate release observed in more than 36 different wheat genotypes (Ryan et al. 1995a; Ryan et al. 1995b), the correlation between Al tolerance and Al-activated citrate exudation in maize roots has usually been studied with a single Al-tolerant genotype, comparing it with one or two Al-sensitive lines. In fact, a comparative study using a panel of six genotypes that capture the range of maize Al tolerance and that differ significantly in their genetic background (three Brazilian and three North American genotypes) indicated that the degree of Al tolerance among the genotypes was not entirely correlated with the magnitude of Al-induced citrate release (Piñeros et al. 2005). Although they found a positive correlation between root tip Al-exclusion (based on root Al content) and Al tolerance, the authors of that study also reported a significant lack of correlation between differential Al tolerance and root citrate exudation for the six maize genotypes, with several of the Al-sensitive lines from Brazil and North America also exhibiting significantly high rates of citrate exudation upon exposure to aluminum. Although this study identified citrate as the only organic acid that was released in an Al-activated manner, large constitutive exudation rates of other potential Al-binding ligands (e.g., malate and phosphate) were reported. Consequently, it is quite likely that in contrast to other species like wheat, Al-induced citrate release is only one of several mechanisms operating simultaneously in the very tolerant maize genotypes. Although root citrate efflux plays a significant role, maize Al tolerance appears as a complex quantitatively inherited trait (see next section) with several physiological

mechanisms operating simultaneously. Although further compelling evidence is still required, additional mechanisms could potentially involve exudation of phenolic compounds (Tolra et al. 2009; Kidd et al. 2001) and/or changes in cell-wall pectin content and degree of methylation (Eticha et al. 2005), which have started to receive some attention in the literature.

Genetics of Maize Al Tolerance

As can be inferred from its physiological characteristics, the genetics of Al tolerance in maize is also quite complex. High genetic variability for Al tolerance has been reported in tropical and temperate maize germplasm, using hydroponic-based phenotyping (Rhue and Grogan 1977; Magnavaca 1982; Furlani et al. 1986) and on acid soils with different level of Al saturation (Bahia Filho et al. 1978; Naspolini Filho et al. 1981), as well as in sand culture irrigated with nutrient solution (Garcia Júnior et al. 1979). Although most of the genetic studies agree that Al tolerance is a quantitative trait in maize, divergent conclusions were reached using different genetic materials. Rhue and colleagues (1978) and Garcia Júnior and Silva (1979) reported that Al tolerance is controlled at a single dominant locus, in which the wide variability for this trait in maize would be explained by a multiple allelic series (Rhue et al. 1978) or by modifiers (Garcia Júnior and Silva 1979). However, Al tolerance in F_2 progenies showed continuous and unimodal frequency distributions, typical for a quantitatively inherited trait (Magnavaca 1982; Magnavaca et al. 1987). Brondani and Paiva (1996) described Al tolerance as a quantitative trait but also reported on dominant allele interactions. In addition to confirming the genetic complexity of this trait, other studies have emphasized the contribution of additive gene effects to the total genetic variation in maize Al tolerance (Sawazaki and Furlani 1987; Pandey et al. 1994; Borrero et al. 1995). Nevertheless, dominance effects may contribute to Al tolerance in maize, as revealed by significant mean square values for specific combining ability (SCA) in diallele crosses (Magnavaca et al. 1987; Paterniani and Furlani 2002; Conceição et al. 2009), in agreement with the identification of Al tolerance QTL showing partial dominance effects (Ninamango-Cárdenas et al. 2003).

Mapping of Al Tolerance QTL in Maize

In the first published study on Al-tolerance loci in maize, Moon and colleagues (1997) used somaclonal variation to generate an Al-sensitive mutant from a highly Al-tolerant inbred line, Cateto 100-6. Using a mapping population generated from these parents, two loci (*Alm1*, on the short arm of chromosome 6; and *Alm2*, on the short arm of chromosome 10) contributing to Al tolerance were identified (Sibov et al. 1999). Subsequently, Ninamango-Cardenas and colleagues (2003) mapped Al-tolerance QTL using a population of 168 $F_{3:4}$ families generated from a cross between a highly Al-tolerant inbred commonly used as a tolerance donor in the breeding programs (L1327, currently named Cateto Al237 or Al237), and Al-sensitive inbred line L53. Five QTL were detected on chromosomes 2, 6, and 8 that could explain 60% of the variance in Al tolerance, measured as net seminal root growth in a hydroponic system. For all but one of the QTL, the tolerant allele was donated by the tolerant parent.

Using the F_3 generation of a cross generated from a different set of inbred lines, Conceição and colleagues (2009) mapped five QTL that together explain 41% of the variation in Al tolerance, in this case measured as root regrowth after Al stress. This work detected SSR markers associated with Al tolerance that could be considered as coincident with QTL previously detected (see Table 6 in Conceicao et al. 2009), with exception for the locus detected in chromosome 4 explaining 10% of the phenotypic variation. QTL were detected on chromosomes 5, 6, and 8, on locations equivalent to those described by Ninamango-Cardenas and colleagues (2003), and the QTL on chromosome 10 is an equivalent

location to the *Alm2* QTL described by Sibov and colleagues (1999).

Molecular Biology of Maize Al Tolerance

Few studies have attempted to elucidate the molecular mechanisms underlying maize Al tolerance and, so far, these studies have not been able to shed light on potential alternative tolerance mechanisms (i.e., other than root tip Al exclusion based on root citrate release). A large body of research on other abiotic stresses indicates that stress can affect the expression of numerous genes and that specific changes in gene expression can play a key role in determining the resistance response. On this basis, two large studies of transcriptional profiling in maize roots have been performed, both based on microarrays (Maron et al. 2008; Mattiello et al. 2010). The two studies also used the same maize line, Cateto 100-6, as the Al-tolerant genotype chosen for comparison. In the first study, Maron and colleagues performed a detailed temporal analysis of root gene expression under Al stress. Al altered the expression of significantly more genes in the Al-sensitive line, possibly as a result of more severe toxicity symptoms. Nevertheless, several Al-regulated genes exhibited higher expression in the Al-tolerant line. Many cell wall-related genes were found to be regulated by Al, as well as genes previously shown to respond to low phosphorus, another stress common to acid soils.

In a second study of global changes in gene expression in response to Al, Mattiello and colleagues (2010) looked at the transcriptome of maize roots growing in acid soils containing toxic levels of Al. The genetic materials used in this study were the highly Al-tolerant Cateto 100-6 and the same somaclonal mutant derived from it by Moon and colleagues (1997). Interestingly, several genes previously reported as up-regulated by Al based on hydroponic experiments were also identified in roots grown in acid soil. Previously unidentified genes were also detected, but follow-up physiological studies that could potentially lead to novel Al-tolerance mechanisms are still lacking.

The study by Maron and colleagues (2008) identified two genes encoding members of the MATE family of transporters that showed expression patterns consistent with a potential role in Al tolerance. These were characterized in a subsequent study (Maron et al. 2010). One of these genes, *ZmMATE1*, was mapped to the telomeric region of chromosome 6, co-localizing with a major Al tolerance QTL that explains 16.2% of the phenotypic variance for Al tolerance. This genomic region was previously associated with Al tolerance in two QTL studies (Sibov et al. 1999; Ninamango-Cárdenas et al. 2003). *ZmMATE1* encodes a 563 amino acid protein sharing significant identity to SbMATE and AtMATE. Transient expression of ZmMATE1::GFP fusions in Arabidopsis protoplasts indicated that the protein is localized to the plasma membrane, while [^{14}C]-Citrate efflux studies in oocytes showed that ZmMATE1 is able to mediate citrate transport. In addition, *ZmMATE1* expression in transgenic Arabidopsis conferred a significant increase in Al tolerance and root citrate exudation in response to Al.

Quantitative real-time PCR (qPCR) showed that *ZmMATE1* expression is concentrated in the root tips and is strongly up-regulated by Al, with much higher Al-induced expression in the Al-tolerant genotypes. Expression in the absence of Al is also significantly higher in Al-tolerant Al237 and C100-6 (used in the original microarray study) compared to Al-sensitive L53. *ZmMATE1* expression is up-regulated by Al as early as one hour after exposure. It is interesting to note that gene up-regulation is also observed in L53, although the relative levels of *ZmMATE1* expression in L53 are not very high even after up-regulation by Al. Cloning the cDNA from the parents of the mapping population (Cateto Al237, Al-tolerant and L53, Al-sensitive) revealed only six nucleotide differences in the coding region, of which two resulted in amino acid substitutions. These small differences in protein sequence are unlikely to cause

significant changes to the functional characteristics of the protein. This was confirmed by the fact that *Xenopus* oocytes injected with cRNA made from either allele showed the same transport properties via electrophysiological analysis (Maron et al. in preparation). These results suggest that it is the level of expression of *ZmMATE1* and not functional differences that underlie the Al-tolerance QTL that *ZmMATE1* represents in this population. Moreover, eQTL mapping of *ZmMATE1* indicates that expression is controlled mostly in cis. Recent data suggests that structural variation in the *ZmMATE1* locus is responsible for the expression differential observed between the parental lines (Maron et al. in preparation). However, the expression pattern of *ZmMATE1* in maize NILs differed from the *SbMATE* expression in sorghum NILs. *SbMATE* expression often was lower when different Alt_{SB} alleles were introgressed into a common Al-sensitive line, suggesting the existence of regulatory factors acting in trans on *SbMATE* (Melo et al. 2012).

Molecular Breeding for Al Tolerance in Maize

Al-tolerance studies in maize have a long history, and genetic materials derived from Cateto have been highlighted as Al-tolerant since the early 1980s, both under field conditions (Naspolini Filho et al. 1981) and in nutrient solution (Magnavaca 1982; Furlani et al. 1986; Sawazaki and Furlani 1987). Subsequently, two major Cateto-derived inbred lines, Cateto Al237 and Cateto 100-6 (or C100-6), were selected as sources of Al tolerance for QTL mapping (Sibov et al. 1999; Ninamango-Cardenas et al. 2003), physiological investigations (Moon et al. 1997; Pineros et al. 2005), and molecular studies in response to Al stress (Maron et al. 2008; 2010; Mattiello et al. 2010). Thus, it has been shown that Cateto is clearly an important source of Al tolerance in tropical maize germplasm.

Cateto constitutes a group of landraces originally cultivated by the native peoples living in coastal areas from Argentina to the Guianas. Cateto was widely adopted by the early European immigrants, representing the most widespread maize racial group in South America. Hence Cateto is classified as an ancient commercial maize race because of its indigenous pre-Columbian origin followed by its extensive commercial use as maize varieties or in local hybrid programs (Paterniani and Goodman 1977). Even without a high-yielding performance, Cateto exhibited high combining ability when crossed with different races and considerable adaptation to specific environments (Paterniani and Goodman 1977). Thus, Al tolerance can be considered an important adaptive trait carried by Cateto, which may have contributed to its overall acceptance in Brazil, a region with large areas of acid soils.

Based on our recent findings, one QTL explaining 16% of the variation for Al tolerance in a recombinant inbred line (RIL) population derived from Cateto Al237 was co-localized with a candidate gene, *ZmMATE1*, which encoded the maize root citrate efflux transporter (Maron et al. 2010). More recently, the genetic map for this RIL population was saturated with markers generated via genotyping-by-sequencing (GBS, Elshire et al. 2011), and this same genomic region harboring *ZmMATE1*, named *qALT6*, was able to improve Al tolerance in maize NILs, when transferred to an Al-sensitive line (L53) using marker-assisted backcrossing (Guimarães et al. in preparation). Nevertheless, Cateto has limited use in a modern maize breeding program because of its divergence from improved materials. Despite the large genetic diversity present in maize, this crop species has experienced one of the most intensive breeding efforts of all cultivated crop species, mainly focused on adapted materials, and this has contributed to widening the distance between breeding lines and germplasm bank materials (Nass and Paterniani 2000). Thus, a pre-breeding program is the most promising alternative for linking the introduction of Al-tolerance alleles from native Brazilian races into elite lines of maize. This strategy could be performed using marker-assisted selection

flanking the target region and should also be expanded to other maize breeding programs aiming to improve maize yield on acidic soils throughout Latin America, Africa, and Asia.

As mentioned above, despite a number of studies indicating that Al tolerance in maize is controlled by several loci with additive effects (Magnavaca 1982; Sawazaki and Furlani 1987), nonadditive effects have also been reported for this trait (Magnavaca and Bahia Filho 1995; Conceição et al. 2009). As hybrid development is one of the main products of a maize breeding program, it will be important to evaluate the combining ability of *qALT6* under field conditions, in order to predict its contribution to grain yield and yield stability on acid soils.

Structure-function Analysis of Membrane Transporters Involved in Root Citrate Exudation and Al Tolerance

Prior to the molecular identification of SbMATE1 and ZmMATE1, earlier studies implementing electrophysiological approaches (e.g., patch clamp) had already identified candidate membrane transporters in the plasma membrane of wheat and maize protoplasts (Ryan et al. 1997; Kollmeier et al. 2001; Piñeros and Kochian 2001; Zhang et al. 2001; Piñeros et al. 2002). The activity of these anion transporters was shown to be modulated by extracellular Al^{3+}. The subsequent identification of TaALMT1 (formerly named ALMT1) from wheat (Sasaki et al. 2004) and SbMATE from sorghum (Magalhaes et al. 2007) represents a pivotal breakthrough, as members from two different families of membrane transporter proteins [ALMT (Al-activated malate transporter) and MATE (multidrug and toxic compound efflux)] mediate the Al-activated root organic acid efflux underlying Al tolerance in a number of plant species. It is interesting to note that members of these two families have similar transport functions but quite different structural properties. Our thinking is that the integration of functional and structural research will enable us to identify and target protein residues/motifs underlying MATE and ALMT transport properties critical for citrate or malate transport and Al activation, and consequently in Al tolerance. This structural-functional information will also be used to develop a platform for bioengineering MATE or ALMT proteins, as a novel way to enhance cereal Al tolerance.

Functional studies in *Xenopus* oocytes have demonstrated that TaALMT1, as well as the subsequently identified homologues in Arabidopsis and rape (Hoekenga et al. 2006; Ligaba et al. 2006), encode malate permeable transporters whose activities can be specifically enhanced by the presence of extracellular Al^{3+} (Sasaki et al. 2004; Piñeros et al. 2008a). The remarkable similarities between the functional characteristics of these ALMT transporters and the root organic acid exudation in response to Al strongly indicate that ALMT-type transporters underlie the exudation process characterized at the whole root level. Recently other members of the ALMT family, including two from maize (ZmALMT1 and ZmALMT2) have also been identified and implicated not only in Al-tolerance responses but in a variety of physiological processes such as mineral nutrition, malate homeostasis, and guard cell function (Kovermann et al. 2007; Piñeros et al. 2008b; Meyer et al. 2010; Meyer et al. 2011; Ligaba et al. 2012). The molecular and functional characterization of ZmALMT1 and ZmALMT2, particularly their expression patterns and the lack of transport enhancement upon exposure to Al, suggest that they are likely to be involved in mediating other mineral nutrition and ion homeostasis processes, rather than mediating Al-enhanced transport responses in maize.

SbMATE is a major sorghum Al-tolerance gene encoding a plasma membrane transporter from a different family of proteins, namely the multidrug and toxin extrusion (MATE) family of transporters. The MATE family is one of five major multidrug resistance (MDR) transporter families, and although they are widely distributed across all kingdoms of living organisms,

their transport properties and functional roles remain largely unknown (Omote et al. 2006; Moriyama et al. 2008). In contrast to the low number of MATE genes per species found in other kingdoms, there are a large number of plant MATE genes (40 to 60 family members in the Arabidopsis, rice, and *Medicago truncatula* genomes) suggesting a wide variety of different biological roles *in planta* (reviewed by Yazaki 2005). Although the transport properties of most plant MATEs remain unknown, recent studies have begun to characterize several plant MATEs, suggesting diverse physiological roles such as the transport of cationic flavonoid, xenobiotic, and alkaloid substrates (Debeaujon et al. 2001; Diener et al. 2001; Li et al. 2002; Otani et al. 2005; Marinova et al. 2007; Morita et al. 2009). Since the discovery of SbMATE, orthologues involved in mediating organic acid efflux in response to Al stress have also been identified in barley, maize, rice bean, rice, and Arabidopsis (Furukawa et al. 2007; Liu et al. 2009; Maron et al. 2010; Yokosho et al. 2011). Additional plant MATE transporters permeable to citrate have also been identified that appear to be involved in Fe translocation in the xylem (as an Fe-citrate complex) and not Al tolerance (Durrett et al. 2007; Yokosho et al. 2009).

As shown in Figure 6.1, phylogenetic analysis of the plant MATEs that have been functionally characterized to date reveals a unique subgroup of functionally characterized plant MATEs that transport citrate (highlighted in red in Figure 6.1b). These transporters include those that have been shown to be involved in sorghum, maize, barley, rye, rice bean, Arabidopsis, and rice Al tolerance, as well as rice and Arabidopsis MATEs localized to the root pericycle and involved in the loading of citrate into the xylem for Fe translocation to the shoot. Structurally, comparison of the amino acid sequences of all of the citrate-permeable plant MATEs in Figure 6.1b indicates that these proteins are members of the NorM-like subset (COG0534) of the MatE superfamily (Pfam01554), sharing a

Fig. 6.1. Phylogenetic analysis of MATE-type transporters for all plant MATE transporters that have been functionally characterized to date. Plant members colored in red represent MATE's that have been shown to mediate citrate transport. The asterisks indicate members mediating citrate release in response to aluminum stress. The tree was built using protein sequences with Geneious Tree Builder software. For a color version of this figure, please refer to the color plate.

common predicted secondary structure consisting of about 500 to 700 amino acids, containing 12 transmembrane helices with a long (~100 residues) cytoplasmic N-tail, and a distinctive long cytoplasmic loop between the second and third transmembrane domain. The presence of the characteristic long cytoplasmic N-tails suggests that these MATEs may interact with other proteins (Moriyama et al. 2008).

Functionally, results from electrophysiological and ^{14}C efflux studies have established that when expressed in *X.* oocytes, the subgroup of SbMATE-like plant transporters mediates H^+ (and possibly Na^+)-coupled citrate efflux. These transport characteristics present intriguing questions regarding substrate recognition, energy coupling, and transport mechanism, given that most other experimentally characterized MATE transporters across all kingdoms show preference for organic cation substrates (see Table I in Omote et al. 2006). Furthermore, while expression of any of the members of this subgroup of SbMATE-like transporters in heterologous systems has resulted in a constitutive electrogenic

transport, their transport is highly sensitive to extracellular Al (except for the barley homologue), being highly inhibited at low micromolar concentrations. This functional property is quite different from the direct Al-enhancement for members of the ALMT family and for the Al-activation of root citrate efflux seen *in planta*. Preliminary findings from our group indicate there are other proteins that may interact with SbMATE to facilitate the Al activation, as well as other post translational modifications of the SbMATE-type transporters that are involved in the regulation of the transporter *in planta.* These preliminary findings will set the stage for basic research on the structure-function of these transporters, quite likely expanding our toolbox for molecular breeding of enhanced cereal Al tolerance.

Conclusions

A major point that we hope we have made in this chapter is that agricultural research has advanced to the point where findings from basic molecular, genomic, and genetic investigations of crop plant traits are now being translated for use in crop improvement programs. In this example, basic research has allowed us to identify physiological mechanisms and the associated genes that confer enhanced Al tolerance in sorghum and maize. We also are beginning to understand the role of genetic diversity and population structure in Al tolerance. This information is now being used to facilitate the effective molecular breeding of improved Al tolerance in both sorghum and maize, in order to improve sorghum and maize yields on acid soils that are prevalent in many countries in the tropics and subtropics.

Acknowledgements

The research described in this chapter was supported by Generation Challenge Programme grants G3007.04, G3008.02, G7010.03.02, G7010.03.05, and G7010.03.03.

References

Bahia Filho AFC., França, GE, Pitta GVE, Magnavaca R, Mendes JF, Bahia FGFTC, Pereira P. 1978. Avaliação de linhagens e populações de milho em condições de elevada acidez. Anais da XI Reunião Brasiileira de Milho e Sorgo, Piracicaba, SP, pp 51–58.

Borgonovi RA, Schaffert RE, Pitta GVE, Magnavaca R, Alves VMC. 1987a. Aluminum tolerance in sorghum. In: Gabelman HW, Loughman BC, eds, *Genetic aspects of plant mineral nutrition*. Dordrecht, Boston, and Lancaster: Martinus Nijhoff Publishers, pp 213–221.

Borgonovi RA, Schaffert RE, Pitta GVE. 1987b. Breeding aluminum-tolerant sorghums. In: Gourley LM, Salinas JG, eds, *Soghum for acid soils, Proceedings of a workshop on evaluating sorghum for tolerance to Al-toxic tropic soils in Latin America*. INTSORMIL – ICRISAT – CIAT, Cali, Colombia, pp 271–292.

Borrero J C, Pandey S, Ceballos H, Magnavaca R, Bahia AFC. 1995. Genetic variances for tolerance to soil acidity in a tropical maize population. *Maydica* 40:283–288.

Boye-Goni SR, Marcarian V. 1985. Diallel analysis of aluminum tolerance in selected lines of grain sorghum. *Crop Science* 25:749–752.

Brown MH, Paulsen IT, Skurray RA.1999. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Molecular Microbiology* 31:393–395.

Caniato FF, Guimarães CT, Schaffert RE, Alves VMC, Kochian LV, Borém A, Klein PE, Magalhaes JV. 2007. Genetic diversity for aluminum tolerance in sorghum. *Theoretical and Applied Genetics* 114:863–876.

Caniato FF, Guimarães CT, Hamblin M, Billot C, Rami J-F, Hufnagel B, Kochian, LV, Liu J, Garcia AAF, Hash CT, Ramu P, Mitchell S, Kresovich S, Oliveira AC, Avellar G, Borém A, Glaszmann J-C, Schaffert RE, Magalhaes JV. 2011. The relationship between population structure and aluminum tolerance in cultivated sorghum. *PloS One* 6:e20830.

Conceição LDHCS, Tessele C, Barbosa Neto JF. 2009. Diallel analysis and mapping of aluminum tolerance in corn inbred lines. *Maydica* 54:51–64.

Debeaujon I, Peeters AJM, Leon-Kloosterziel KM, Koornneef M. 2001. The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *The Plant Cell* 13:853–871.

Delhaize E, Craig S, Beaton CD, Bennet RJ, Jagadish VC, Randall PJ. 1993a. Aluminum tolerance in wheat *.Triticum asetivum* L. I. Uptake and distribution of aluminum in root apices. *Plant Physiology* 103:685–693.

Delhaize E, Ryan PR, Randall PJ. 1993b. Aluminum tolerance in wheat *.Triticum aestivum* L. II. Aluminum-stimulated excretion of malic acid from root apices 5379. *Plant Physiology* 103:695–702.

Deu M, Rattunde F, Chantereau J. 2006. A global view of genetic diversity in cultivated sorghums using a core collection. *Genome* 49:168–180.

Diener AC, Gaxiola RA, Fink GR. 2001. Arabidopsis ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. *The Plant Cell* 13:1625–1638.

Duncan RR, Clark RB, Furlani PR. 1983. Laboratory and field evaluations of sorghum for response to aluminum and acid soil. *Agronomy Journal* 75:1023–1026.

Duncan RR. 1988. Sequential development of acid soil tolerant sorghum genotypes under field stress conditions. *Communications in Soil Science and Plant Analysis* 19:1295–1305.

Durrett TP, Gassmann W, Rogers EE. 2007. The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. *Plant Physiology* 144:197–205.

Elshire, RJ et al. 2011. A robust, simple genotyping-by-sequencing .GBS. approach for high diversity species. *PLoS One* 6:e19379 .2011.

Eticha D, Stass A, Horst WJ. 2005. Cell-wall pectin and its degree of methylation in the maize root-apex: Significance for genotypic differences in aluminium resistance. *Plant Cell and Environment* 28:1410–1420.

Flores CI, Gourley LM, Pedersen JF, Clark RB. 1991. Inheritance of acid-soil tolerance in sorghum .Sorghum bicolor. grown on an ultisol. In: RJ Wright, VC Baligar, RP Murrman, eds, *Plant-Soil Interactions at Low pH*. Dordrecht: Kluwer Academic Publishers, pp 1081–1093.

Foy CD, Duncan RR, Waskom RM, Miller DR. 1993. Tolerance of sorghum genotypes to an acid, aluminum toxic tatum subsoil. *Journal of Plant Nutrition* 16:97–127.

Furlani PR, Lima M, Miranda LT De, Miranda LEC De, Sawazaki E, Magnavaca R. 1986. Avaliação de linhagens, materiais comerciais e duas populações de milho para tolerância ao alumínio. *Pesquisa Agropecuária Brasileira* 21(6):655–660.

Furlani PR, Bastos CR, Borgonovi RA, Schaffert RE.1987. Resposta diferencial de genótipos de sorgo para tolerância ao alumínio em solução nutritiva. *Pesquisa Agropecuária Brasileira* 22:323–330.

Furlani PR, Bastos CR. 1990. Genetic control of aluminum tolerance in sorghum. In: N El Bassam, M Dambroth, BC Loughman, eds, *Genetic Aspects of Plant Mineral Nutrition*. Dordrecht: Kluwer Academic Publishers, pp 215–219.

Furukawa J, Yamaji N, Wang H, Mitani N, Murata Y, Sato K, Katsuhara M, Takeda K, Ma JF. 2007. An aluminum-activated citrate transporter in barley. *Plant and Cell Physiology* 48:1081–1091.

Garcia Júnior O, Silva WJ. 1979. Análise genética da tolerância ao alumínio em milho. *Ciênc. Cult.* 31:585.

Garcia Júnior O, Silva WJ, Massei MAS. 1979. An efficient method for screening maize inbreds for Al tolerance. *Maydica* 24:75–82.

Garvin DF, Carver BF. 2003. Role of the genotype in tolerance to acidity and aluminum toxicity. In: Rengel Z, ed, *Handbook of Soil Acidity*. New York and Basel: Marcel Dekker, Inc., pp 387–406.

Gourley LM, Rogers SA, Ruiz-Gomez C, Clark RB. 1990. Genetic aspects of aluminum tolerance in sorghum. *Plant and Soil* 123:211–216.

Hoekenga OA, Maron LG, Piñeros MA, Cancado GMA, Shaff J, Kobayashi Y, Ryan PR, Dong B, Delhaize E, Sasaki T, Matsumoto H, Yamamoto Y, Koyama H, Kochian LV. 2006. AtALMT1, which encodes a malate transporter, is identified as one of several genes critical for aluminum tolerance in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 103:9738–9743.

Jorge RA, Arruda P. 1997. Aluminum-induced organic acids exudation by roots of an aluminum-tolerant tropical maize. *Phytochemistry* 45:675–681.

Kidd PS, Llugany M, Poschenrieder C, Gunse B, Barcelo J. 2001. The role of root exudates in aluminium resistance and silicon-induced amelioration of aluminium toxicity in three varieties of maize .Zea mays L... *Journal of Experimental Botany* 52:1339–1352.

Kochian LV, Hoekenga OA, Piñeros MA. 2004. How do crop plants tolerate acid soils? - Mechanisms of aluminum tolerance and phosphorous efficiency. *Annual Review of Plant Biology* 55:459–493.

Kollmeier M, Dietrich P, Bauer CS, Horst WJ, Hedrich R. 2001. Aluminum activates a citrate-permeable anion channel in the aluminum-sensitive zone of the maize root apex. A comparison between an aluminum-sensitive and an aluminum-resistant cultivar. *Plant Physiology* 126:397–410.

Kovermann P, Meyer S, Hörtensteiner S, Picco C, Scholz-Starke J, Ravera S, Lee Y, and Martinoia E. 2007. The Arabidopsis vacuolar malate channel is a member of the ALMT family. *The Plant Journal* 52:1169–1180.

Krill AM, Kirst M, Kochian LV, Buckler ES, Hoekenga OA. 2010. Association and linkage analysis of aluminum tolerance genes in maize. *Plos One* 5:1–11.

Li L, He Z, Pandey GK, Tsuchiya T, Luan S. 2002. Functional Cloning and Characterization of a Plant Efflux Carrier for Multidrug and Heavy Metal Detoxification. *Journal of Biological Chemistry* 277:5360–5368.

Ligaba A, Katsuhara M, Ryan PR, Shibasaka M, Matsumoto M. 2006. The BnALMT1 and BnALMT2 Genes From Brassica napus L. Encode Aluminum-Activated Malate Transporters that Enhance the Aluminum Resistance of Plant Cells. *Plant Physiology* 142:1294–1303.

Ligaba A, Maron L, Shaff J, Kochian L, Piñeros M. 2012. Maize ZmALMT2 is a root anion transporter that mediates constitutive root malate efflux. *Plant, Cell & Environment* 35:1185–1200.

Liu J, Magalhaes JV, Shaff J, Kochian LV. 2009. Aluminum-activated citrate and malate transporters from the MATE and ALMT families function independently to confer Arabidopsis aluminum tolerance. *The Plant Journal* 57:389–399.

Lopes MA, Magnavaca R, Bahia Filho AFC, Gama EEG. 1987. Avaliação de populações de milho e seus cruzamentos para tolerância à toxidez de alumínio em solução nutritiva. *Pesquisa Agropecuária Brasileira*, 22:257–263.

Magalhaes JV. 2002. Molecular genetic and physiological investigations of aluminum tolerance in sorghum .*Sorghum bicolor* L. Moench. PhD dissertation, Cornell University.

Magalhaes JV, Garvin DF, Wang Y, Sorrells ME, Klein PE, Schaffert RE, Li L, Kochian LV. 2004. Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. *Genetics* 167:1905–1914.

Magalhaes JV, Liu J, Guimaraes CT, Lana UGP, Alves VMC, Wang YH, Schaffert RE, Hoekenga OA, Pineros MA, Shaff JE, Klein PE, Carneiro NP, Coelho CM, Trick HN, Kochian LV. 2007. A gene in the multidrug and toxic compound extrusion .MATE. family confers aluminum tolerance in sorghum. *Nature Genetics* 39:1156–1161.

Magalhaes JV. 2010. How a microbial drug transporter became essential for crop cultivation on acid soils: Aluminium tolerance conferred by the multidrug and toxic compound extrusion .MATE. family. *Annals of Botany* 106:199–203.

Magnavaca R. 1982. Genetic variability and the inheritance of aluminum tolerance in maize .Zea mays L... PhD dissertation, University of Nebraska - Lincoln.

Magnavaca R, Gardner C, Clark. 1987. Evaluation of inbred maize lines for aluminum tolerance in nutrient solution. In: Gabelman HLB, ed., *Genetic Aspects of Plant Mineral Nutrition*. Dordrecht: Martinus Nijhoff, pp 255–265.

Magnavaca R, Bahia Filho AFC. 1995. Seleção de milho para tolerância ao alumínio. In: *Simpósio Internacional sobre Estresse Ambiental. Anais: O milho em perspectiva*. Mexico, CIMMYT, pp 369–394.

Mariano ED, Keltjens WG. 2003. Evaluating the role of root citrate exudation as a mechanism of aluminium resistance in maize genotypes. *Plant and Soil* 256: 469–479.

Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, Routaboul JM, Debeaujon I, Klein M. 2007. The Arabidopsis MATE Transporter TT12 Acts as a Vacuolar Flavonoid/H+-Antiporter Active in Proanthocyanidin-Accumulating Cells of the Seed Coat. *The Plant Cell* 19:2023–2038.

Maron LG, Kirst M, Mao C, Milner MJ, Menossi M, Kochian LV. 2008. Transcriptional profiling of aluminum toxicity and tolerance responses in maize roots. *New Phytologist* 179:116–128.

Maron L, Piñeros MA, Guimarães C, Magalhaes J, Pleiman J, Mao C, Shaff J, Belicuas SNJ, Kochian L. 2010. Two functionally distinct members of the MATE .multidrug and toxic compound extrusion. family of transporters potentially underlie two major Al tolerance QTL in maize. *The Plant Journal* 61:728–740.

Mattiello L, Kirst M, Silva FR, Jorge RA, Menossi M. 2010. Transcriptional profile of maize roots under acid soil growth. *BMC Plant Biology* 10:196.

Melo JO, Lana UGP, Pineros M, Alves VMC, Guimarães CT, Liu J, Zheng Y, Zhong S, Fei Z, Maron LG, Schaffert RE, Kochian LV, Magalhaes JV. 2012. Incomplete transfer of accessory loci influencing SbMATE expression underlies genetic background effects for aluminum tolerance in sorghum. *The Plant Journal*, online.

Meyer S, Mumm P, Imes D, Endler A, Weder B, Al-Rasheid KAS, Geiger D, Marten I, Martinoia E, Hedrich R. 2010. AtALMT12 represents an R-type anion channel required for stomatal movement in Arabidopsis guard cells. *The Plant Journal* 63:1054–1062.

Meyer S, Scholz-Starke J, De Angeli A, Kovermann P, Burla B, Gambale F, Martinoia E. 2011. Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. *The Plant Journal* 67:247–257.

Minella E, Sorrells ME. 1992. Aluminum tolerance in barley: Genetic relationships among genotypes of diverse origin. *Crop Science* 32:593–598.

Miranda LT De, Furlani PR, Miranda LEC De, Sawazaki E. Genetics of environmental resistance and super genes: Latente aluminum tolerance. 1984. *Maize Genetics Cooperation News Letter* 58:46–48.

Moon D, Ottoboni LM, Souza AP, Sibov S, Gaspar M, Arruda P. 1997. Somaclonal-variation-induced aluminum-sensitive mutant from an aluminum-inbred maize tolerant line. *Plant Cell Reports* 16:686–691.

Morita M, Shitan N, Sawada K, Van Montagu MCE, Inze D, Rischer H, Goossens A, Oksman-Caldentey KM, Moriyama Y, Yazaki K. 2009. Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion .MATE. transporter in Nicotiana tabacum. *Proceedings of the National Academy of Sciences* 106:2447–2452.

Moriyama Y, Hiasa M, Matsumoto T, Omote H. 2008. Multidrug and toxic compound extrusion MATE.-type proteins as anchor transporters for the excretion of metabolic waste products and xenobiotics. *Xenobiotica* 38:1107–1118.

Nass LL, Paterniani E. 2000. Pre-breeding: A link between genetic resources and maize breeding. *Scientia Agricola* 57.3:581–587.

Ninamango-Cárdenas F, Guimarães CT, Martins PR, Parentoni SN, Carneiro NP, Lopes MA, Moro JR, Paiva E. Mapping QTLs for aluminum tolerance in maize. 2003. *Euphytica* 130:223–232.

Naspolini Filho W, Bahia Filho AFC, Vianna RT, Gama EEG, Vasconcellos CA, Magnavaca R. 1981. Comportamento de linhagens e híbridos simples de milho (*Zea mays* L.) em solo sob vegetação de cerrado. *Ciência e Cultura* 33(5):722–727.

Omote H, Hiasa M, Matsumoto T, Otsuka M, Moriyama Y. 2006. The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends in Pharmacological Sciences* 27:587–593.

Otani M, Shitan N, Sakai K, Martinoia E, Sato F, Yazaki K. 2005. Characterization of vacuolar transport of the

endogenous alkaloid Berberine in *Coptis japonica*. *Plant Physiology* 138:1939–1946.

Pandey S, Ceballos H, Magnavaca R, Bahia AFC, Duque-vargas J, Vinasco LE. 1994. Genetics of tolerance to soil acidity in tropical maize. *Crop Science* 34:1511–1514.

Paterniani MEAGZ, Furlani PR. 2002. Tolerância à toxicidade de alumínio de linhagens e híbridos de milho em solução nutritiva. *Bragantia* 61(1):11–16.

Paterniani E, Goodman MM 1977. Races of maize in Brazil and adjacent areas. 95 pp. Centro Internacional de Mejoramiento de Maiz y Trigo. Texcoco, Mexico.

Pellet DM, Grunes DL, Kochian LV. 1995. Organic-acid exudation as an aluminum-tolerance mechanism in maize (*Zea mays* L.). *Planta* 196:788–795.

Piñeros MA, Cancado GMA, Kochian LV. 2008a. Novel properties of the wheat aluminum tolerance organic acid transporter .TaALMT1. revealed by electrophysiological characterization in Xenopus oocytes: Functional and structural implications. *Plant Physiology* 147:2131–2146.

Piñeros MA, Cançado GMA, Maron LG, Lyi SM, Menossi M, Kochian LV. 2008b. Not all ALMT1-type transporters mediate aluminum-activated organic acid responses: The case of *ZmALMT1* – an anion selective transporter. *The Plant Journal* 53:352–367.

Piñeros MA, Kochian LV. 2001. A patch-clamp study on the physiology of aluminum toxicity and aluminum tolerance in maize: Identification and characterization of Al3+-induced anion channels. *Plant Physiology* 125:292–305.

Piñeros MA, Magalhaes JV, Carvalho AVM, Kochian LV. 2002. The physiology and biophysics of an aluminum tolerance mechanism based on root citrate exudation in maize. *Plant Physiology* 129:1194–1206.

Piñeros MA, Shaff JE, Manslank HS, Alves VMC, Kochian LV. 2005. Aluminum resistance in maize cannot be solely explained by root organic acid exudation. A comparative physiological study. *Plant Physiology* 137:231–241.

Rhue RD, Grogan CO. 1977. Screening corn for Al tolerance using different Ca and Mg concentrations. *Agronomy Journal* 69:755–760.

Rhue RD, Grogan CO, Stockmeyer EW, Everett HL. 1978. Genetic control of Al tolerance in corn. *Crop Science* 18:1063–1067.

Riede, CR, Anderson JA. 1996. Linkage of RFLP markers to an aluminum tolerance gene in wheat. *Crop Science* 36:905–909.

Ryan PR, Delhaize E, Randall PJ. 1995a. Characterization of Al-stimulated efflux of malate from the apices of Al-tolerant wheat roots. *Planta* 196:103–110.

Ryan PR, Delhaize E, Randall PJ. 1995b. Malate efflux from root apices and tolerance to aluminum are highly correlated in wheat. *Australian Journal of Plant Physiology* 22:531–536.

Ryan PR, Skerrett M, Findlay GP, Delhaize E, Tyerman SD. 1997. Aluminum activates an anion channel in the apical cells of wheat roots. *Proceedings of the National Academy of Sciences* 94:6547–6552.

Sasaki T, Yamamoto Y, Ezaki B, Katsuhara M, Ahn SJ, Ryan PR, Delhaize E, Matsumoto H. 2004. A wheat gene encoding an aluminum-activated malate transporter. *The Plant Journal* 37:645–653.

Sawazaki E, Furlani PR. Tolerância ao alumínio em milho Cateto. 1987. *Bragantia*, 46.2:269–278.

Sibov ST, Gaspar MJ, Ottoboni LMM, Arruda P, Souza AP. 1999. Two genes controlling aluminum tolerance in maize: Genetic and molecular mapping analyses. *Genome* 42: 475–482.

Tang Y, Sorrells ME, Kochian LV, Garvin DF. 2000. Identification of RFLP markers linked to the barley aluminum tolerance gene *Alp*. *Crop Science* 40:778–782.

Tolra R, Barcelo J, Poschenrieder C. 2009. Constitutive and aluminium-induced patterns of phenolic compounds in two maize varieties differing in aluminium tolerance. *Journal of Inorganic Biochemistry* 103:1486–1490.

von Uexküll HR, Mutert E. 1995. Global extent, development and economic-impact of acid soils. *Plant and Soil* 171:1–15.

Wood S, Sebastian K, Scherr SJ. 2000. Soil resource condition. In: *Pilot analysis of global ecosystems: Agroecosystems*. Washington: International Food Policy Research Institute and the World Resources Institute, pp 45–54.

Yazaki K. 2005. Transporters of secondary metabolites. *Current Opinion in Plant Biology* 8:301–307.

Yokosho K, Yamaji N, Ueno D, Mitani N, Ma JF. 2009. OsFRDL1 is a citrate transporter required for efficient translocation of iron in rice. *Plant Physiology* 149:297–305.

Yokosho K, Yamaji N, Ma JF. 2010. Isolation and characterisation of two MATE genes in rye. *Functional Plant Biology* 37:296–303.

Yokosho K, Yamaji N, Ma JF. 2011. An Al-inducible MATE gene is involved in external detoxification of Al in rice. *The Plant Journal* 68:1061–1069.

Zhang W-H, Ryan PR, Tyerman SD. 2001. Malate-permeable channels and cation channels activated by aluminium in the apical cells of wheat roots. *Plant Physiology* 125:1459–1472.

Chapter 7

Freezing Tolerance in the Triticeae

Galiba Gabor, Eric J. Stockinger, Enrico Francia, Justyna Milc, Gabor Kocsy, and Nicola Pecchioni

Abstract

In the face of growing food demand, major variation in cereal yield from year to year resulting from drought and low temperatures poses a serious problem for agricultural production and for the stability of agricultural trade. Furthermore, in the coming decades climate change is predicted to cause dramatic shifts in temperature and rainfall, posing further challenges to agriculture. Accordingly, maintaining high and stable yields in the driest and coldest cereal-growing areas is becoming one of the most urgent aims for plant breeding. From a biological point of view, the strategy of escape relies on developmental plasticity (Levitt 1980), that is, successful completion of reproduction before the onset of severe stress such as drought and frost. However, escape is probably not the best strategy for maximizing agricultural productivity, or at least, it should not be the only route pursued. In fact, given the limited area of cultivatable land, a tendency to perennialism coupled with an increasing tolerance to abiotic stresses over the growth cycle could be for herbaceous crops a significant genetic alternative to the agronomic one of double-cropping. As an example, because of their longer growing period, winter cereals usually have higher yield potential than spring varieties, which are planted later in the spring. In order to withstand stresses, a plant tolerant to freezing — just as one tolerant to drought — has to cope with dehydration. Ice formation in the intercellular space subtracts water from the cell; at –10° C more than 90% of the osmotically active water typically moves out of the cells. At the same time, plasma membranes, as well as the organellar ones, are damaged and can lose integrity. Survival at freezing temperatures is a phenomenon dependent on a number of factors, including frost duration and severity, alternation of frost and thaw periods, and synthesis of toxic substances affecting recovery capacity. However, the ability to survive freezing is based on the effectiveness of the cold acclimation process (Thomashow 1999). Cold acclimation is a relatively slow, adaptive response that takes place during the fall, when the temperature, length of the day, and light intensity decrease gradually. All these factors are important for attaining a genetically determined freezing tolerance. The physiological process comprises biochemical changes that enable tissues to enhance their frost resistance. Plant species that do not acclimate, generally those of tropical origin, are severely damaged even at temperatures just below 0° C. For this reason, studies aimed at understanding the molecular

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

mechanisms that protect the plant cell from frost events were mostly focused on the acclimation process. Species of the Triticeae tribe of the Poaceae, such as wheat and barley, able to acclimate to and to tolerate frost, are one of the best models for studying freezing tolerance in herbaceous, non-woody plants. The present chapter reviews in detail the genetic and genomic knowledge accumulated over the last twenty years in these model species, in terms of genetic loci and sequence variation able to confer higher tolerance to frost. Lastly, the use of genetic resources, as well as new genomic tools for producing freezing tolerant varieties, is discussed.

Major Determinants of Frost Tolerance in the Triticeae (QTLs and Genes)

In temperate cereals the ability of fall-sown varieties to survive winter depends on a complex trait referred to as winter hardiness. In turn, this is composed of three different but interconnected components: vernalization requirement, photoperiod sensitivity, and frost tolerance (Galiba et al. 2009). The latter can be defined as the ability of plants to survive freezing temperatures, prevent damage to tissues, and minimize the negative effects of freezing on eventual yield. During the last 20 years, several field and growth chamber screening methods have been proposed to study and assist breeding for frost-tolerant varieties. According to Prášil and colleagues (2007) such methods can be classified as direct (monitoring plant survival, LT_{50}, i.e., at a temperature lethal for 50% of the plants, and chlorophyll fluorescence through the *Fv/Fm* ratio) or indirect (selection of molecular markers correlated with the trait). In barley, Tòth and colleagues (2004) and Rapacz and colleagues (2010) developed PCR-based markers for assisted selection, and Rapacz and colleagues (2008) proposed markers based on *COR* (COld-Regulated) gene expression. Many direct assays have also been integrated with molecular markers, to complement the frost tolerance evaluation and improve the prediction of winter hardiness (Rizza et al. 2011). On the other hand, conventional breeding strategies based on winter survival possibly coupled to visual rating of damage in the past have been rather inefficient in improving frost resistance in winter cereals (Limin and Fowler 1991). Figure 7.1 is an example from barley. In samples of winter and spring germplasm pools, there was basically no consistently significant improvement in frost tolerance during the progression from ancient local landraces, to old cultivars, then to modern ones (released after 1980).

This slow or non-existent improvement can be partly attributed to the fluctuating nature of winter injuries, which would not allow constant selection for tolerance across breeding generations. The minimum temperature that Triticeae members are able to survive is typically in the range of –10 to –20° C. However, although cold acclimation and frost tolerance are still considered complex polygenic traits, the ever-expanding accumulation of "omics" data is leading to an increased understanding of the mechanisms that plants have evolved in order to tolerate this stress condition (Pecchioni et al. 2012; Pecchioni et al. 2013).

QTL and Genes Responsible for Vegetative Frost Tolerance

In temperate cereal growing areas, frost is normally experienced during vegetative development. In crops such as wheat and barley, linkage mapping has been the method of choice for genetic studies, starting from experimental bi-parental crosses. In the wake of vertebrate genetics, association mapping has more recently begun to be adopted in plants. This involves searching for genotype-phenotype correlations in collections of individuals in which relatedness is unknown. However, as emphasized by Myles

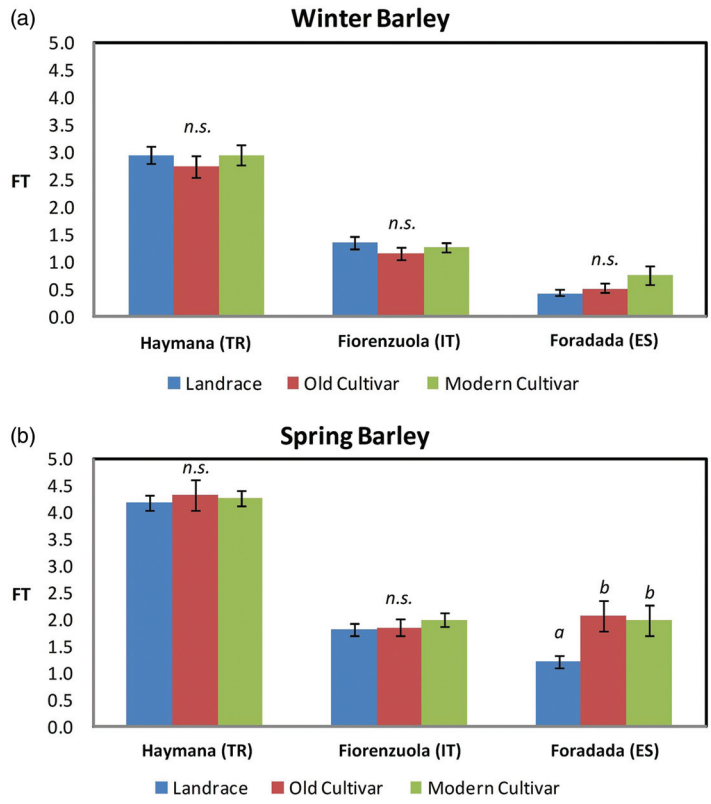


Fig. 7.1. Average values of frost tolerance, FT, expressed as visual score ranging from 0 = very tolerant (no visible frost injury) to 5 = very susceptible (>80% plants killed), for (a) 91 winter, and (b) 59 spring barleys. Genotypes are divided into ancient local landraces, old cultivars (released before 1980), and modern cultivars (released after 1980). Data were recorded in winter 2004/2005 from field trials in three locations in the Mediterranean basin. Vertical bars represent standard error, while letters indicate the result of multiple comparison tests (LSD and Tukey's Procedure, $p < 0.05$). For a color version of this figure, please refer to the color plate.

and colleagues (2009), the two approaches are more similar than is often assumed and should be regarded as complementary rather than alternative methods for identifying the genes underlying phenotypic variation. The identification of winter hardiness/frost tolerance QTLs (quantitative trait loci) in Triticeae genomes was thus initially based mainly on mapping populations of recombinant inbred lines in wheat and of doubled haploid lines in barley. The first reports identified a genomic region of the long arm of homoeologous chromosome group 5 spanning about 50 cM as containing loci for major effects on vernalization requirement (*VRN1*) and frost tolerance (then defined as *FR-1*). In hexaploid wheat, chromosome 5A was identified as responsible for most variation in the two traits (Sutka and Snape 1989; Hayes et al. 1993; Pan et al. 1994; Galiba et al. 1995; Sutka et al. 1999). About ten years later, improved statistical methods for QTL detection, as well as new mapping populations, allowed the frost tolerance effects to be resolved into two loci (*Frost Resistance-1*, *FR-1*, co-segregating with *VRN1*, and *Frost Resistance-2*, *FR-2*), approximately 30 cM apart (Francia et al. 2004; Skinner et al. 2006). It has also been shown that *FR-1* and *FR-2* on chromosome group 5 are present in wheat and barley

(Galiba et al. 2009; Tondelli et al. 2011). Interestingly, the barley 'Nure' x 'Tremois' genetic system developed by Francia and colleagues (2004) remains the only bi-parental population in the Triticeae where both *FR*- QTL are segregating, since in wheat *FR-A1* (Galiba et al. 1995) and *FR-A2* (Vágújfalvi et al. 2003) were individually mapped. The best candidate for *FR-1* is currently a MADS box gene coding for a protein similar to the meristem identity AP1 transcription factor in Arabidopsis. The gene is known as *TaAP1* in wheat and *HvBM5A* in barley, it is up-regulated (directly or indirectly) by vernalization and by long days (for a review see Galiba et al. 2009), and was proved to be the determinant for *VRN-1,* the major locus involved in the vegetative-to-reproductive transition (Danyluk et al. 2003; Trevaskis et al. 2003; Yan et al. 2003; von Zitzewitz et al. 2005). Recent evidence suggests that coincident *VRN-1* and *FR-1* QTL are the pleiotropic effects of this same gene, and that *VRN-1* allelic variation influences the expression duration of low temperature-induced genes (Dhillon et al. 2010). In particular, mutations in the *VRN-1* promoter, resulting in high *VRN-1* transcript levels under both long and short days, dampen the expression of *COR* genes, and lowers resistance, especially under long-day conditions (Galiba et al. 2009; Trevaskis 2010). Further support for *VRN-1* being a determinant of freezing tolerance derives from the work by Limin and colleagues (2007). These authors demonstrated in barley cultivars used as parents in mapping populations that timing of maximum frost tolerance is usually coincident with the timing of vernalization saturation. As such, the hypothesis of pleiotropy would explain the old general observation of breeders that winter-type genotypes, in wheat and in barley, carrying a vernalization sensitive ("winter") allele at *VRN-1*, are more tolerant than spring-type cultivars. However, there is evidence that frost resistance is not necessarily a function of vernalization, as in the case of facultative varieties that are frost tolerant but not vernalization sensitive (see also Limin et al. 2007). Moreover, a direct interaction of *VRN-1/FR-1* with the promoter of genes involved in the development of tolerance has not been demonstrated yet. First attempts to describe the relationships between vernalization- and photoperiodically-regulated genes/loci and factors that regulate the development of frost tolerance upon cold acclimation have been recently reviewed (Galiba et al. 2009; Pecchioni et al. 2013), and a possible advance in this direction has been reported by Eagles and colleagues (2011) in winter wheats. These authors found interesting implications for freezing tolerance of a C-T transition located in exon 4 of the *Vrn-A1* gene. The SNP (single nucleotide polymorphism) was already predicted to cause a leucine-to-phenylalanine substitution in the amino acid sequence of *VRN-A1* and to affect a conserved K domain of the protein (Chen et al. 2009d). Since K domains typically mediate interaction and/or dimerization between MADS box proteins, the C-T allelic variation was hypothesized to alter vernalization response, freezing tolerance, and other physiological traits (Eagles et al. 2011). Although understanding the molecular basis of the genetic network that interconnect vernalization and frost resistance is far from being complete, we propose in Figure 7.2 a hypothetical model whereby *VRN-1/FR-1* and *FR-2* interact with each other and with effector genes, leading to low-temperature tolerance and reproductive development.

The frost tolerance conferred by *FR-2*, appears to be ruled by one or more genes from the family C-repeat binding factor (*CBF*), also known as DRE binding protein 1 (*DREB1*), members of which occur in clusters of at least 13 elements (Francia et al. 2004, 2007; Skinner et al. 2005; Miller et al. 2006). At present, determination of whether the *FR-2* effect is a result of either several CBFs acting in an additive manner, or the action of only one, is still under way. In diploid einkorn wheat, *T. monococcum*, a deletion of five amino acids in the AP2 DNA binding domain of *CBF12* was found to co-segregate with frost tolerance between the spring, frost-sensitive accession DV92, and the winter, frost-tolerant G3116

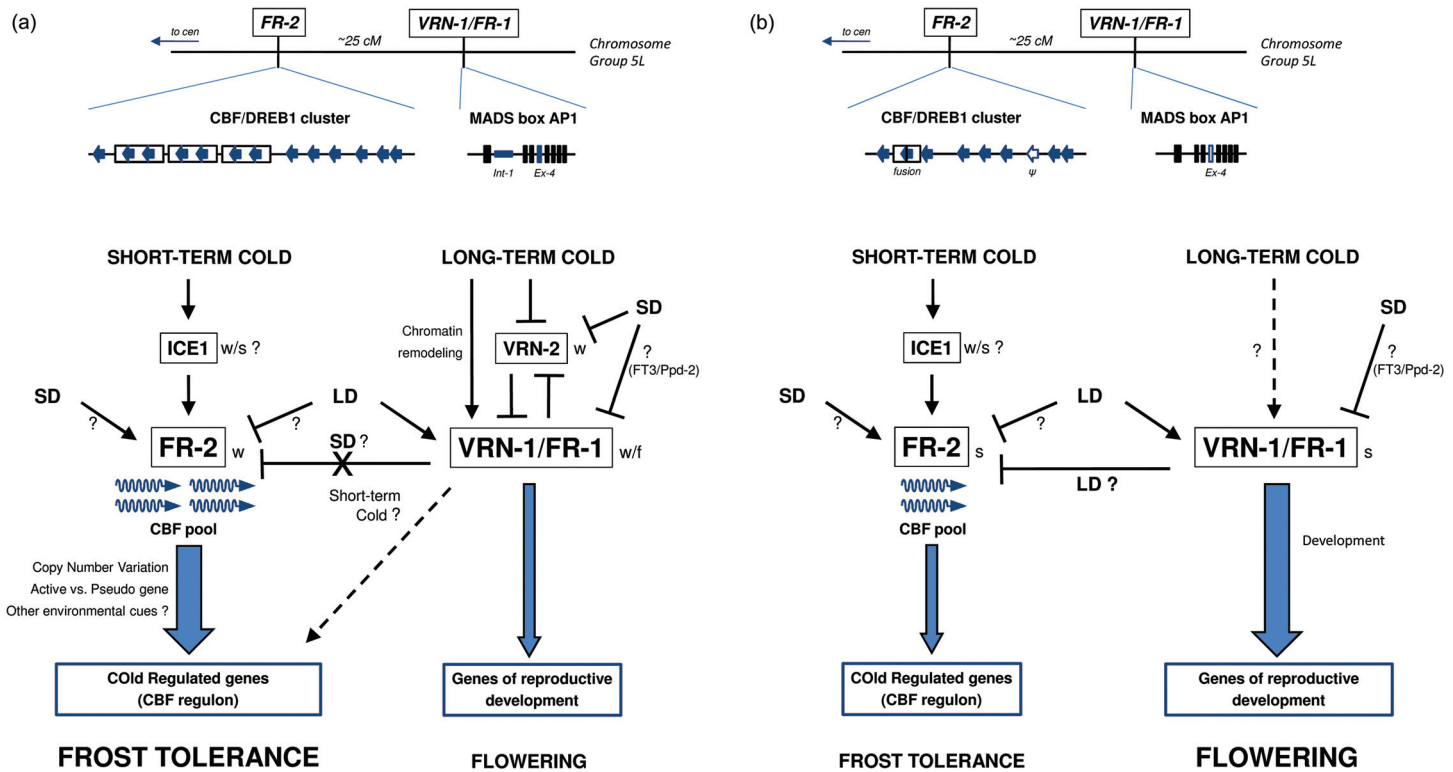


Fig. 7.2. Hypothetical model for the action of *VRN-1/FR-1* and *FR-2* in response to different environmental stimuli. The genetic structure of the two loci cis-acting on the long arms of homoeologous group 5 chromosomes in Triticeae is shown separately for winter/facultative (**a**) and spring (**b**) genotypes. SD, short-day conditions; LD, long-day conditions; w/f/s, winter/facultative/spring alleles. For a color version of this figure, please refer to the color plate.

(Knox et al. 2008). In barley the data suggest that an increase in the copy-numbers of *CBF2* and *CBF4* might be the underlying genetic basis for *FR-H2* tolerance alleles (Stockinger et al. 2007; Knox et al. 2010). Further reinforcing the existence of the prominent role played by the CBFs on frost tolerance in cultivated germplasm, Tondelli and colleagues (2006) mapped to two loosely linked positions on chromosome 7H, barley orthologs of the *AtFRY1* and *AtICE1* genes, putative upstream regulators of *CBF* genes in Arabidopsis (Thomashow 2010). No cold tolerance QTL have been mapped on chromosome 7H suggesting that, at least in barley, allelic variation at these two CBF regulators are not important for the trait. In the studies of Francia and colleagues (2004) and von Zitzewitz and colleagues (2011), the proportion of phenotypic variance accounted for by *FR-H1* ranged from 15% to 37%, whereas for *FR-H2* it ranged from 9% to 22%. Similar major effects were also observed at orthologous chromosome 5A loci in wheat (for a review see, e.g., Galiba et al. 2009).

In the presence of such effects, which are of similar weight at 5A loci, that explained a large part although not all the phenotypic variance for the trait, studies that attempt to identify other QTLs contributing to vegetative frost tolerance in Triticeae would deserve attention. These studies, on one hand allowed the investigation of contributions of other wheat chromosome 5 homoeologous to frost tolerance. Snape and colleagues (1997) identified a *FR-D1* locus on chromosome 5D, with a maximum likelihood peak about 6 cM proximal to *VRN-D1* (Snape et al. 2001). *FR-D1* might therefore be homoeologous to *FR-A2*. A similar situation exists for the *FR-B1* locus (Tóth et al. 2003), for which a recombinant substitution-line population derived from the cross 'Chinese Spring' x 'Chinese Spring'('Cheyenne' 5B) located *FR-B1,* close to but separate and proximal to *VRN-B1*. This was redesignated as *FR-B2* in the 2004 supplement of the Catalogue of Gene Symbols for Wheat (McIntosh et al. 2004). However, it has still not been verified experimentally whether this locus contains any *CBF* cluster like its *Fr-A2* homoeolog. Effects from 5B chromosome seemed to be less important for the trait than 5A and 5D loci, at least from results in Chinese Spring/Cheyenne (susceptible/resistant) chromosome substitution lines (Sutka 2001). In the same study, when comparing frost testing in phytotron and in the field, it was found that chromosome substitutions of the 5th homoeologous group and the 2B and 4B ones played an important role in both environments (Sutka 2001). A few other QTLs were mapped to other chromosomes, mainly in studies where at least *VRN-1*/*FR-1* was not segregating. A pioneering study by Tuberosa and colleagues (1997) found a slightly more complex situation in a "winter x winter" type barley cross 'Arda' x 'Opale.' Nine freezing-tolerance QTLs were mapped on chromosomes 2H, 3H, 6H, and 5H, by screening in a controlled environment, with the one on 5H roughly coinciding with *FR-1* (Tuberosa et al. 1997). It is also possible that this represented *FR-H2* instead, although in this case it couldn't be tested by comparative mapping (Cattivelli et al. 2002). In the wheat 'Chinese Spring' x 'Synthetic' population, Börner and colleagues (2002) found a region on chromosome arm 6AS that could be orthologous to the barley 6H-Bin6 segment which is associated with low-temperature tolerance (Tuberosa et al. 1997) and salt tolerance (Cattivelli et al. 2002). Another attempt to find additional loci in hexaploid wheat was made by Båga and colleagues (2007), by developing two doubled haploid populations with no allelic variation at *VRN-1/FR-1*. A major frost tolerance locus was again coincident with the position of *FR-A2* in the "winter x winter" population 'Norstar' x 'Winter Manitou' (which carries the recessive alleles *vrn-A1*, *vrn-B1*, *vrn-D1* from 'Norstar'). Two *CBF* genes (i.e., *CBF14* and *CBF15*) were located under the QTL peak, with the favorable allele contributed by the tolerant parent 'Norstar.' Interestingly, besides *FR-A2*, a weaker QTL was highlighted in a large region (about 20 cM) on chromosome 1D, with the favorable allele from 'Norstar.'

Latest advances in genomic tools have allowed fast and cost-effective mining of genome-wide single nucleotide polymorphisms in crops, the so-called GWA (genome-wide association) studies, with the aim of finding associations between SNP alleles and phenotypes of interest. Currently only three association-mapping approaches for studying frost tolerance have been published on the Triticeae: two in barley and one in rye, the most frost-tolerant species of the tribe. In a first attempt, not genome-wide, Fricano and colleagues (2009) investigated the allelic variation of only four barley *CBF* genes out of 13 in a panel of 216 accessions composed of European cultivars, landraces and *H. spontaneum* genotypes. Two nucleotide variants of *HvCBF14* and one nucleotide variant of *HvBM5A* were identified as statistically associated with frost tolerance (Fricano et al. 2009). Von Zitzewitz and colleagues (2011) performed a genome-wide association-mapping study of winter hardiness traits in 148 barley accessions consisting of advanced breeding lines and cultivars, assembled as part of the USDA barley Coordinated Agricultural Project (www.barleyCAP.org). Scoring freezing tolerance as percent winter survival, all the significant associations were on chromosome 5H, and only a few associations with markers on 3H, 4H, and 6H approached the threshold. In the *FR-H2* region, two SNPs were significantly associated with the trait, one in the *HvCBF9* gene, and the other located in a gene encoding a heat-shock transcription factor (HSF). At the *FR-H1* locus a specific *HvBM5A* intron 1 amplicon showed the most significant association, whereas a third significant SNP, about 8 cM proximal to *HvBM5A*, was in a *Glu-tRNA aminotransferase subunit C*, a gene with no obvious relationship to frost tolerance (von Zitzewitz et al. 2011). Finally, Li et al. (2011) studied eleven candidates involved in frost response (including several *ScCBFs*, *ScDREB2*, *ScICE2*, *and ScVRN1*) in a panel of 201 lines selected from Eastern and Middle European rye populations. High levels of nucleotide diversity and a fast intragenic LD (linkage disequilibrium) decay (within approximately 520 bp on average) were revealed, indicating that the low linkage disequilibrium in rye compared to self-pollinating species promises a high resolution in genome-wide association mapping.

Significant associations between freezing tolerance and SNPs/haplotypes of candidate genes were identified. In particular, two SNPs in *ScCBF15* and one in *ScCBF12*, all leading to amino acid changes, were related to frost tolerance. Although no significant SNP-frost tolerance association was observed for *ScVRN1*, this same gene had significant interaction effects with six other candidates (four *ScCBF*s plus *ScDREB2* and *ScDHN3*), underlining the important role of *ScVRN1* in the frost responsive network (Li et al. 2011). Taken as a whole, the association-mapping studies conducted so far in the Triticeae support the findings of bi-parental linkage mapping and expression studies, that is, that the *FR-2* and *FR-1* loci play crucial roles in frost resistance, with only a minor contribution of a few, still unknown other genes, residing at loci on other chromosome groups.

QTL and Genes Responsible for Reproductive Frost Tolerance

The degree of tolerance to freezing shown by temperate cereals depends largely on the stage of development at which the stress occurs. This is particularly true for winter cultivars at the vegetative stage, which become virtually dormant under the influence of steadily reduced temperature and day length. In later growth stages (i.e., from pre-heading to flowering) the plants become more susceptible to low temperatures (Fuller et al. 2009). In areas experiencing subtropical/Mediterranean climates, such as the Australian cereal belt, West Asia, and North Africa, wheat and barley are grown to reach maturity during winter, when conditions are most favorable for growth but occasional nighttime frost events in the order of –1 to –5° C occur (Reinheimer et al. 2004; Chen et al. 2009a). In

these environments, the choice of sowing date is constrained both by the higher probability of frost early in the season and by the hot and dry conditions that typically limit growth late in the season. In temperate regions, mid- to late-spring freezes may also cause spike damage to the winter wheat crops that flower during spring and summer (Whaley et al. 2004). These frost events can cause floret and spike abortion, as well as damage to the developing grain (Reinheimer et al. 2004; Frederiks et al. 2011). Efforts over the last few decades to improve tolerance of Australian cereal varieties to low temperatures in reproductive tissues (LTR tolerance) have been largely unsuccessful, owing to various difficulties of working with the trait (Fuller et al. 2009). Variation in heading time represents one such hindrance, because spikes at different growth stages differ in their frost susceptibility and flowering time can enable escape. However, this factor, which can potentially confound the measurement of genuine LTR tolerance, can be accounted for to some extent by scoring heads that were at a specific developmental stage at the time of frosting (Reinheimer et al. 2004). By adopting this strategy, Reinheimer et al. (2004) identified several lines with potential tolerance in the field, including the Japanese barley cultivars 'Amagi Nijo' and 'Haruna Nijo.' In two mapping populations derived from them (i.e., 'Amagi Nijo' x 'WI2585' and 'Haruna Nijo' x 'Galleon'), QTLs implicated in the genetic control of reproductive frost tolerance were found on chromosomes 2HL and 5HL, with tolerance alleles deriving from the Japanese parents (Reinheimer et al. 2004; Chen et al. 2009a). The frost resistance locus on chromosome 2HL was linked to *Flt-2L*, a locus controlling flowering time, spike density, and plant height (Chen et al. 2009b), whereas the QTL on chromosome 5HL was in the same general area (indeed co-segregation was not tested) of *VRN-H1/FR-H1* and was responsible for resistance to frost-induced floret sterility and frost-induced grain damage in all the studied populations (Reinheimer et al. 2004). In another experiment aimed at identifying potential LTR tolerance mechanisms, the 'Amagi Nijo' x 'WI2585' and 'Haruna Nijo' x 'Galleon' populations were examined for flowering time and spike morphology traits associated with the 2HL and 5HL LTR tolerance loci (Chen et al. 2009c). In spring-type progeny of both crosses, winter alleles at *VRN-H1/FR-H1* were linked in coupling with LTR tolerance and were associated with earlier flowering. On the other hand, tolerance on 2HL was coupled with late flowering alleles at *Flt-2L*. Both chromosome regions influenced chasmogamy/cleistogamy (open/closed florets), although tolerance was associated with cleistogamy at the 2HL locus and chasmogamy at the 5HL locus. LTR tolerance controlled by both loci was accompanied by shorter spikes, which were due to fewer florets per spike on 5HL, but shorter rachis internodes on 2HL (Chen et al. 2009c). Taken together, these results suggest that mechanisms of LTR tolerance differ from those proposed to control freezing tolerance at the vegetative stage, involving an extended vegetative growth phase. However, it remains to demonstrate, whether also in the case of LTR tolerance, the CBF-response pathway is activated and at which extent.

Copy-Number Variation and Winter Hardiness Co-selection

Copy-number variation is the term used to describe where allelic diversity occurs in the form of variable unit numbers of a given genomic segment. Typically the segment length varies in size from about 1 kb to 1 Mb (Scherer et al. 2007). Greatest insight into this phenomenon comes primarily from animal systems, in which there are whole genome reference sequences and genomes of other individuals have been resequenced. One of the main means by which copy-number variation has been revealed is array comparative genomic hybridization (CGH), in which reference and test genomic DNAs are hybridized to arrays representing the reference genome (Alkan et al. 2011). Regions common to both genomes but differing in

copy-number between the reference and the test sample will produce different signal intensities, which when expressed as ratios provide estimates of copy-number differences (Alkan et al. 2011). Newer sequence-based methods are expected to overtake CGH in facilitating both detection of copy-number variable regions and genotyping other individuals once copy-number variable regions are defined (Alkan et al. 2011). Estimates suggest about 12% of the human genome may be subject to variable copy-numbers (Redon et al. 2006). Structural variation across the genome is so prevalent that the National Center for Biotechnology Information (NCBI) now maintains a database solely devoted to genomic structural variation (http://www.ncbi.nlm.nih.gov/dbvar/).

A number of studies in plants have shown that these genomes also have variable numbers of segments. Hybridizing total genomic DNA of maize lines B73 and Mo17 to high-density oligonucleotide microarrays based on the B73 genome revealed several hundred sequences that differed in copy-number between the two genomes. About 180 gene sequences referred to as presence/absence variation (PAV) were also detected in one maize line but not in the other maize line (Springer et al. 2009). Resequencing of maize lines, including Mo17 and five additional elite lines, indicated that this phenomenon is widespread (Lai et al. 2010). Interestingly most of the CNV and PAV features in the maize genome are also variable in teosinte, the wild ancestor of maize (Swanson-Wagner et al. 2010). Copy-number variable regions have also been detected in soybean, and CGH analyses of introgression lines have indicated that this copy-number variation is stably inherited (Haun et al. 2011). In a different study it was asked whether structural variation arose de novo in response to selection pressure. The *Arabidopsis thaliana* Columbia ecotype was passed through five generations at three different temperatures (16° C, 22° C, and 28° C), and at each generation the lines producing the highest seed yield were selected for the next generation CGH experiments. This revealed numerous structural changes occurring at the upper and lower temperature relative to the 22° C reference temperature.

Detection of *CBF* Gene Copy-Number Variation in Triticeae Cereals

Copy-number variable regions have been found at wheat and barley loci for freezing tolerance and winter hardiness. These discoveries involved expression analyses of *CBF* genes and sequencing of bacteriophage lambda genomic clones encompassing the *CBF* genes. More than 20 *CBF* genes were identified in the genome of a single barley genotype, 13 of which mapped to *FR-H2* (Skinner et al. 2005; 2006; Francia et al. 2007). Two *CBF* genes, *CBF2* and *CBF4*, exhibited much higher transcript levels in 'Nure' than in 'Tremois' (Stockinger et al. 2007). Expression analyses were then carried out using pools of doubled haploid (DH) recombinants from the 'Nure' x 'Tremois' population (NT) in which individuals carried either the 'Nure' *Fr-H2* allele or the 'Tremois' *fr-H2* allele. DH individuals having the 'Nure' allele also expressed *CBF2* and *CBF4* to significantly higher levels than those having the 'Tremois' allele (Stockinger et al. 2007). Thus rather than a qualitative difference, there was a quantitative difference in *CBF* expression levels. Sequencing genomic clones in bacteriophage lambda vectors encompassing *CBF2* and *CBF4* from both 'Nure' and 'Tremois' and from 'Dicktoo' and 'Morex,' parents of another mapping population segregating for freezing tolerance (Skinner et al. 2006), indicated that the two frost-hardy barleys 'Nure' and 'Dicktoo,' which have a winter *vrn-H1* allele at *VRN-H1*, harbor two *CBF2* paralogs and multiple copies of the *CBF2A-CBF4B* genome segment, whereas the two non-frost-hardy spring barleys 'Tremois' and 'Morex,' which have a *Vrn-H1* spring allele at *VRN-H1*, harbor only single paralogs of *CBF2* and *CBF4* (Figure 7.2). It is important to point out that the transcribed regions are identical and would not have been revealed as being different through

cDNA sequencing strategies. Even the genomic clones having the different sequenced *CBF2A-CBF4B* genome segments were indistinguishable when they were RFLP(restriction fragment length polymophism)-fingerprinted with a standard set of enzymes and coding sequence probes.

Winter Hardiness and Association between *VRN-1* Allelic State and *CBF* Copy-Numbers

The relationship in these four barley genotypes between the allelic state at *VRN-H1*/*FR-H1* and the copy-numbers of the *CBF2A-CBF4B* genomic segment at *FR-H2* raised the question of whether this association occurs in other barleys, or in wheat. To address these questions, a wider sampling of barley and surveys of wheat were carried out using DNA blot hybridization strategies. The wider sample of barley was scored with a *CBF2* gene-specific probe (Knox et al. 2010). This *CBF2* probe cross hybridizes to both *CBF2A* and *CBF2B*, the latter a single copy gene, and as such the copy-numbers of *CBF2A* can be estimated by determining the ratio of *CBF2A* to *CBF2B* signal intensity. While this is by no means an exhaustive survey of barley, the data strongly supported the notion that there are two *CBF2* paralogs in the genomes of *vrn-H1* winter allele genotypes, one of which is amplified to multiple copies, and that there is only a single *CBF2* paralog in the genomes of spring allele genotypes (Figure 7.2). Just as analysis of *CBF*s in barley revealed that *CBF2* and *CBF4* expression levels correlated with gene copy-numbers, a similar strategy in wheat revealed that *CBF14* expression levels seem to correlate with copy-numbers (Knox et al. 2010). Using single chromosome recombinant lines Vágújfalvi and colleagues (Vágújfalvi et al. 2005) detected significantly higher transcript levels of *CBF14*, *CBF15*, and *CBF16* (*CBF7*, *CBF1A*, and *CBF1C*, respectively) in recombinants having the 'Cheyenne' 5A homoeolog segment than in recombinants having the 'Chinese Spring' 5A homoeolog segment. In separate experiments

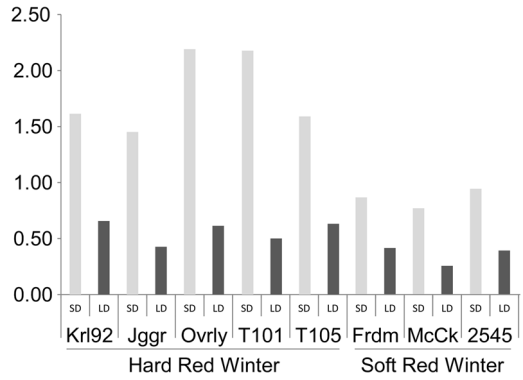


Fig. 7.3. *CBF14* expression in eight winter wheat cultivars. *CBF14* transcript levels are normalized to actin transcript levels. Wheats are grouped by their functional class. Hard red winter: 'Karl 92' (Krl92), 'Jagger' (Jggr), 'Overly' (Ovrly), TAM101 (T101), TAM105 (T105); soft red winter: 'Freedom' (Frdm), 'McCormick' (McCk), and 'Pioneer 2545' (2545).

in which a panel of nine winter wheats and three spring wheats were assayed, transcripts for *CBF1*, *CBF2*, *CBF9,* and *CBF14* accumulated to much higher levels in winter wheats than in spring wheats (Stockinger et al. 2007). However, across the group of nine winter wheats large differences in *CBF14* transcript levels were detected (Figure 7.3). Scoring this wheat germplasm with a *CBF14* gene-specific probe revealed that *CBF14* copy-numbers did indeed vary (Knox et al. 2010). Moreover, *CBF14* copy-number differences parallel expression level differences; that is, the wheats expressing *CBF14* at higher levels had greater signal intensity for two of the *CBF14* cross hybridizing fragments relative to levels detected in lines having greater signal intensity for only one of the *CBF14* cross hybridizing fragments (Knox et al. 2010). At this time it is unclear whether the other *CBF*s exhibiting expression-level differences across the different wheat lines also differ in copy-number.

Copy-number variable regions are thought to be a source of allelic diversity in terms of base pair differences within a species that is greater than that of single nucleotide polymorphisms (Alkan et al. 2011). In Triticeae cereals the genomic regions encompassing the *CBF*

genes show a trend that copy-numbers are higher in genotypes having a *vrn-1* winter allele than in genotypes having a *Vrn-1* spring allele. As *VRN-1* is the key factor required to make the reproductive transition from the vegetative growth phase, it will be interesting to better understand this association between *VRN-1* allelic state and *CBF* copy-numbers. The apparent increase in copy-numbers of the hard red winter wheats over the soft red winter wheats is also associated with the greater capacity of the hard red winter wheats to survive the semi-arid winters of the Great Plains (Knox et al. 2010). It will be interesting to determine whether this greater level of winter hardiness is associated only with the increase in *CBF14* copy-number or whether other loci are similarly affected and perhaps are even playing a role in potentiating the capacity of *CBF14* to activate its target genes. The increase in copy-number together with a vernalization-responsive *VRN-1* allele likely could have been co-selected by breeders, unconsciously, or consciously, since it leads to higher frost hardiness.

Wheat vs. Barley Genomics of Frost Tolerance: Common and Specific Mechanisms

In the first part of this chapter, genomics of freezing tolerance in Triticeae was compared with an overview of the results obtained by mapping and sequencing approaches. In addition to the static aspects of the genomic information such as DNA sequence or structural variation, functional genomics focuses on the dynamic aspects such as gene transcription and translation. Although several authors studied the changes in the transcriptome during cold acclimation (Gulick et al. 2005; Svensson et al. 2006; Monroy et al. 2007; Skinner 2009; Winfield et al. 2009, 2010; Kocsy et al. 2010; Ganeshan et al. 2011; Greenup et al. 2011) or proteome (Sarhadi et al. 2010; Rinalducci et al. 2011; Vitámvás et al. 2012) of wheat and barley, so far only two comparative studies investigating transcriptome in both species have been reported (Cho et al. 2006; Schreiber et al. 2009). However, the comparison of cold-induced changes in the transcriptome and metabolome of the two species is expected to provide new insights into the consequences of their speciation and polyploidization. The complete sequence of the barley and wheat genomes will be available in the near future, allowing this work to progress much more efficiently. From transcriptome results, it appears that most of the cold-affected genes are related to general defense, photosynthetic, and metabolic processes. In wheat both the effect of an abrupt decrease in temperature (Monroy et al. 2007; Kocsy et al. 2010) and of a gradual reduction in temperature (Winfield et al. 2009) were studied. The rationale of the latter system is the discovery of genes having different cold-induction threshold temperatures, which was shown in the case of *COR14b* gene, by comparison of wheat genotypes with different freezing tolerance (Vágújfalvi et al. 2003; Galiba et al. 2009). Transcriptome analysis of cold-acclimated shoots revealed differential regulation of genes encoding proteins involved in regulatory and metabolic processes in winter vs. spring genotypes (Gulick et al. 2005; Monroy et al. 2007). As between winter and spring wheats, such differences may also exist between wheat (more tolerant) and barley, due to their differential freezing tolerance. The number of low temperature-induced genes also depends on the level of freezing tolerance as it was described in wheat (Monroy et al. 2007; Kocsy et al. 2010; Winfield et al. 2010), and a species-dependent difference in this parameter is also probable. In wheat many genes (related to flowering and gibberellin pathway) were affected by cold either only in the crown or only in leaves during the vegetative-to-generative transition as was shown by microarray analysis (Winfield et al. 2009).

Similarly, different cold-induced changes were observed in gene expression in leaf and crown tissues of wheat by the cDNA-AFLP method (Ganeshan et al. 2011). The cold-induced alterations in the transcriptome were studied much less extensively in barley than in hexaploid wheat. Prolonged cold response was

detected in the expression of genes related to vernalization and signaling in barley (Greenup et al. 2011). Comparison of transcriptomes of wild type barley and mutants affected in chloroplast development revealed the major role of the chloroplasts in the control of the molecular adaptation to cold (Svensson et al. 2006). It is well established that cold acclimation in the Triticeae involves large transcriptome reconfiguration through major "regulatory hubs" (Svensson et al. 2006; Greenup et al. 2011; Laudencia-Chingcuanco et al. 2011). Evidence of an additional response mechanism that results in further modulation of the transcriptome was obtained in previously cold-acclimated wheat plants exposed to the "more severe conditions" (second-phase hardening) of –3 and –12° C, by Herman and colleagues (2006) and Skinner (2009) respectively. This treatment resulted in additional freezing tolerance, up-regulation of stress-related genes, and down-regulation of the ones related to photosynthesis in shoots (Herman et al. 2006). Although it is still unclear how much metabolic activity is retained by the plant as it cools below –10° C, a quarter of the 423 genes that were significantly up/down-regulated in the report by Skinner (2009) were not similar to any gene of known function.

On the other hand, as certain genes were differentially regulated in wheat cultivars with different tolerance during cold treatment, the identification of particular alleles related to a greater freezing survival can be determined in this experimental system. The transcriptional profile of wheat and barley was compared only at optimal growth temperature (not during cold treatment), and a high correlation was found between the two species (Schreiber et al. 2009). It was found that highly expressed genes are evolutionarily conserved, both in sequence and transcriptional activity. Alterations in the transcript levels resulting from the evolution of new roles of duplicated genes, which may derive from changes in the promoter sequence or in the trans regulation, was investigated in this study too. Gene activity levels across tissues after gene duplication positively correlated in wheat and barley. However, expression profiles of duplicated wheat genes were less similar to their barley orthologs than those of the duplicated barley genes to their wheat orthologs (Schreiber et al. 2009). This controversy can be explained by the following possibilities: (1) Different genes were duplicated in wheat and barley; (2) In the hexaploid wheat a greater functional divergence of the duplicated genes occurred than in the diploid barley because of the presence of the three genomes and the replacement of the functions by the homoeolog genes on the other genomes. Based on the observed correlation in the gene expression between the two species in the individual tissues, such a relationship can be also assumed between them if they are exposed to similar environmental conditions, such as low temperature.

A special way of exploring the differences between the barley and wheat transcriptomes was the characterization of wheat–barley chromosome addition lines (Cho et al. 2006). With a barley chip containing 22,792 probe sets, 4,014 transcripts could be detected specifically from barley, and, from these 4,014, 1,787 from 5 wheat–barley disomic chromosome addition lines. Thus, this system based on sequence and expression differences can be used for large-scale physical mapping of genes. Although cold-induced alterations in the transcript profile were not compared in barley and wheat, such a comparison was made in the case of CBF transcription factors (Campoli et al. 2009), which play a major role in the regulation of cold acclimation due to their effects on a large set of genes (Galiba et al. 2009). From a comparison of the expression of several *CBF* genes in Triticeae (rye, wheat, barley), it turned out that sample timing, induction temperature and light-related factors have significant effects on transcript levels, but there were no clear differences between the genotypes of a certain species (Campoli et al. 2009). At the proteome level, no comparison was made between barley and wheat. Cold-induced changes in the proteome were studied only in bread wheat (Sarhadi et al. 2010;

Rinalducci et al. 2011; Vítámvás et al. 2012). In leaves of cold-treated wheat the concentration of those proteins involved in defense (ascorbate recycling) and protein processing increased, and the concentration of those related to the Krebs cycle and photosynthesis decreased (Rinalducci et al. 2011).

A comparison of two wheat genotypes with different freezing tolerance and vernalization requirement showed a close association between the vernalization fulfillment and the start of a decline in the protein accumulation during growth at low temperature (Sarhadi et al. 2010). The authors also observed that the concentration of proteins related to defense, photosynthesis and metabolism was affected by cold. Comparison of chromosome 5A substitution lines having this chromosome from genotypes with different frost tolerance to each other revealed the effects of chromosome 5A and 5B on protein accumulation during long-term cold acclimation (Vitámvás et al. 2012). Similarly to the transcriptome level (Herman et al. 2006), at the proteome level a decrease in the amount of several photosynthetic proteins indicated the impairment of the photosynthetic processes (Vitámvás et al. 2012). A time lag between a down-regulation of cold-inducible transcripts and cold-inducible proteins was proposed in vernalized winter wheat plants, especially when COR/LEA proteins were investigated. A down-regulation of cold-inducible proteins in vernalized, but still cold-treated, winter wheat plants immediately after the vegetative/reproductive transition would have fatal effects on their survival (Sarhadi et al. 2010, Vítámvás et al. 2012). The time lag between the down-regulation of transcripts and proteins would ensure the survival of the plants at low temperature after this transition for a certain period, due to the prolonged existence of the protective proteins. The importance of several COR/LEA proteins in the protection of cell content during frost may be indicated by the fact that their relative abundance is correlated with the level of freezing tolerance of cold-treated spring and winter wheat genotypes (Houde et al. 1992). Transcriptomic and proteomic studies with wheat and barley indicate that cold acclimation induces freezing tolerance-dependent changes in the level of several mRNAs and proteins. Future comparative investigations are necessary in order to clarify which similarities and differences exist in the cold-induced changes in the transcriptome and proteome between barley and wheat.

Genetic Resources

The Middle East region known as the Fertile Crescent is considered the center of origin for cultivated Triticeae, based on the presence in this area of their wild progenitors and the archeological evidence (Zohary and Hopf 1993). Climatic conditions of this region are reflected in the winter growth habit characterizing the majority of wild progenitors of the tribe (Kosová et al. 2008). Together with a quantitative long-day response, winter growth habit is thus considered ancestral for the cultivated wheats and barley. After agriculture spread outside the Fertile Crescent, arriving at the Balkans around 6000 B.C., and to Europe over the following 2,000–3,000 years (Lister et al. 2009), novel adaptive traits suited to the new environments were selected. One of them was the spring growth habit that enabled expansion of cereal cultivation at higher latitudes without damage by cold winters (Cockram et al. 2009) and taking advantage of long growing seasons due to reduced photoperiod response (Peng et al. 2011). This brought into existence two more groups of barley and wheat, apart from winter types, divided according to the vernalization requirement: spring types and rarer facultative types. Satisfaction of vernalization is necessary for winter Triticeae to flower, although not strictly so; a strong delay in flowering occurs under prolonged long-day conditions, even without any period of cold, while the spring types have no vernalization requirement. There is more than one description for the facultative type; according to the International Union for the Protection of New Varieties of Plants (UPOV)

they are characterized by an intermediate flowering time (in comparison to winter and spring types) when grown under inductive photoperiods in the absence of vernalization (Cockram et al. 2009), while von Zitzewitz and colleagues (2005) classified them simply as cold-tolerant and vernalization unresponsive.

The vernalization requirement has therefore probably been the most important adaptative trait driving the differentiation of the cultivated gene pools of Triticeae throughout the world. Beginning with the Industrial Revolution, traditional agricultural systems were replaced by modern methods, and landraces by modern cultivars. At the beginning of the twentieth century the first pure-line varieties were released, and dwarfism genes were first introgressed into bread wheat by N. Strampelli in Italy (Borojevic and Borojevic 2005). Selection led however to a profound narrowing of the genetic basis of crops, leaving behind many potentially useful genes, as demonstrated by molecular marker techniques (Tanksley and McCouch 1997). Investigation of the nucleotide diversity in a sample of accessions corresponding to different stages in wheat evolution showed a strong diversity reduction of domesticated *T. dicoccum* (–70%), durum (–84%) and bread wheats (–69%), with respect to the wild *T. dicoccoides* (Haudry et al. 2007). In barley, molecular studies showed that only 40% of alleles found in the wild progenitor *Hordeum vulgare ssp. spontaneum* could be still found in cultivated genotypes (Ellis et al. 2000). These findings are not surprising, if we consider that the majority of barley 6-row winter varieties grown in Europe were developed from only five progenitors (Cockram et al. 2007), while most of hard red winter wheat varieties in the USA originated from just two lines imported from Poland and Russia (Tanksley and McCouch 1997).

The need to preserve endangered genetic resources such as landraces and wild species, coupled with an awareness that crop improvements are based on availability of diverse germplasm, led in 1983 to the founding of the Commission on Genetic Resources for Food and Agriculture. In 2001 the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) was adopted, followed by the establishment in 2004 of the Global Crop Diversity Trust, a public and private partnership that supports crop collection maintenance with individual and government funding (http://www.croptrust.org). Germplasm available today for breeding and research can be classified into six groups: modern cultivars, obsolete cultivars, landraces, genetic stocks, breeding lines, and wild relatives; and these can be subdivided according to the "three gene pool" concept introduced by Harlan and de Wet (1971). If we consider the potential usefulness of those gene pools for improving frost resistance, hexaploid spelt (*T. spelta* L.) is known to be more winter hardy than bread wheat, and could be sown in regions with a higher probability of frost damage; however, it yields less and is more difficult to harvest than bread wheat (Milisauskas 2011). A number of *A. tauschii* accessions were found to be frost hardy, but not as much as winter wheat cultivars (Limin and Fowler 1991). The research section at the International Maize and Wheat Improvement Center (CIMMYT) developed more than 1,000 synthetic hexaploid wheats, by using different accessions of *A. tauschii* as D genome donors. Synthetic hexaploids were resistant/tolerant to different abiotic and biotic stresses (Dreisigacker et al. 2008). However, the potential of synthetic wheat for improving frost tolerance at the vegetative stage was limited, that is, their hardiness levels were similar or equal to that of their hardy parent, with no transgressive recombinants, as highlighted in early attempts to deploy cold tolerance of amphidiploids (Limin and Fowler 1982; Le et al. 1986). Nevertheless, more recently some synthetic hexaploids were reported to suffer less frost damage during the reproductive period (Maes et al. 2001). In barley, high variability in tolerance to different abiotic stress, cold tolerance included, could be introgressed from the wild progenitor *H. vulgare*

ssp. spontaneum, representing the primary gene pool. Accordingly, weak frost tolerance alleles at QTLs on chromosomes 2H, 4H, and 6H were contributed by the wild parent in the cross 'Arta' x *Hordeum v. ssp. spontaneum* 41-1 (Baum et al. 2003).

A well known success story was the production of the interspecific fertile allopolyploid triticale in the 1960s, obtained by crossing wheat and the most frost-tolerant species among small grain cereals, that is, rye (Hömmö 1994). Since then triticale production grew and expanded, until the most frost-resistant cultivar obtained, 'Aubrac,' was demonstrated to perform much better at freezing temperatures than winter wheat and, under the specific test conditions, as well as a frost resistant rye (Rizza et al. 1997). The few examples of using wild relatives to improve tolerance to frost and other abiotic and biotic stresses, and the large amount of diversity stored in gene banks, leaves the impression that there is still a lot of novel germplasm with potential for exploitation in plant breeding. These few examples also pose questions about how to most efficiently use gene bank accessions and mine their sequence diversity in the genomic era. Despite great efforts to coordinate their activities, the quality of existing collections (availability of pure, properly characterized and documented materials) is still to be improved. Gene banks are expected to significantly improve their usability when they adopt systematic genotyping at key adaptation genes, or even genome-wide genotyping by sequencing (Kilian and Graner 2012), of pure germplasm stocks, coupled with their precise phenotyping. Molecular markers have begun to unravel genetic diversity (Spooner et al. 2005). For example, a set of 373 bread-wheat accessions reflecting the maximum of both the number of observed alleles and the number of geographical origins was selected by screening nearly 4,000 bread-wheat accessions from 73 countries with SSR markers, and by using passport data. Then, the seeds of all of these core accessions were made available, together with their DNAs stored on a single 384-well plate, ready for different molecular applications (Balfourier et al. 2007). Not only will proper procedures for extraction and long-term storage of DNA be required, the quality of gene bank information systems also still varies considerably, from poorly completed, to functional online databases with descriptors, molecular data, and geo-referenced data (Khoury et al. 2010). From a breeding point of view, once key loci for useful agronomic traits have been cloned, genotypic characterization can uncover further allelic variants that can be tested for superior performance. Allele mining of collections can be accomplished in two major ways: (1) a modified TILLING (targeting induced local lesions in genome) method called EcoTilling that detects the mutations naturally occurring in the primary and secondary gene pools; and (2) sequencing-based allele mining (Kumar et al. 2010).

Another approach to improving the efficiency of genetic resource utilization is the focused identification of germplasm strategy (FIGS), currently under development and evaluation. This approach uses the information about the environment where the accession was collected to select the accessions most likely to contain the traits that confer adaptation to that environment (http://www.figstraitmine.org). The relevance of the FIGS approach in better targeting accessions held in gene banks for valuable traits was demonstrated by El Bouhssini and colleagues (2011) when they compared random vs. FIGS-based screening of the International Center for Agricultural Research in the Dry Areas (ICARDA) collection for Russian wheat aphid resistance. While the screening of 5,000 randomly selected bread wheats did not identify resistant accessions, screening 510 accessions selected by means of FIGS enabled identification of 12 promising lines with different levels of resistance. A good example of how *ex-situ* germplasm collections can be mined for useful alleles by FIGS is a large-scale project that enabled the cloning of 7 new *Pm3* powdery mildew resistance alleles from a focused subset of 1,300 wheat landraces. Doubling the

known functional allelic diversity at the locus, FIGS selected landraces that were rich in new functional diversity demonstrating the success of the strategy (Bhullar et al. 2009). This strategy could be implemented for abiotic stress tolerance as well, for example, for identifying more allelic diversity for genes involved in winter hardiness. Most reports deal with allelic diversity of *VRN* genes in cultivated wheat and barley; however, wild progenitor accessions were also investigated in barley. For example, a large-scale survey of haplotype diversity at *VRN-H1* and *VRN-H2* loci in 429 spring, winter and facultative barley commercial European varieties enabled identification of three novel *VRN-H1* alleles present at low frequencies within EU germplasm (Cockram et al. 2007). Analysis of *VRN-H1* intron 1 In/Del variation showed that, while 98% of accessions carried winter alleles, three wild barley accessions (all originating in Israel) harbored spring alleles, one of which was not previously described in cultivated materials (Cockram et al. 2011). The development in less than ten years of the NGS (next-generation sequencing) platforms has led to a significant improvement in the throughput and cost of sequencing. This has made feasible the resequencing of candidate genes, transcriptomes and entire genomes, as well as comparisons with a reference genome sequence (Kumar et al. 2010). The availability of sequence information should assist in the management of germplasm resources, for example, to reduce redundancy of "core collections" (Glaszmann et al. 2010), assure homogeneity, and enable correct taxonomic classification. Once generated, sequence data will have to be followed by the development of appropriate bioinformatics tools; in addition, common standards and biobank information management systems (BIMs) will need to be adopted to manage such complex sets of data (Kilian and Graner 2012). Finally, high-throughput genotyping should be coupled with an effective and accurate phenotyping that combines conventional with novel technologies such as noninvasive imaging, spectroscopy, robotics, and high-performance computing, giving birth to phenomics, which could be described as "high-throughput plant physiology" (Furbank and Tester 2011).

From Sequence to Varieties: Advances in Assisted Selection of Large Genome Cereal Crops

Until now, a prerequisite for any application of DNA-based technologies in plant breeding has been the knowledge of the genetic basis of the trait, that is, the inheritance of the trait, number and position of loci, contribution of each locus, and locus interactions. Additional prerequisites should be markers associated with (in linkage disequilibrium with) the loci, together with a knowledge of their information content, alleles, and distance from the locus. A recent improvement in these applications has been brought about by gene and QTL cloning, including cloning for abiotic stresses. In many cases cloning has allowed moving from associated, linked, markers, to "perfect" candidate gene-derived markers (Salvi and Tuberosa 2005; Varshney et al. 2005). Vinocur and Altman (2005) reviewed how plants could be engineered for tolerance to abiotic stresses, using the increasing number of available cloned genes, with expected achievements and limitations; however this could be a practical option only for those countries having a permissive regulatory system.

A second step that should improve the precision of the DNA-based interventions in plant breeding is the implementation of high-throughput whole genome surveys, such as high-throughput SNP platforms and genotyping by sequencing using NGS techniques (Varshney et al. 2009). However, such advanced tools will be used routinely in plant breeding only when costs per sample can be lowered sufficiently. As detailed in the previous paragraphs, there have been significant gains in understanding the genetic bases for freezing tolerance in the Triticeae, which has turned out to be

relatively simple. The QTLs found generally act in an additive way, without significant epistasis. Moreover, the major loci acting in the different species appear to be located in syntenic chromosome locations (Galiba et al. 2009; Pecchioni et al. 2013). It is possible that new populations obtained by crossing among winter, facultative, and spring types could identify other QTLs for the trait; other contributions could come from multi-parent-derived populations, and from genome-wide association (GWA) studies. However, until present, breeding for freezing tolerance could be simply effective as marker-assisted selection (MAS) at a few selected loci. According to Kumar and colleagues (2012), another form of MAS should be named LD-MAS (i.e., MAS using markers in LD with a QTL), and a modern way to pursue LD-MAS would be the substitution of linked markers with candidate gene (CG) "perfect" markers (Varshney et al. 2005).

On the other hand, in a short/medium-term perspective, at least three facts suggest genomic selection (GS) could be possible and beneficial in selecting for freezing tolerance. The first of these is the progress made by international initiatives in sequencing wheat and barley (Feuillet et al. 2011). Secondly, field conditions most often impose combined stresses on a plant throughout its life cycle. Mittler (2006) reviewed evidence that plants have responses to a combination of two different abiotic stresses that cannot be deduced by applying each stress individually. Lastly, the traditional "reductionist science" approach is evolving into a "systems biology" approach. Such a systemic approach could tell us about interactions between loci and move marker-assisted breeding from the differential view of genetics to a comprehensive integrative approach. Interactions between loci, coding and non-coding, and between these and other factors such as cellular structures could significantly contribute to the phenotype. As a demonstration of such an integrative hypothesis, Joosen and colleagues (2009) reported a metabolome study where more than one third of the compounds present in RILs of Arabidopsis were not detected in either parent, but were the most likely result of the recombination of loci contributing to biosynthesis pathways.

LD-MAS for Freezing Tolerance

It seems clear that for freezing tolerance only a small battery of master regulators (i.e., *FR-1* and *FR-2*) should be included as priority targets in LD-MAS efforts aimed at improving the trait in temperate cereal crops. A strikingly successful example of how the selection of a single gene variant can consistently improve tolerance to an abiotic stress has been reported by Neeraja and colleagues (2007). They showed in rice how the simple introgression by marker-assisted backcrossing of the cloned *Sub1* QTL of submergence tolerance (a regulator of the MYB family), successfully improved the trait in a rice cultivar widely grown in flood-prone Asiatic regions. In the case of *FR-1*, the possibility of co-selecting growth habit together with frost tolerance is evident. Akar and colleagues (2009) reported the *VRN-H1* gene as the best predictor for marker-assisted selection for frost tolerance within Turkish highly frost-tolerant barley accessions and other winter, facultative, and spring barley germplasm. Similar indications came from the work of Rapacz and colleagues (2010), where allelic variation in the promoter region of *VRN-H1* (*HvBM5a*) was reported to be significantly related to freezing tolerance of plants that had been partially de-acclimated in the field. Moreover, simple diagnostic markers for *VRN* genes could be deployed in breeding programs to increase genetic diversity by utilizing winter × spring crosses (Cockram et al. 2009). For example, programs for developing freezing-tolerant facultative barley germplasm could also be pursued by a very simple strategy: fix winter alleles at *VRN-H1/FR-H1* and *FR-H2* on chromosome 5H, with the large deletion on chromosome 4H where *VRN-H2* resides. Photoperiod responses could then be targeted to

specific environments by selecting for appropriate alleles at the major loci determining long- or short-day sensitivity (that is *PPD-H1* on chromosome 2H and *PPD-H2* on chromosome 1H, respectively). Long-day sensitivity could maximize yield by delaying maturity in those environments not affected by terminal drought stress, while short-day sensitivity is appropriate for all environments, as it will delay the vegetative-to-reproductive transition. By contrast, LD-MAS of functional polymorphisms in *CBF* genes has still not yet been validated. In barley, Akar and colleagues (2009) found that only one out of the three markers designed on *CBF* genes was moderately associated with frost tolerance in barley, and Rapacz and colleagues (2010) did not find significant associations between a *HvCBF4* polymorphism and frost tolerance. Due to the observed CNV of specific CBF family elements differentiating tolerant and susceptible wheat and barley (Knox et al. 2010), fast and reliable methods for CNV detection could also be sought for application in LD-MAS. Other LD-MAS studies used PCR-based markers associated with *FR-1* and *FR-2* (Tóth et al. 2004), or the dehydrins *Wcs120* and *Dhn13* (Holková et al. 2009).

Genomic Selection (GS)

Genomic selection was proposed as a means of overcoming the limits of LD-based MAS to polygenic trait selection. It was defined by Meuwissen and colleagues (2001) as a method of predicting the breeding value of genotypes by analyzing phenotype together with high-density marker scores. Basically, GS simultaneously estimates all locus, haplotype, or marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs). The index incorporates all marker information in a prediction model, thereby avoiding biased marker-effect estimates and capturing more of the variation resulting from small-effect QTLs. By using high-density SNP panels, the genotype that would best fit with the genomic prediction model should be selected, containing cumulated positive effects from all contributing genes and minor effect QTLs. Heffner and colleagues (2011) demonstrated that for 13 agronomic traits in a population of 374 winter wheats, the average prediction accuracies for GS would be 28% higher than MAS. Since a high-density SNP panel could be excessively costly per single analysis, the adoption of GS by breeders strongly depends on the increasing availability of cheap high-throughput marker systems. Moreover, GS can be better proposed for species where genomic constitution is known, in terms of sequences and their physical position, and for which cultivar resequencing projects are in progress, as in apple (Kumar et al. 2012). In this view, the success in sequencing all gene-containing regions of barley and wheat is a necessary requirement to allow GS-based schemes. Recently, Paux and colleagues (2011) reported that GS methods are under evaluation for crops such as maize and wheat and, in some cases, are being applied in commercial breeding programs, although details have yet to be published. Genomic selection could be particularly useful for accumulating durable (quantitative) disease resistance QTLs in wheat, as proposed by Rutkoski and colleagues (2011) for stem rust, where the multigenic nature of adult plant resistance hampers the efficiency of MAS-based pyramiding. Because of the lack of mapped minor QTLs affecting the final level of freezing tolerance, GS for freezing tolerance could be an option for Triticeae, albeit together with genome-wide selection for other abiotic stress tolerances and agronomically relevant traits. Once high-density SNP panels can be made available and at reasonable assay costs, it should not be necessary to know all the (minor) QTL positions to select associated markers. While GS should substantially accelerate the breeding cycle, it would also dramatically change the role of phenotyping (including automated phenomics facilities), which could be used more to update the prediction models driving GS than to select lines (Heffner et al. 2009).

Conclusions and Perspectives

Global climate change will demand increasing versatility of crops to acclimate to multiple and dramatically fluctuating abiotic stress factors (Fedoroff et al. 2010). As underlined above, selection for an enhanced frost tolerance will remain an important aim of cereal breeding programs in temperate climate zones, together with tolerance to drought, salt, flooding, and improved nutrient efficiencies. Moreover, selection for multiple stress tolerances should be coupled with fine tuning of the growth cycle, by manipulating *VRN*, *PPD*, and *EPS* (*Earliness per se*) genes, as suggested by Francia and colleagues (2011). Since the late 1970s, when Olien (1979) and then Levitt (1980) described the events leading to frost damage in plants, an impressive amount of experimental evidence has accumulated for genetic and molecular mechanisms that allow plants to tolerate freezing temperatures. Many pieces of the puzzle are in place, as described in this chapter, and together will allow wheat and barley improvement to be driven by means of simple DNA-based technologies. Some as yet unresolved issues could be addressed in the near future. First of all, little is known about the molecular mechanisms of low temperature sensing by the Triticeae cells. A few regulatory steps above the *CBF/DREB* genes are known in the model plant Arabidopsis, with a role for calcium and calmodulin signaling (Doherty et al. 2009), but the information is far from complete. Knowledge of this process could perhaps help to improve the speed of the acclimation cascade. For example, it is already known that frost tolerant wheat and barley genotypes are able to induce the cold acclimation machinery at higher temperatures than the susceptible ones. This, firstly observed in barley (Crosatti et al. 1995), spawned the hypothesis of threshold induction temperatures (Galiba et al. 2009). According to the hypothesis, frost-tolerant genotypes start decreasing their LT_{50} values and accumulating COR transcripts at higher growth temperatures than the frost-sensitive ones. In addition, the frost-tolerant genotypes can also acclimate faster in comparison to the frost-sensitive ones (Fowler 2008). This behavior could be further studied, since it would give an adaptive mechanism to plants subject to frost in reproductive (sensitive) phases or to plants subject to fast and dramatic fluctuations of temperature. Lastly, research on sensing mechanisms, as on speed of acclimation cascade, could be coupled with integrative biology research in order to achieve a better understanding of *FR-1* and *FR-2* loci interaction. It would then be important to know how different molecular networks interact in response to various abiotic stresses. Epigenetic studies until now have revealed how the methylation patterns of histones at *VRN-1* change in response to vernalization, thus mediating its environmentally regulated expression (Oliver et al. 2009), but, for example, no epigenetic studies have been done on CBF regulation. If we take an optimistic view, for Triticeae, modern high-throughput screening methods could soon enable us to carry out genome-based selection of genotypes possessing desired genomic structures. The new approaches should, one would hope, lead to the development of new cultivars with physiologically tailored characteristics and with an improved versatility against abiotic stresses, underlying the plant's ability to cope with rapidly changing environments.

Acknowledgements

This research was supported by the Hungarian Scientific Research Fund (CNK80781, OTKA K83642) and by by the National Development Agency grant 1021 TÁMOP-4.2.2/B-10/1-20100025 Doctoral School of Molecular- and Nanotechnologies, 1022 Faculty of Information Technology, University of Pannonia.

Research was supported by Fondazione Cassa di Risparmio di Modena, Progetto di Ricerca Internazionale "FROSTMAP – Physical mapping of the barley Frost resistance-H2 (Fr-H2) locus". Funding from "GENOMORE" project

of the Fondazione Cassa di Risparmio di Carpi is also acknowledged.

Salaries and research support in the Stockinger lab provided by state and federal funds appropriated to The Ohio State University, Ohio Agricultural Research and Development Center, the Ohio Plant Biotechnology Consortium, and USDA-CSREES subaward CO396A-F. Authors are grateful to Professor Ignacio Romagosa and to Dr. Taner Akar for having provided Spain and Turkey frost tolerance data, generated within the European Union-INCO-MED program MABDE (ICA3-CT2002-10026).

References

Akar, Taner, Enrico Francia, Alessandro Tondelli, Fulvia Rizza, A. Michele Stanca, and Nicola Pecchioni. 2009. Marker-assisted characterization of frost tolerance in barley (Hordeum vulgare L.). *Plant Breeding* 128(4):381–386.

Alkan, Can, Bradley P. Coe, and Evan E. Eichler. 2011. Genome structural variation discovery and genotyping. *Nature Reviews Genetics* 12(5):363–376.

Båga, Monica, Sanjay V. Chodaparambil, Allen E. Limin, Marin Pecar, D. Brian Fowler, and Ravindra N. Chibbar. 2007. Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. *Functional & Integrative Genomics* 7(1):53–68.

Balfourier, François, Valérie Roussel, Pjotr Strelchenko, Florence Exbrayat-Vinson, Pierre Sourdille, Gilles Boutet, Jean Koenig, Catherine Ravel, Olga Mitrofanova, Michel Beckert, and Gilles Charmet. 2007. A worldwide bread wheat core collection arrayed in a 384-well plate. *Theoretical and Applied Genetics* 114(7):1265–127.

Baum, Michael, Stefania Grando, Gunter Backes, Ahmed Jahoor, A. Sabbagh, and Salvatore Ceccarelli. 2003. QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' x *H. spontaneum* 41-1. *Theoretical and Applied Genetics* 107(7):1215–1225.

Bhullar, Navreet K., Kenneth Street, Michael Mackay, Nabila Yahiaoui, and Beat Keller. 2009. Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proceedings of the National Academy of Sciences* 106(23):9519–9524.

Börner, Andreas, Erika Schumann, A. Fürste, H. Cöster, B. Leithold, Marion S. Röder, and Wilhelm Eberhard Weber. 2002. Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 105(6-7):921–936.

Borojevic, Katarina, and Ksenija Borojevic. 2005. The transfer and history of 'reduced height genes' (Rht) in wheat from Japan to Europe. *Journal of Heredity* 96(4):455–459.

Campoli, Chiara, Maria A. Matus-Cádiz, Curtis J. Pozniak, Luigi Cattivelli, and D. Bryan Fowler. 2009. Comparative expression of Cbf genes in the Triticeae under different acclimation induction temperatures. *Molecular Genetics and Genomics* 282(2):141–152.

Cattivelli, Luigi, Paolo Baldi, Cristina Crosatti, Natale Di Fonzo, Primetta Faccioli, Maria Grossi, Anna M. Mastrangelo, Nicola Pecchioni, and A. Michele Stanca. 2002. Chromosome regions and stress-related sequences involved in resistance to abiotic stress in *Triticeae*. *Plant Molecular Biology* 48(5-6):649–665.

Chen, Andrew, Lawrence V. Gusta, Anita Brûlé-Babel, Richard Leach, Ute Baumann, Geoffrey B. Fincher, and Nicholas C. Collins. 2009a. Varietal and chromosome 2H locus-specific frost tolerance in reproductive tissues of barley (*Hordeum vulgare* L.) detected using a frost simulation chamber. *Theoretical and Applied Genetics* 119(4):685–694.

Chen, Andrew, Ute Baumann, Geoffrey B. Fincher, and Nicholas C. Collins. 2009b. *Flt-2L*, a locus in barley controlling flowering time, spike density, and plant height. *Functional and Integrative Genomics* 9:243–254.

Chen, Andrew, Jason Reinheimer, Anita Brûlé-Babel, Ute Baumann, Margaret Pallotta, Geoffrey B. Fincher, and Nicholas C. Collins. 2009c. Genes and traits associated with chromosome 2H and 5H regions controlling sensitivity of reproductive tissues to frost in barley. *Theoretical and Applied Genetics* 118:1465–1476.

Chen, Yihua, Brett F. Carver, Shuwen Wang, Fengqiu Zhang, and Liuling Yan. 2009d. Genetic loci associated with stem elongation and winter dormancy release in wheat. *Theoretical and Applied Genetics* 118(5):881–889.

Cho, Seungho, David F. Garvin, and Gary J. Muehlbauer. 2006. Transcriptome analysis and physical mapping of barley genes in wheat–barley chromosome addition lines. *Genetics* 172(2):1277–1285.

Cockram, James, Elena Chiapparino, Scott A. Taylor, Konstantina Stamati, Paolo Donini, David A. Laurie,·and Donal M. O'Sullivan. 2007. Haplotype analysis of vernalization loci in European barley germplasm reveals novel VRN-H1 alleles and a predominant winter VRN-H1/VRN-H2 multi-locus haplotype. *Theoretical and Applied Genetics* 115(7):993–1001.

Cockram, James, Carol Norris, and Donald M. O'Sullivan. 2009. PCR-based markers diagnostic for spring and winter seasonal growth habit in barley. *Crop Science* 49(2):404–310.

Cockram, James, Huw Hones, and Donal M. O'Sullivan. 2011. Genetic variation at flowering time loci in wild and

cultivated barley. *Plant Genetic Resources: Characterization and Utilization* 9(2):264–267.

Crosatti, Cristina, Chiara Soncini, A. Michele Stanca, and Luigi Cattivelli. 1995. The accumulation of a cold-regulated chloroplastic protein is light-dependent. *Planta* 196(3):458–463.

Danyluk, Jean, Ndjido A. Kane, Ghislain Breton, Allen E. Limin, D. Brian Fowler, and Fathey Sarhan. 2003. TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiology* 132(4):1849–1860.

Dhillon, Taniya, Stephen P Pearce, Eric J. Stockinger, Assaf Distelfeld, Chengxia Li, Andrea K. Knox, Ildikó Vashegyi, Attila Vágújfalvi, Gabor Galiba, and Jorge Dubcovsky. 2010. Regulation of freezing tolerance and flowering in temperate cereals: the VRN-1 connection. *Plant Physiology* 153(4):1846–1858.

Doherty, Colleen J, Heather A Van Buskirk, Susan J Myers, and Michael F Thomashow. 2009. Roles for Arabidopsis CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. *The Plant Cell* 21(3):972–984.

Dreisigacker, Susanne, Masahiro Kishii, Jacob Lage, and Marilyn Warburton. 2008. Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. *Australian Journal of Agricultural Research* 59(5):413–420.

Eagles, Howard A., Karen Cane, and Ben Trevaskis. 2011. Veery wheats carry an allele of Vrn-A1 that has implications for freezing tolerance in winter wheats. *Plant Breeding* 130(4):413–418.

El Bouhssini, Mustapha, Kenneth Street, Ahmed Amri, Michael Mackay, Francis C. Ogbonnaya, Adnan Omran, Osman Abdalla, Michael Baum, Dabbous and Fawzi Rihawi. 2011. Sources of resistance in bread wheat to Russian wheat aphid (*Diuraphis noxia*) in Syria identified using the Focused Identification of Germplasm Strategy (FIGS). *Plant Breeding* 130:96–97.

Ellis, Roger P., Brian P. Forster, David Robinson, Linda Handley, Dennis C. Gordon, Joanne R. Russell, and Wayne Powell. 2000. Wild barley: A source of genes for crop improvement in the 21st century? *Journal of Experimental Botany* 51(342):9–17.

Fedoroff, Nina V., David S. Battisti, R.N. Beachy, Peter J.M. Cooper, David A. Fischhoff, C.N. Hodges, Vic C. Knauf, David B. Lobell, Barbara J. Mazur, David J. Molden, Matthew P. Reynolds, Pamela C. Ronald, Mark W. Rosegrant, Pedro A. Sanchez, A. Vonshak and Jinag-Kang Zhu. 2010. Radically rethinking agriculture for the 21st century. *Science* 327(5967):833–834.

Feuillet, Catherine, Jan E. Leach, Jane Rogers, Patrick S. Schnable, and Kellye Eversole. 2011. Crop genome sequencing: Lessons and rationales. *Trends in Plant Science* 16(2):77–88.

Fowler, D. Brian. 2008. Cold Acclimation Threshold Induction Temperatures in Cereals. *Crop Science* 48 (3):1147–1154.

Francia, Enrico, Fulvia Rizza, Luigi Cattivelli, A. Michele Stanca, Gabor Galiba, B. Tóth, Patrick M. Hayes, Jeffrey S. Skinner, and Nicola Pecchioni. 2004. Two loci on chromosome 5H determine low-temperature tolerance in a 'Nure' (winter) x 'Tremois' (spring) barley map. *Theoretical and Applied Genetics* 108(4):670–680.

Francia, Enrico, Delfina Barabaschi, Alessandro Tondelli, Giovanni Laidò, Fulvia Rizza, A. Michele Stanca, Matteo Busconi, Corrado Fogher, Eric J. Stockinger, and Nicola Pecchioni. 2007. Fine mapping of a Hv*CBF* gene cluster at the frost resistance locus *Fr-H2* in barley. *Theoretical and Applied Genetics* 115(8): 1083–1091.

Francia, Enrico, Alessandro Tondelli, Fulvia Rizza, Franz W. Badeck, Orazio Li Destri Nicosia, Taner Akar, Stefania Grando, Adnan Al-Yassin, Abdelkader Benbelkacem, William T.B. Thomas, Fred van Eeuwijk, Ignacio Romagosa, A. Michele Stanca, and Nicola Pecchioni. 2011. Determinants of barley grain yield in a wide range of Mediterranean environments. *Field Crops Research* 120 (1):169–178.

Frederiks, Troy M., Jack Christopher, S.E.H. Fletcher, and Andrew K. Borrell. 2011. Post head-emergence frost resistance of barley genotypes in the northern grain region of Australia. *Crop & Pasture Science* 62(9): 736–745.

Fricano, Agostino, Fulvia Rizza, Primetta Faccioli, Donata Pagani, Paolo Pavan, Alessandra Stella, Laura Rossini, Pietro Piffanelli, and Luigi Cattivelli. 2009. Genetic variants of HvCbf14 are statistically associated with frost tolerance in a European germplasm collection of Hordeum vulgare. *Theoretical and Applied Genetics.* 119(7):1335–1348.

Fuller, Mick P., Jack Christopher, and Troy M. Frederiks. 2009. Low-temperature damage to wheat in head - Matching perception with reality. In Lawrence V. Gusta, Michael E. Wisniewski, and Karen K. Tanino (eds.), *Plant cold hardiness: From the laboratory to the field*, Cambridge and Madison WI: CABI Publishing, pp. 12–18.

Furbank, Robert T., and Mark Tester. 2011. Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* 16(12):635–644.

Galiba, Gábor, Attila Vágújfalvi, Chengxia Li, Alexandra Soltész, and Jorge Dubcovsky. 2009. Regulatory genes involved in the determination of frost tolerance in temperate cereals. *Plant Science* 176(1):12–19.

Galiba, Gabor, Steve A. Quarrie, Jozsef Sutka, Alexei I. Morgounov, and John W. Snape. 1995. RFLP mapping of the vernalization (Vrn1) and frost resistance (Fr1) genes on chromosome 5A of wheat. *Theoretical and Applied Genetics* 90(7-8):1174–1179.

Ganeshan, Seedhabadee, Pallavi Sharma, Lester Young, Ashwani Kumar, D. Brian Fowler, and Ravindra N. Chibbar. 2011. Contrasting cDNA-AFLP profiles between crown and leaf tissues of cold-acclimated wheat plants indicate differing regulatory circuitries for low

temperature tolerance. *Plant Molecular Biology* 75(4-5):379–398.

Glaszmann, Jean-Christophe, Benjamin Kilian, Hari D Upadhyaya, and Rajeev Varshney. 2010. Accessing genetic diversity for crop improvement. *Current Opinion in Plant Biology* 13(2):167–173.

Greenup, Aaron G., Sharyar Sasani, Sandra N. Oliver, Sally A. Walford, Anthony A. Millar, and Ben Trevaskis. 2011. Transcriptome analysis of the vernalization response in barley (Hordeum vulgare) seedlings. *PloS One* 6(3):e17900.

Gulick Patrick J., Simon Drouin, Zhihua Yu, Jean Danyluk, Guylaine Poisson, Antonio F. Monroy, and Fathey Sarhan. 2005. Transcriptome comparison of winter and spring wheat responding to low temperature. *Genom*e 48(5):913–923.

Harlan, Jack, and Jan de Wet. 1971. Towards a rational classification of cultivated plants. *Taxon* 20:509–517.

Haudry, Annabelle, Alberto Cenci, Catherine Ravel, Thomas Bataillon, Dominique Brunel, Charles Poncet, Isabelle Hochu, Sebastien Poirier, Sylvain Santoni, Sylvain Glémin, and Jacques David. 2007. Grinding up wheat: A massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolutoin* 24(7):1506–1517.

Haun, William J., David L. Hyten, Wayne W. Xu, Daniel J. Gerhardt, Thomas J. Albert, Todd Richmond, Jeffrey A. Jeddeloh, Gaofeng Jia, Nathan M. Springer, Carroll P. Vance, and Robert M. Stupar. 2011. The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiology* 155(2):645–655.

Hayes, Patrick M., Tom K. Blake, Tony H. H. Chen, Somvong Tragoonrung, Fuqiang Chen, Aihong Pan, and Ben Hui Liu. 1993. Quantitative trait loci on barley (*Hordeum vulgare* L.) chromosome 7 associated with components of winterhardiness. *Genome* 36(1): 66–71.

Heffner, Elliot L., Jean-Luc Jannink, and Mark E. Sorrells, M. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome* 4:65–75.

Heffner, Elliot L., Mark E. Sorrells, and Jean-Luc Jannink. 2009. Genomic Selection for Crop Improvement. *Crop Science* 49 (1):1–12.

Herman, Eliot M., Kelsi Rotter, Ramaswamy Premakumar, G. Elwinger, Hanhong Bae, Rino Bae, Linda Ehler-King, Sixue Chen, and David P. Livingston 3rd. 2006. Additional freeze hardiness in wheat acquired by exposure to -3 degreesC is associated with extensive physiological, morphological, and molecular changes. *Journal of Experimental Botany* 57(14):3601–3618.

Holková, Ludmila, Ilja Tom Prášil, Marta Bradáčová, Pavel Vítámvás, and Oldřich Chloupek. 2009. Screening for frost tolerance in wheat using the expression of dehydrin genes Wcs120 and Wdhn13 at 17°C. *Plant Breeding* 128(4):420–422.

Hömmö, Leena Maarit. 1994. Hardening of some winter wheat (Triticum aestivum L.), rye (Secale cereals L.), triticale (Triticosecale Wittmack) and winter barley (Hordeum vulgare L.) cultivars during autumn and the final winter survival in Finland. *Plant Breeding* 112(4):285–293.

Houde, Mario, Rajinder S. Dhindsa, and Fathey Sarhan. 1992. A molecular marker to select for freezing tolerance in *Gramineae*. *Molecular and General Genetics* 234(1):43–48.

Joosen, Ronny V.L., Wilco Ligterink, Henk W.M. Hilhorst, and Joost J.B. Keurentjes. 2009. Advances in genetical genomics of plants. *Current Genomics* 10(8):540–549.

Khoury, Colin, Brigitte Laliberté, and Luigi Guarino. 2010. Trends in *Ex Situ* conservation of plant genetic resources: A review of global crop and regional conservation strategies. *Genetic Resources and Crop Evolution* 57(4):625–639.

Kilian, Benjamin, and Andreas Graner. 2012. NGS technologies for analyzing germplasm diversity in genebanks. *Briefings in Functional Genomics* 11 (1):38–50.

Knox, Andrea K., Chengxia Li, Attila Vágújfalvi, Gabor Galiba, Eric J. Stockinger, and Jorge Dubcovsky. 2008. Identification of candidate CBF genes for the frost tolerance locus Fr-A^m2 in Triticum monococcum. *Plant Molecular Biology* 67(3):257–270.

Knox, Andrea K., Taniya Dhillon, Hongmei Cheng, Alessandro Tondelli, Nicola Pecchioni, and Eric J. Stockinger. 2010. *CBF* gene copy number variation at *Frost Resistance-2* is associated with levels of freezing tolerance in temperate-climate cereals. *Theoretical and Applied Genetics* 121(1):21–35.

Kocsy, Gábor, Benedikt Athmer, Dragan Perovic, Axel Himmelbach, Attila Szűcs, Ildikó Vashegyi, Patrick Schweizer, Gábor Galiba, and Nils Stein. 2010. Regulation of gene expression by chromosome 5A during cold hardening in wheat. *Molecular Genetics and Genomics* 283(4):351–363.

Kosová, Klára, Ilja Tom Prášil, and Pavel Vítámvás. 2008. The relationship between vernalization- and photoperiodically-regulated genes and the development of frost tolerance in wheat and barley. *Biologia Plantarum* 52(4):601–615.

Kumar, G. Ram, Karuppusamy Sakthivel, Raman M. Sundaram, Chirravuri N. Neeraja, Sena.M. Balachandran, Nallathiga Shobha Rani, Basvaraj C. Viraktamath, and Maganti S. Madhav. 2010. Allele mining in crops: Prospects and potentials. *Biotechnology Advances* 28:451–461.

Kumar, Satish, Marco Bink, Richard Volz, Vincent Bus, and David Chagné. 2012. Towards genomic selection in apple (Malus ×domestica Borkh.) breeding programmes: Prospects, challenges and strategies. *Tree Genetics & Genomes* 8(1): 1–14.

Lai, Jinsheng, Ruiqiang Li, Xun Xu, Weiwei Jin, Mingliang Xu, Hainan Zhao, Zhongkai Xiang, Weibin Song, Kai Ying, Mei Zhang, Yinping Jiao, Peixiang

Ni, Jianguo Zhang, Dong Li, Xiaosen Guo, Kaixiong Ye, Min Jian, Bo Wang, Huisong Zheng, Huiqing Liang, Xiuqing Zhang, Shoucai Wang, Shaojiang Chen, Jiansheng Li, Yan Fu, Nathan M. Springer, Huanming Yang, Jian Wang, Jingrui Dai, Patrick S. Schnable, and Jun Wang. 2010. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics* 42(11): 1027–1030.

Laudencia-Chingcuanco, Debbie, Seedhabadee Ganeshan, Frank You, Brian Fowler, Ravindra Chibbar, and Olin Anderson. 2011. Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). *BMC Genomics* 12:299.

Le, Hoan T., David A. Reicosky, and Robert C. Olien. 1986. Freezing hardiness in amphiploids of *Triticum tauschii* (Coss.) Schmal. and *T. turgidum* L. *var. durum*. *Canadian Journal of Genetics and Cytology* 28(3):390–394.

Levitt, Jacob. 1980. *Responses of Plants to Environmental Stresses: Water, radiation, salt, and other stresses*, 2nd ed. New York: Academic Press.

Li, Yongle, Andreas Böck, Grit Haseneyer, Viktor Korzun, Peer Wilde, Chris-Carolin Schön, Donna P Ankerst, and Eva Bauer. 2011. Association analysis of frost tolerance in rye using candidate genes and phenotypic data from controlled, semi-controlled, and field phenotyping platforms. *BMC Plant Biology* 11:146.

Limin, Allen E., and D. Brian Fowler. 1982. The expression of cold hardiness in Triticum species amphiploids. *Canadian Journal of Genetics and Cytology* 24(1):51–56.

Limin, Allen E., and D. Brian Fowler. 1991. Breeding for cold hardiness in winter wheat: Problems, progress, and alien gene expression. *Field Crops Research* 27(3):201–218.

Limin, Allen E., Ann Corey, Patrick M. Hayes, and D. Brian Fowler. 2007. Low-temperature acclimation of barley cultivars used as parents in mapping populations: response to photoperiod, vernalization and phenological development. *Planta* 226(1):139–146.

Lister, Diane L., Susan Thaw, Mim A. Bower, Huw Jones, Michael P. Charles, Glynis Jones, Lydia M.J. Smith, Christopher J. Howe, Terence A. Brown, and Martin K. Jones. 2009. Latitudinal variation in a photoperiod response gene in European barley: Insight into the dynamics of agricultural spread from 'historic' specimens. *Journal of Archaeological Science* 36: 1092–1098.

Maes, Benoit, Richard M. Trethowan, Matthew P. Reynolds, Maarten van Ginkel, and Bent Skovmand. 2001. The influence of glume pubescence on spikelet temperature of wheat under freezing conditions *Australian Journal of Plant Physiology* 28(2):141–148.

McIntosh, Robert A., Katrien M. Devos, Jorge Dubcovsky, and William J. Rogers. 2004. Catalogue of gene symbols for wheat: 2004 supplement. Wheat Information Service [online]. Available from: http://wheat.pw.usda.gov/ggpages/awn/50/Textfiles/WGC.html.

Meuwissen, T.H.E., Ben J. Hayes, and Michael E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829.

Milisauskas, Saraunas. 2011. Early Neolithic, the first farmers in Europe, 7000–5500/5000 BC. In: Sarunas Milisauskas (ed.), *European Prehistory*, 2nd ed., New York: Springer-Verlag, pp. 153–222.

Miller, Andrea K., Gabor Galiba, and Jorge Dubcovsky. 2006. A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-Am2 in Triticum monococcum. *Molecular Genetics and Genomics* 275(2):193–203.

Mittler, Ron. 2006. Abiotic stress, the field environment and stress combination. *Trends in Plant Science* 11(1):15–19.

Monroy, Antonio F., Ani Dryanova, Brigitte Malette, Daniel H. Oren, Mohammed Ridha Farajalla, Wucheng Liu, Jean Danyluk, Lasantha W. C. Ubayasena, Khalil Kane, Graham J. Scoles, Fathey Sarhan, and Patrick J. Gulick. 2007. Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. *Plant Molecular Biology* 64(4):409–423.

Myles, Sean, Jason Peiffer, Patrick J. Brown, Elhan S. Ersoz, Zhiwu Zhang, Denise E. Costich, and Edward S. Buckler. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. *The Plant Cell* 21(8):2194–2202.

Neeraja, Chiruvuri Naga, Reycel Maghirang-Rodriguez, Alvaro M. Pamplona, Sigrid Heuer, Bertrand C.Y. Collard, Endang M. Septiningsih, Georgina V. Vergara, Darlene L. Sanchez, Kenong Xu, Abdelbagi M. Ismail, and David J. Mackill. 2007. A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theoretical and Applied Genetics* 115(6):767–776.

Charles R. Olien. 1979. Physiology of winter hardiness in barley. In Dickson, A.D., J.R. Harlan, D.L. Klingman, J.G. Moseman, and C.R. Olien (eds.), *Barley: Origin, botany, culture, winter hardiness, genetics, utilization, pests*, Hyattsville: USDA Agriculture Handbook no. 338, pp. 147–154

Oliver, Sandra N., E. Jean Finnegan, Elizabeth S. Dennis, W. James Peacock, and Ben Trevaskis. 2009. Vernalization-induced flowering in cereals is associated with changes in histone methylation at the VERNALIZATION1 gene. *Proceedings of the National Academy of Sciences* 106(20):8386–8391.

Pan, Aihong, Patrick M. Hayes, Fuqiang Chen, Tony H.H. Chen, Tom K. Blake, Stephen I. Wright, Ildiko Karsai, and Zoltan Bedo. 1994. Genetic analysis of the components of winterhardiness in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 89(7-8):900–910.

Paux, Etienne, Pierre Sourdille, Ian Mackay, and Catherine Feuillet. 2011. Sequence-based marker development in wheat: Advances and applications to breeding. *Biotechnology Advances* http://dx.doi.org/10.1016/j.biotechadv.2011.09.015.

Pecchioni, Nicola, Justyna A. Milc, Marianna Pasquariello, and Enrico Francia. 2012. Barley: Omics approaches for

abiotic stress tolerance. In Narendra Tuteja, Sarvajeet Singh Gill, Antonio F. Tiburcio, and Renu Tuteja (eds.), *Improving crop resistance to abiotic stress*. 1st ed. Weinheim: Wiley-VCH, pp. 777–882

Pecchioni, Nicola, Klára Kosová, Pavel Vítámvás, Ilja T. Prášil, Justyna A. Milc, Enrico Francia, Zsolt Gulyás, Gábor Kocsy, and Gábor Galiba. 2013. Genomics of low-temperature tolerance for an increased sustainability of wheat and barley production. In *Genomics of plant genetic resources to improve crop production, food security and nutritional quality. Vol. 2. Advances in genomics of plant genetic resources*. In press. Springer.

Peng, Junhua H., Dongfa Sun, and Eviatar Nevo. 2011. Domestication evolution, genetics and genomics in wheat. *Molecular Breeding* 28:281–301.

Prášil, Ilja T., Pavla Prášilová, and Pavel Mařík. 2007. Comparative study of direct and indirect evaluations of frost tolerance in barley. *Field Crops Research* 102(1): 1–8.

Rapacz, Marcin, Barbara Wolanin, Katarzyna Hura, and Miroslaw Tyrka. 2008. The effects of cold acclimation on photosynthetic apparatus and the expression of COR14b in four genotypes of barley (Hordeum vulgare) contrasting in their tolerance to freezing and high-light treatment in cold conditions. *Annals of Botany* 101(5):689–699.

Rapacz, Marcin, Miroslav Tyrka, Magdalena Gut, and Wojciech Mikulski. 2010. Associations of PCR markers with freezing tolerance and photosynthetic acclimation to cold in winter barley. *Euphytica* 175(3):293–301.

Redon, Richard, Shumpei Ishikawa, Karen R. Fitch, Lars Feuk, George H. Perry, T. Daniel Andrews, Heike Fiegler, Michael H. Shapero, Andrew R. Carson, Wenwei Chen, Eun Kyung Cho, Stephanie Dallaire, Jennifer L. Freeman, Juan R. Gonzalez, Monica Gratacos, Jing Huang, Dimitrios Kalaitzopoulos, Daisuke Komura, Jeffrey R. Macdonald, Christian R. Marshall, Rui Mei, Lyndal Montgomery, Kunihiro Nishimura, Kohji Okamura, Fan Shen, Martin J. Somerville, Joelle Tchinda, Armand Valsesia, Cara Woodwark, Fengtang Yang, Junjun Zhang, Tatiana Zerjal, Jane Zhang, Lluis Armengol, Donald F. Conrad, Xavier Estivill, Chris Tyler-Smith, Nigel P. Carter, Hiroyuki Aburatani, Charles Lee, Keith W. Jones, Stephen W. Scherer, and Matthew E. Hurles. 2006. Global variation in copy number in the human genome. *Nature* 444(7118):444–54.

Reinheimer, Jason L., Andrew R. Barr, and Jason K. Eglinton. 2004. QTL mapping of chromosomal regions conferring reproductive frost tolerance in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 109(6): 1267–1274.

Rinalducci, Sara, Maria Giulia Egidi, Ghasem Karimzadeh, Ferdous Rastgar Jazii, and Lello Zolla. 2011. Proteomic analysis of a spring wheat cultivar in response to prolonged cold stress. *Electrophoresis* 32(14): 1807–1818.

Rizza, Fulvia, Paolo Baldi, Luigi Cattivelli, and Giovanni Delogu. 1997. Cold hardening in triticale in comparison with rye and wheat. *Cereal Research Communications* 25(4):947–954.

Rizza, Fulvia, Donata Pagani, Magdalena Gut, Ilja T. Prášil, Chiara Lago, Alessandro Tondelli, Luigi Orrù, Elisabetta Mazzucotelli, Enrico Francia, Franz-Werner Badeck, Cristina Crosatti, Valeria Terzi, Luigi Cattivelli, A. Michele Stanca. 2011. Diversity in the response to low temperature in representative barley genotypes cultivated in Europe. *Crop Science* 51:2759–2779.

Rutkoski, Jessica E., Elliot L. Heffner, and Mark E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. *Euphytica* 179(1):161–173.

Salvi, Silvio, and Roberto Tuberosa. 2005. To clone or not to clone plant QTLs: Present and future challenges. *Trends in Plant Science* 10(6):297–304.

Sarhadi, Elham, Siroos Mahfoozi, Seyed Abdollah Hosseini, and Ghasem Hosseini Salekdeh. 2010. Cold acclimation proteome analysis reveals close link between the up-regulation of low-temperature associated proteins and vernalization fulfillment. *Journal of Proteome Research* 9(11):5658–5667.

Scherer, Stephen W., Charles Lee, Ewan Birney, David M. Altshuler, Evan E. Eichler, Nigel P. Carter, Matthew E. Hurles, and Lars Feuk. 2007. Challenges and standards in integrating surveys of structural variation. *Nature Genetics* 39(7 Suppl):S7–15.

Schreiber, Andreas W., Tim Sutton, Rico A. Caldo, Elena Kalashyan, Ben Lovell, Gwenda Mayo, Gary J. Muehlbauer, Arnis Druka, Robbie Waugh, Roger P. Wise, Peter Langridge, and Ute Baumann. 2009. Comparative transcriptomics in the *Triticeae*. *BMC Genomics* 10:285.

Skinner, Daniel Z. 2009. Post-acclimation transcriptome adjustment is a major factor in freezing tolerance of winter wheat. *Functional & Integrative Genomics* 9(4):513–523.

Skinner, Jeffrey S., Péter Szucs, Jarislav von Zitzewitz, Luis Marquez-Cedillo, Tanya Filichkin, Eric J. Stockinger, Michael F. Thomashow, Tony H.H. Chen, and Patrick M. Hayes. 2006. Mapping of barley homologs to genes that regulate low temperature tolerance in Arabidopsis. *Theoretical and Applied Genetics* 112(5): 832–842.

Skinner, Jeffrey S., Jarislav von Zitzewitz, Péter Szucs, Luis Marquez-Cedillo, Tanya Filichkin, Keenan Amundsen, Eric J. Stockinger, Michael F. Thomashow, Tony H. H. Chen, and Patrick M. Hayes. 2005. Structural, functional, and phylogenetic characterization of a large *CBF* gene family in barley. *Plant Molecular Biology* 59(4): 533–551.

Snape, John W., A. Semikhodskii, L. Fish, R. N. Sarma, Steve A. Quarrie, Gábor Galiba, József Sutka. 1997. Mapping frost resistance loci in wheat and comparative mapping with other cereals. *Acta Agronomica Hungarica* 45:265–270.

Snape John W., R.N. Sarma, Steve A. Quarrie, R.N. Fish, Gábor Galiba, József Sutka. 2001. Mapping genes for

flowering time and frost tolerance in cereals using precise genetic stocks. *Euphytica* 120: 309–315.

Spooner, David M., Rob van Treuren, and M.Carmen de Vicente. 2005. *Molecular Markers for Genebank Management*. IPGRI Technical Bulletin No. 10. Rome: International Plant Genetic Resources Institute.

Springer, Nathan M., Kai Ying, Yan Fu, Tieming Ji, Cheng-Ting Yeh, Yi Jia, Wei Wu, Todd Richmond, Jacob Kitzman, Heidi Rosenbaum, A. Leonardo Iniguez, W. Brad Barbazuk, Jeffrey A. Jeddeloh, Daniel Nettleton, and Patrick S. Schnable. 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genetics* 5(11):e1000734.

Stockinger, Eric J., Jeffrey S. Skinner, Kip G. Gardner, Enrico Francia, and Nicola Pecchioni. 2007. Expression levels of barley Cbf genes at the Frost resistance-H2 locus are dependent upon alleles at Fr-H1 and Fr-H2. *The Plant Journal* 51(2):308–321.

Sutka, József, and John W. Snape. 1989. Location of a gene for frost resistance on chromosome 5A of wheat. *Euphytica* 42(1-2):41–44.

Sutka, József. 2001. Genes for frost resistance in wheat. *Euphytica* 119:167–172.

Sutka, József, Gábor Galiba, Attila Vágújfalvi, Bikram S. Gill, and J. W. Snape. 1999. Physical mapping of the Vrn-A1 and Fr1 genes on chromosome 5A of wheat using deletion lines. *Theoretical and Applied Genetics* 99(1-2):199–202.

Svensson, Jan T., Cristina Crosatti, Chiara Campoli, Roberto Bassi, A. Michele Stanca, Timothy J. Close, and Luigi Cattivelli. 2006. Transcriptome analysis of cold acclimation in barley albina and xantha mutants. *Plant Physiology* 141(1):257–270.

Swanson-Wagner, Ruth A., Steven R. Eichten, Sunita Kumari, Peter Tiffin, Joshua C. Stein, Doreen Ware, and Nathan M. Springer. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Research* 20(12):1689–1699.

Tanksley, Steven D., and Susan R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277(5329):1063–1066.

Thomashow, Michael F. 1999. Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. *Annual Review of Plant Physiology and Plant Molecular Biology* 50:571–599.

Thomashow, Michael F. 2010. Molecular basis of plant cold acclimation: Insights gained from studying the CBF cold response pathway. *Plant Physiology* 154(2):571–577.

Tondelli, Alessandro, Enrico Francia, Delfina Barabaschi, Alessio Aprile, Jeffrey S. Skinner, Eric J. Stockinger, A. Michele Stanca, and Nicola Pecchioni. 2006. Mapping regulatory genes as candidates for cold and drought stress tolerance in barley. *Theoretical and Applied Genetics* 112(3):445–454.

Tondelli, Alessandro, Enrico Francia, Delfina Barabaschi, Marianna Pasquariello, and Nicola Pecchioni. 2011. Inside the CBF locus in Poaceae. *Plant Science* 180(1):39–45.

Tóth, Balázs, Gabor Galiba, Emma Fehér, József Sutka, and John W. Snape. 2003. Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. *Theoretical and Applied Genetics* 107(3):509–514.

Tóth, Balázs, Enrico Francia, Fulvia Rizza, A. Michele Stanca, Gábor Galiba, and Nicola Pecchioni. 2004. Development of PCR-based markers on chromosome 5H for assisted selection of frost-tolerant genotypes in barley. *Molecular Breeding* 14(3):265–273.

Trevaskis, Ben, David J. Bagnall, Marc H. Ellis, W. James Peacock, and Elizabeth S. Dennis. 2003. MADS box genes control vernalization-induced flowering in cereals. *Proceedings of the National Academy of Sciences of the United States of America* 100(22):13099–13104.

Trevaskis, Ben. 2010. The central role of the VERNALIZATION1 gene in the vernalization response of cereals. *Functional Plant Biology* 37(6):479–487.

Tuberosa, Roberto, Gábor Galiba, Maria Cristina Sanguineti, Enrico Noli, and József Sutka. 1997. Identification of QTL influencing freezing tolerance in barley. *Acta Agronomica Hungarica* 45:413–417.

Vágújfalvi, Attila, Gábor Galiba, Luigi Cattivelli, and Jorge Dubcovsky. 2003. The cold-regulated transcriptional activator *Cbf3* is linked to the frost-tolerance locus *Fr-A2* on wheat chromosome 5A. *Molecular Genetics and Genomics* 269(1):60–67.

Vágújfalvi, Attila, Alessio Aprile, Andrea Miller, Jorge Dubcovsky, Giovanni Delugu, Gábor Galiba, and Luigi Cattivelli. 2005. The expression of several *Cbf* genes at the *Fr-A2* locus is linked to frost resistance in wheat. *Molecular Genetics and Genomics* 274(5):506–514.

Varshney, Rajeev K, Andreas Graner, and Mark E Sorrells. 2005. Genomics-assisted breeding for crop improvement. *Trends in Plant Science* 10(12):621–630.

Varshney, Rajeev K, Spurthi N Nayak, Gregory D May, and Scott A Jackson. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27(9):522–530.

Vinocur, Basia, and Arie Altman. 2005. Recent advances in engineering plant tolerance to abiotic stress: Achievements and limitations. *Current Opinion in Biotechnology* 16(2):123–132.

Vítámvás, Pavel, Ilja T. Prášil, Klára Kosová, Sébastien Planchon, and Jenny Renaut. 2012. Analysis of proteome and frost tolerance in chromosome 5A and 5B reciprocal substitution lines between two winter wheats during long-term cold acclimation. *Proteomics* 12(1):68–85.

von Zitzewitz, Jarislav, Péter Szucs, Jorge Dubcovsky, Liuling Yan, Enrico Francia, Nicola Pecchioni, Ana Casas, Tony H. H. Chen, Patrick M. Hayes, and Jeffrey S. Skinner. 2005. Molecular and structural characterization of barley vernalization genes. *Plant Molecular Biology* 59(3):449–467.

von Zitzewitz, Jarislav, Alfonso Cuesta-Marcos, Federico Condon, Ariel J. Castro, Shiaoman Chao, Ann Corey, Tanya Filichkin, Scott P. Fisk, Lucia Gutierrez, Kale Haggard, Ildikó Karsai, Gary J. Muehlbauer, Kevin P. Smith, Otto Veisz, and Patrick M. Hayes. 2011. The genetics of winterhardiness in barley: Perspectives from genome-wide association mapping. *The Plant Genome Journal* 4(1):76.

Winfield, Mark O., Chungui Lu, Ian D. Wilson, Jane A. Coghill, and Keith J. Edwards. 2009. Cold- and light-induced changes in the transcriptome of wheat leading to phase transition from vegetative to reproductive growth. *BMC Plant Biology* 9:55.

Winfield, Mark O., Chungui Lu, Ian D. Wilson, Jane A. Coghill, and Keith J. Edwards. 2010. Plant responses to cold: Transcriptome analysis of wheat. *Plant Biotechnology Journal* 8(7):749–771.

Whaley, Michael J., Michael E.J. Kirby, John H. Spink, John M. Foulkes, and D.L. Sparkes. 2004. Frost damage to winter wheat in the UK: The eVect of plant population density. *European Journal of Agronomy* 21:105–115.

Yan, L, A Loukoianov, G Tranquilli, M Helguera, T Fahima, and J Dubcovsky. 2003. Positional cloning of the wheat vernalization gene *VRN1*. *Proceedings of the National Academy of Sciences of the United States of America* 100(10):6263–6268.

Zohary, Daniel, and Maria Hopf. 1993. *Domestication of plants in the Old World:The origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley*. Oxford: Clarendon Press.

Chapter 8

Molecular Breeding for Stay-Green: Progress and Challenges in Sorghum

Vincent Vadez, Santosh Deshpande, Jana Kholova, Punna Ramu, and C. Tom Hash

Abstract

The stay-green trait is regarded as the best characterized characteristic conferring drought adaptation in several crops including sorghum. Quantitative trait loci (QTLs) for stay-green have been identified using several bi-parental populations. Several of these QTLs are currently being used for introgression in a number of genetic backgrounds. Part of the challenge in the introgression of these QTLs lays in the limited polymorphism between donor and recurrent parents. As a consequence, certain QTL can't always be distinguished, such as Stg3 and StgB which are on the same chromosome, SBI-02. Current progress in marker technology is contributing to enhancing the marker coverage of QTL intervals and this would improve breeding efficiency. Despite the knowledge of genomic regions conferring the stay-green trait, it is surprising that knowledge of the physiological mechanisms explaining stay-green are still relatively unknown. Early explanations focused on a role of stay-green as maintaining photosynthetic activity. It has also been hypothesized that the stay-green trait relates to the plant nitrogen balance and in particular to the capacity to absorb nitrogen during the post-anthesis period. It is only relatively recently that water availability during the post-anthesis period, that is, when the stay-green phenotype expresses itself, has been proposed as a possible cause for the stay-green phenotype. However, the reasons that water is left for absorption are still unexplained and could be accounted for by either a deeper soil extraction depth or water saving traits operating at early stages. As the mechanisms responsible for stay-green become more evident and as DNA-sequencing technologies offer denser genome coverage, the likelihood is that the future of manipulating the stay-green trait will be about manipulating its physiological components.

Introduction

The capacity of certain genotypes in several annual crop species to maintain green leaves during the grain-filling period (the "stay-green" phenotype, SG) is an intriguing crop feature that has long been studied and included in breeding programs in several crops, especially under water-limited conditions. Indeed, the maintenance of green leaf area has been reported

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

to improve the quality of residues (van Oosterom et al. 1996), support the continuation of carbon fixation and supply of starch (McBee et al. 1983), prevent premature death and lodging (Rosenow and Clark 1981), sustain grain-filling under water stress (Rosenow et al. 1983; Rajcan and Tollenaar 1999a, 199b), and improve grain yield under stress (Borrell and Douglas 1996). Here we focus on using stay-green as a breeding target under water-limited conditions and review recent progress in different areas of stay-green research, with a particular focus on sorghum, where this trait has been most studied.

Given the potential benefit of stay-green, genotypes displaying this trait have been used to identify the genomic regions responsible for this phenotype. Several QTLs have been identified, using different breeding populations and stay-green QTL donors, and different types of drought stress. This information is reviewed and the most important QTLs are identified. We also review the experimental conditions in which phenotyping for stay-green has taken place and the different ways of assessing this phenotype, either from leaf senescence curves or leaf greenness assessments. A following section then summarizes current work being done at ICRISAT (International Crops Research Institute for the Semi-Arid Tropics, in Andhra Pradesh, India) to introgress several known QTLs for stay-green into various agronomically elite genetic backgrounds.

Our understanding of the stay-green trait and of the genetic regulation of mechanisms that lead to the expression of a stay-green phenotype in sorghum is still very incomplete. Early works considered the benefit of stay-green in terms of extending the period during which a leaf could actively fix carbon (McBee et al. 1983). Subsequent work also related to the carbon economy of the plant addressed the nitrogen status of the plant and in particular the balance between nitrogen demand and nitrogen capture (Borrell et al. 2001). A large amount of work, mostly involving transgenics, has addressed the question of maintaining the production of cytokinins to prevent leaf senescence (Gan and Amasino 1995). These views, which try to address the "symptoms" of stay-green, see in the degradation of the photosynthetic pigments the key entry point for manipulating the stay-green trait. These approaches are probably the complete opposite to more recent work that addresses the "cause" of stay-green and looks at stay-green from the angle of water supply, taking the view that stay-green expression is a consequence of having water available in the soil profile during grain-filling, when stay-green is actually measured (Vadez et al. under review). Therefore, two sections of this review will deal with these "early" and more recent considerations related to the stay-green trait.

As we progress in our understanding of the physiological mechanisms and genetic regulation of the stay-green phenotype in sorghum, manipulation of the trait is likely to evolve from the current introgression of genomic regions involved in expression of the stay-green phenotype, which we now know are likely explained by mechanisms of varied nature, to the introgression of these mechanistic components individually. For instance, it was found recently that B35 (= BTx642) donor parent alleles at stay-green QTL *Stg1* contributed to increased water extraction by moderately senescent caudatum variety S 35 (Vadez et al. 2011), but did not do this in the genetic background of highly-senescent durra variety R 16. Therefore, work will be needed to identify the best germplasm options as donors for each of the components of the stay-green phenotype, and these may vary with the genetic backgrounds and specific soil, water, and temperature regimes in which improved drought tolerance is desired. Work will also be needed to measure the "baseline" component trait value of potential recipient genotypes. This would involve both the development of high through-put phenotyping methods for assessing these traits, and the refinement of molecular tools for deciphering the genetic basis of these key traits. Current efforts in sorghum are exploiting

genotyping-by-sequencing data to provide full-genome scans across >100,000 SNP (single-nucleotide polymorphism) loci in each member of a portion of a global reference collection of sorghum germplasm that has been phenotyped with a lysimetric system to explore the allelic variation available for key components of the stay-green phenotype. Similar phenotype and genoptype data for available sets of stay-green QTL introgression lines in several genetic backgrounds are also being used to better characterize the genomic regions associated with each of six stay-green QTLs having favorable alleles from donor B35 [= BTx642].

Last but not least, the phenotypic evaluation of traits involved in drought adaptation is limited by the number of years and sites in which these experimentations can be carried out. This is so because many of the productive processes eventually generating yield are influenced substantially by environmental cues (e.g., Reymond et al. 2003). Given the large variability of agro-climatic (including weather and soil) conditions, and the difficulty of acquiring reliable yield estimates across environments, it is risky to rely on only experimental data to assess the value of a given trait, such as stay-green and its components, for yield improvement under water-limited conditions. The literature is riddled with reports of genotype-by-environment interactions, in one recent study, in the yield performance of a reference collection of groundnut (Hamidou et al. 2012): other studies simply address the environment-specific identification of stay-green QTLs across locations and years (Tuinstra et al. 1997; Crasta et al. 1999; Subhudi et al. 2000; Tao et al. 2000; Xu et al. 2000; Kebede et al. 2001; Haussmann et al. 2002; Sanchez et al. 2002). Accordingly, here we review how the use of crop-simulation modeling can contribute to an ex-ante assessment of the likely impact of traits shown to be associated with stay-green, in term of the probability of success, the possible range of yield increase, and geographic determination of where gains can be made.

QTL Identification

The late 1990s and early 2000s have seen a plethora of studies aimed at identifying stay-green QTLs. Several studies used as the stay-green donor parent B35 (= BTx642), a BC1 derivative of IS12555, a durra sorghum from Ethiopia (Rosenow et al. 1983). Six QTLs for pre-flowering stress tolerance were mapped (Tuinstra et al. 1996), with B35 as a source of stay-green, in a cross with pre-flowering drought tolerant line Tx7078. The stress was imposed by withholding irrigation for several weeks during the vegetative period, until flowering of about 80% of the lines. Drought tolerance was assessed either by the absolute yield under stress, or by the ratio of yield, seed number, or plant height under stress to the same parameters measured under fully-irrigated conditions. In a later study, using the same mapping population, three trials were conducted. Although two of these trials did not have any drought effect and the water stress was applied by withholding irrigation at flowering, a 40% grain yield reduction was achieved under imposed drought treatment. Stay-green was evaluated by scoring each plot for this trait, using a 1-to-5 scale, weekly from flowering until maturity. Several QTLs for stay-green were mapped, with two of these, on linkage group F (SBI-10) and I (SBI-10), also co-mapped with yield under either only drought or under both drought and fully-irrigated conditions (Tuinstra et al. 1997). Crasta and colleagues (1999) also used B35 as a stay-green donor parent and Tx430 as a senescent parent to produce a recombinant inbred line (RIL) population of 96 individuals to map seven different QTLs (StgA, StgD, StgG, as major QTLs, and StgB, StgI.1, StgI.2 and StgJ, as minor QTLs). The scoring of stay-green was done by visual assessment, with the experiment carried out in the field, with plants exposed to post-flowering stress. Tao and colleagues (2000) identified five QTLs for stay-green in trials that were conducted in five locations over three years. They used QL41 as a donor source, and QL39 as a drought-sensitive elite parent. QL41 is a derivative from a

cross between QL33 and B35. Their work clearly showed that QTLs varied across environments and years, and that three stay-green QTLs were each detected in more than two environments.

Xu and colleagues (2000) identified four stay-green QTLs (*Stg1*, *Stg2*, *Stg3*, and *Stg4*) in a mapping population based on a cross between B35 and Tx7000. Two trials were conducted in two locations and two years, and stress was imposed by stopping irrigation at anthesis. Stay-green was assessed with a plot score of 1 to 5 at physiological maturity. *Stg1* and *Stg2*, found on linkage group A (SBI-03), were consistently identified across locations and in both years, whereas *Stg3* and *Stg4* were on linkage group D (SBI-02) and J (SBI-05) and were found in specific seasons only. Using the same mapping population in other field trials, Subhudi and colleagues (2000) compared their QTL results to those of Crasta and colleagues (1999) and Tuinstra and colleagues (1997), and showed consistency of QTLs in different genetic backgrounds. Subudhi and colleagues (2000) showed that *Stg2*, *Stg3* and *Stg4* of the current populations corresponded to *StgA*, *StgD*, and *StgJ* of Crasta and colleagues (1999) and asserted that *Stg2* was likely the most important for the expression of the stay-green phenotype. Although *Stg1* of Subhudi and colleagues (2000) found no equivalent in Crasta and colleagues (1999), it was likely very closely related to *StgB* and *Stg I.1* of Crasta and colleagues (1999). Sanchez and colleagues (2002) reported on the development of near-isogenic introgression lines for four stay-green QTLs (*Stg1*, *Stg2*, *Stg3* and *Stg4* of Subudhi et al., 2000) in a marker-assisted backcrossing program involving B35 as donor and Tx7000 as recurrent parent.

Kebede and colleagues (2001) have mapped QTLs for stay-green with another donor parent, SC56, a conversion line (3-dwarf plant height and reduced photoperiod sensitivity) derived from a Sudanese caudatum-nigricans landrace. Another donor parent for stay-green, E 36-1, a cultivar of Ethiopian origin, has also been used to map QTLs for the stay-green trait in two RIL mapping populations from which a total of seven QTLs were identified (Haussmann et al., 2002), with three of them being common to both populations. Kassahun and colleagues (2010) reported results validating *StgB*, which appears to be identical to a stay-green QTL on the long arm of SBI-02 that has been incorporated into elite breeding material in Australia by conventional pedigree selection (D. Jordan and A. Borrell, pers. comm.). More recently, a third stay-green QTL on SBI-02, in addition to the previously reported *Stg3* and *StgB*, was reported by Haryarimana and colleagues (2010) as mapping to the interval flanked by SSR markers *Xtxp19* and *Xtxp298* on the short arm of SBI-02. Finally, Sabadin and colleagues (2012) have mapped two stay-green QTLs on SBI-02 (*St2-1* and *St2-2*) as well as single stay-green QTLs on SBI-03 (*St3*), SBI-04 (*St4*), SBI-05 (*St5*), SBI-06 (*St6*), SBI-08 (*St8*), and SBI-09 (*St9*), together explaining 65 to 69% of the genetic variance for stay-green in two water-stressed environments, using an SSR-anchored, DArT-saturated (Diversity Arrays Technology) linkage map for a modest-sized RIL population based on the cross of BR007, a breeding line from the Embrapa Maize and Sorghum program in Brazil, and SC283, a USDA sorghum conversion line based on guinea landrace IS7173C, from Tanzania. When flowering time and plant height were used as cofactors in the QTL analysis, many of the originally detected stay-green QTLs in this population were no longer statistically significant; on the other hand, *St3*, *St4*, *St8* and newly detected *St10* were statistically significant and together explained 30 to 35% of genetic variation for stay-green in the water-stressed environments in the two years of testing.

Overall, six sources of the stay-green trait (B35 = BTx642, E 36-1, QL41, SC56, SC283, and SDS 1948-3) have so far been used for the identification of QTLs for this phenotype in sorghum. Several of the stay-green QTLs identified have been validated in different backgrounds. However, there is to date only scant understanding and knowledge of the

physiological mechanisms underlying each of these stay-green QTLs and their interactions – at least little that has been published towards trait expression.

QTL Introgression – Current Progress at ICRISAT

The initial stay-green QTL mapping studies used RFLPs (restriction fragment length polymorphisms) and AFLPs (amplified fragment length polymorphisms; Tuinstra et al. 1997), but later studies extensively used SSRs (simple sequence repeats; Subudhi et al. 2000, Haussmann et al. 2002, Harris et al. 2007) and DArTs (Sabadin et al. 2012). Most of the stay-green QTL introgression and QTL validation studies reported to date (Tuinstra et al. 1996, Harris et al. 2007) have used RFLPs and/or SSRs as flanking markers for foreground selection – along with RAPD (random amplified polymorphic DNA) and/or AFLP markers.

Researchers at ICRISAT-Patancheru selected six candidate QTLs for the stay-green trait from donor B35, including *Stg1*, *Stg2*, *Stg3*, and *Stg4* reported by Subudhi and colleagues (2000), Sanchez and colleagues (2002), and Harris and colleagues (2007), as well as additional QTLs on SBI-01 (*StgA*) and SBI-02 (*StgB*), and initiated marker-assisted backcross (MABC) transfer of these into a number of genetically diverse, tropically-adapted elite sorghum varieties having a range of drought tolerance (Hash et al. 2003). Recurrent parents included highly senescent post-rainy season-adapted durra variety R 16, short 2-dwarf tan white-grained caudatum variety ISIAP Dorado, and tall 2-dwarf tan, white-grained, sweet-stemmed caudatum sister-line varieties S 35 and ICSV 111. The positions of flanking SSRs for these six target QTLs were putatively inferred from stay-green QTL mapping studies published in the late 1990s and early 2000s (Table 8.1). The unavailability of alternate SSRs and lack of polymorphism was a major constraint for this activity in the early 2000s, when this work was initiated. As a result, most of the stay-green QTLs targeted for this introgression work were characterized by large confidence intervals between flanking markers and scant availability of flanking SSR polymorphisms between the donor and recurrent parents. These limitations on readily available flanking marker polymorphism had many implications for this attempt to introgress stay-green QTLs into several target genetic backgrounds. The lack of enough polymorphic SSRs spread across each QTL region resulted in a high probability of losing the QTL even after flanking marker confirmation because of the possibility of recombination occurring within one or more of the putative QTL target regions, linkage drag with unfavourable traits, and ultimately a lower level of recurrent parent genome recovery. Similarly lack of polymorphic SSRs between B35 (donor) and recurrent parents (especially Indian *durras*) meant that no progress in stay-green QTL introgression work was possible in many target backgrounds until more markers were available. Accordingly, the MABC project was focused on two genetic backgrounds, R 16 and S 35=ICSV 111.

For the R 16-background, BC_2-progenies were developed with foreground selection each generation to confirm the presence of alleles from the donor parent at the SSR loci flanking each of the six putative stay-green QTLs, combined with limited background selection, in order to hasten the recovery of recurrent parent alleles in genomic regions distant from one or more of the target QTLs. These progenies were evaluated for stay-green expression (Kassahun et al. 2010), and one of the backcross progenies, RSG 04005, was confirmed as carrying three QTLs (*Stg3*, *Stg4* and *StgB*). This entry was used as the stay-green donor in another round of backcrosses to R 16 to derive single-QTL introgression lines. Two of the stay-green QTLs, *StgB* and *Stg3*, are mapped on sorghum chromosome SBI-02. These two QTLs are linked to the morphological marker gene *Z/z* controlling the important grain quality trait of mesocarp thickness, with the stay-green alleles from B35 linked with

Table 8.1. Details of sorghum SSRs used for foreground selection for marker-assisted backcrossing of stay-green QTLs at ICRISAT.

Staygreen QTLs	Sorghum chromosome[1]	Physical distance (Mbp)[2]	Linkage distance (cM)[3]
StgA markers			
Xtxp357	1	23.8	48.0
Xtxp43	1	50.3	65.8
Xtxp149	1	50.7	67.8
Xtxp88	1	50.7	67.7
Xtxp032	1	NA	77.1
Stg3 markers			
Xtxp019	2	NA	80.0
Xtxp013	2	56.0	82.0
Xtxp298	2	57.1	92.9
XSbAGB03	2	58.1	97.7
Mesocarp thickness		NA	89.9-104.4[4]
Xtxp001	2	61.4	126.0
Xtxp056.1	2	61.6	124.0
Xcup63	2	59.1	NA
Xgap084	2	NA	133.0
Xtxp286	2	NA	139.1
StgB markers			
Xtxp100	2	69.6	166.1
Xtxp7	2	70.3	167.5
Xtxp207	2	70.3	167.5
Xtxp296	2	71.1	171.8
Xcup26	2	70.3	167.5
Xtxp8	2	NA	193.8
BI139914	2	71.8	NA
Stg2 markers			
Xtxp033	3	51.4	61.3
Xgap236	3	52.3	62.8
Xtxp205	3	NA	68.1
Xtxp031	3	55.2	71.1
Xtxp183	3	NA	75.2
Xtxp336	3	55.4	73.2
R (pericarp color)	3	NA	87.9[4]
Xtxp120(Cba)	3	57.3	88.4
Xtxp231	3	NA	101.1
Xtxp59	3	58.5	103.2
Between Stg1 and Stg2			
Xtxp218	3	NA	110.9
Xtxp114	3	60.8	113.7
Stg1 markers			
Xtxp285	3	67.8	136.6
Xtxp034	3	69.7	147.8
Awns: A/a	3		157.9-161.6[4]
Stg4 markers			
Xtxp15	5	42.0	64.2
Xtxp14	5	42.3	64.1
Xtxp299	5	NA	64.1
Xtxp225	5	NA	59.4
Xtxp23	5	54.5	75.9

[1]Sorghum chromosome nomenclature as per Kim et al. (2005).

[2]Physical map distance (in MbP) as estimated by BLAST search of primer pair sequence of individual SSR with sorghum genome sequence as described in Ramu and Deshpande et al. (2010).

[3]Linkage distance (in cM) of SSRs as estimated in consensus map developed by Mace et al. (2009).

[4]Predicted linkage map position (putative interval in cM) of morphological markers as per Mace et al. (2010).

alleles for recessively inherited thick mesocarp. Similarly *Stg1* and *Stg2* had overlapping confidence intervals on SBI-03, and exhibited linkage with one of two epistatically interacting genes controlling pericarp pigmentation (*Y*/*y* on SBI-01 and *R*/*r* on SBI-03) and a gene controlling the absence/presence of awns (*A*/*a*). The Indian post-rainy season (*rabi*) sorghum has a very specific grain quality requirement of thin mesocarp with creamy/lustrous bold grain. Utilizing the intermediate BC_2-progeny for further backcrossing helped to generate segregation for desirable phenotypes at the mesocarp and pericarp loci, recombined with B35 alleles at stay-green QTL-flanking SSR markers. This permitted isolation of white-grained, thin mesocarp progenies such as 'K369white' in later backcross generations. Such "clean" single-QTL introgression lines can now be used individually as donor parents of specific stay-green QTLs shown to have the largest favourable phenotypic effects across several genetic backgrounds.

Similarly, for stay-green QTL introgression in S 35 = ICSV 111-background, backcrossing was continued up to BC_4 and BC_5 with complete foreground selection and incomplete background selection during each backcross generation. This has finally produced sets of $BC_nF_4/BC_{n+1}F_3$ progenies. These progenies have been evaluated across locations and years for their stay-green expression in field, pot, and lysimeter studies under fully-irrigated non-stress conditions and rain-fed or supplementally-irrigated terminal drought-stress conditions during the cool, dry post-rainy (*rabi*) season in peninsular India. Several progenies still segregated for plant height and flowering time, and so were advanced by selfing for several additional generations following pedigree selection. Even though it took four to five seasons more to isolate clean versions of stay-green QTL introgression lines in this genetic background, the fact that it was achieved with limited SSR availability in the early days of marker-assisted backcrossing must be appreciated. These recently developed and validated "clean" stay-green QTL introgression lines are now available in several genetic backgrounds and can be used as parents directly in breeding programs to feed the current and future product pipeline.

We also tried utilizing another well characterized stay-green source, namely Ethiopian landrace germplasm accession E 36-1 (Haussmann et al. 2002). The major problem for utilizing this source was lack of marker (SSR) polymorphism, as E 36-1 belongs to same set of of zera-zera landraces that have contributed extensively in development of the agronomically elite caudatum and guinea-caudatum derivatives across global sorghum breeding programs over the past fifty years. Further, the major stay-green QTL from E 36-1 maping to SBI-10 had a very large confidence interval and is linked with unfavourable alleles at neighboring shoot fly resistance component QTLs associated with seedling leaf blade trichome density and seedling glossines score. We are currently developing a fine-mapping population to break this unfavorable linkage by crossing an intermediate derivative of E 36-1 with E 36-1 alleles at the stay-green QTL on SBI-10 chromosome and a QTL introgression line with shoot fly resistance QTLs, utilizing newer marker techniques such as DArT, and single-nucleotide polymorphisms (SNPs) identified with a genotyping-by-sequencing (GbS) pipeline exploiting one of the new generation sequencing (NGS)-platforms (Elshire et al. 2011).

Fine-mapping populations for the best of stay-green QTLs validated in R 16 and S 35 backgrounds are being developed. These populations will be screened for several physiological parameters. This, along with newer marker systems such as DArTs and GbS-SNPs, will help to achieve deeper genome coverage for tracking important recombinants across large genomic region(s) present between the QTL-flanking SSR markers. These recombinants will help to identify near-isogenic lines for each of the stay-green QTLs, with no or minimal negative linkage drag, for direct utilization by breeding programs and as well for use in pyramiding multiple QTLs

to confirm whether or not their epistatic interactions are of economic importance. Currently a few selected best-performing stay-green QTL ingtrogression lines are being used in microarray assay and proteomics studies seeking up- and/or down-regulated genes and to identify gene products specific to these gene combinations.

Mechanisms Explaining Stay-Green

The Nitrogen and Carbohydrate Route

The potential benefit of stay-green was initially viewed from the angle of the maintenance of photosynthetic activity (Rosenow et al. 1983; Thomas and Smart 1993; Borrell et al. 2000). Results showed that indeed delayed senescence of fully-irrigated *Lolium temulentum* leaves would increase the carbon fixation by 11% over the entire life of the leaf, simply by delaying senescence by two days (Thomas and Howarth 2000). Other results also showed that levels of basal stem sugars (Duncan 1984) or carbohydrate contents (McBee et al. 1983) were higher in stay-green sorghum genotypes. Sanchez and colleagues (2002) also made the assumption that the delayed leaf senescence from stay-green would sustain photosynthetic activity. Hence, a number of studies have documented the worthiness of maintaining photosynthetic activity of the leaf for more time. While this may be true in situations where there is no water limitation, and where there is a light-capture interest of delaying leaf senescence, this may be less of a value in situations where water is limited and photosynthetic activity is bound to be regulated by stomata opening. Therefore, we would argue that the contribution of stay-green in terms of carbon fixation under water-stress conditions may likely be very limited.

Another approach to explaining stay-green differences has been to assess their role in the nitrogen balance of the plant. In crops producing grain, the most important nutrient required to fill up grain is nitrogen and it is remobilized from the N-rich leaf tissues (Sinclair and Vadez 2002). As rubisco, a central enzyme for the conversion of CO_2 into carbohydrates, accounts for about half the nitrogen in leaves of C3 plants and about 25% of the leaves of C4 plants, remobilizing N from rubisco and photosynthetic pigments implies that the photosynthetic rate is bound to decrease during grain filling. For instance, Borrell and Hammer (2000) showed that senescent and stay-green sorghum hybrids differed in the supply-demand balance for N, with stay-green having a shortfall in N that is about 25% lower than that in senescent hybrids, and explaining a slower rate of leaf senescence in the stay-green genotypes. A similar case was reported in maize, where a stay-green hybrid acquired up to 60% of its N supply during the grain-filling period, whereas a senescent hybrid acquired only 40% of its total N during the same period (Rajcan and Tollenaar 1999a). This showed the importance of maintaining N uptake during grain filling in staygreen lines across different species. Subedi and Ma (2005) also clearly showed that in both stay-green and senescent maize hybrids, stopping the supply of N from V8 to maturity dramatically accelerated the decrease in leaf greenness, measured by SPAD readings, compared to a treatment in which N supply was maintained. Another study also showed that under low-N conditions, there were genotypic differences in sorghum in the capacity to extract N from the soil profile (Nakamura et al. 2002). For these reasons, the N status of a plant is still considered an important factor in the expression of stay-green. Among the five cases of stay-green reviewed by Thomas and Howarth (2000), the type E stay-green is a case where senescence initiates at a similar date and follows a similar rate to a senescent type, but the higher initial N content in the leaves buffers the grain-filling-induced decline in leaf-N. That is, the current view is that an increased N uptake by roots during grain-filling leads to longer duration of leaves, and the higher specific leaf N (SLN) levels maintains the photosynthesic activity of these leaves at high levels for a longer period.

Given the tight N supply-demand balance, it has been shown that de-topping plant panicles can indeed delay leaf senescence. For example Rajcan and Tollenaar (1999b) showed that green leaf area maintenance was higher in situations of high source:sink ratios, achieved by partial or full prevention of maize cob fertilization. A similar phenomenon occurs in the case of genotypes having a poor grain yield potential and therefore a poor sink for N, leading to the expression of a "yield-resistant" stay-green phenotype. This association of stay-green expression with a low grain-yield trend has been one of the main criticisms of the use of the stay-green trait in crop improvement (Ludlow and Muchow 1990). However, Borrell and colleagues (2000) and Haussmann and colleagues (2002) showed a positive correlation between stay-green expression and grain yield under terminal drought-stress conditions; although in the latter study, one of the two RIL populations that was used showed no correlation between grain yield and stay-green expression. Tuinstra and colleagues (1997) also identified two QTLs for sorghum stay-green that co-mapped with grain-yield QTLs. So it appears that the relationship between stay-green expression and grain yield under terminal drought-stress conditions depends both on the environment and on the background that are considered. One of the future challenges will surely be to identify the genotypes and the environments in which stay-green expression is not at the expense of grain and/or stover yield potential. The potential of stay-green genotypes to accumulate N during the grain-filling period, provided that grain yield potential is not compromised, is the most promising hypothesis to explain an N effect on the expression of stay-green that could have agronomic relevance. Further work is needed to understand the processes that allow N absorption to be sustained during grain-filling, especially under water-limited conditions. As discussed in the next section, the capacity to absorb N during grain-filling under such conditions is bound to be closely related to water status issues (having water left in the profile to allow N absorption).

Addressing the Symptoms or Addressing the Causes?

Thomas and Howarth (2000) reviewed different modalities resulting in the display of a stay-green phenotype. They have suggested five ways to stay green. Of these, four were concerned with the rate and onset of pigment decline. Pigment degradation is a self-programmed process in plants during maturation and there are a number of factors affecting N remobilization that alter that natural process. For example, de-topping the panicle delays leaf senescence. However, even after removing the effect of N-status-altering processes, pigment degradation continues its natural way. Research initiated to counter that process found that enhancing cytokinin production delayed pigment degradation in tobacco leaves (Gan and Amasino 1995; Roitsch and Ehneβ 2000). A number of studies have followed in which transgenics were developed, which contained a gene contributing to enhanced cytokinin production to retard pigment degradation (Rivero et al. 2007; Peleg et al. 2011), and which were reported to be drought tolerant. While the maintenance of green leaf area may be beneficial in situations when water is not limiting, as earlier shown in *Lolium temulentum* (Thomas and Howarth 2000), the overall approach and hypothesis of its value under water-limited situation remains questionable. Plants exchange carbon dioxide for water through the stomata and under water-limited situations the degree of stomatal opening is what sets the photosynthetic rate, because of the absolute necessity to match stomatal opening to the limited water available. Therefore, the cytokinin-related maintenance of green pigmentation under water-limited conditions may be assigned to a "type-C" stay-green as defined by Thomas and Howarth (2000), that is, a type that stays green but in which the photosynthetic functionality is equivalent to a senescent line because of the effects of water limitation on stomatal opening. Although the work on the overexpression of cytokinins to retard leaf senescence is intriguing, we would argue that it may fit

situations of mild water stress or no water stress, where indeed maintaining longer leaf life could be beneficial.

Surprisingly, the past 25 years of research on the stay-green phenotype have only lately led to examination of the possible association between stay-green expression and plant water status, although some relation between stay-green expression and plant water status had been hypothesized early (Tuinstra et al. 1998). More surprising is the fact that several reports had shown that stay-green was likely associated with maintenance of root growth (Hatlitligil et al. 1984; MacKay and Barber 1986), with the hypothesis that enhanced root growth would contribute to enhanced N absorption. As we saw in the previous example, the maintenance of a functional stay-green under water-limited conditions, that is, a plant type having both green leaf area remaining *and* active photosynthetic activity, depends on having water available in the soil profile at the time of leaf senescence. The difficulty in testing this hypothesis is concerned with methods that can be precise enough to assess plant water extraction at a fairly late stage of plant growth when stay-green expression is at its maximum. Recently, a lysimetric system has been developed at ICRIDSAT-Patancheru (Vadez et al. 2008), which consists of long and large plastic tubes in which plants are grown with the spacing and soil exploration volume they would have in a natural field conditions. This system has allowed the measurement of the pattern of water uptake to support transpiration in several crops, including sorghum (Vadez et al. 2011), chickpea (Zaman-Allah et al. 2011), and peanut (Ratnakumar and Vadez 2011). Using this system, a set of pearl millet topcross hybrids contrasting for their level of terminal-drought tolerance were assessed under conditions of terminal-drought stress, imposed by stopping irrigation at flowering time. The results clearly showed that hybrids differed in their stay-green expression as the stress developed and showed highly significant correlation ($R^2 = 0.76$-0.79) with the water extracted three weeks after panicle emergence (Vadez et al. unpublished). These results have been confirmed in several experiments of pearl millet and offer an outstanding demonstration that stay-green directly relates to the water availability during the grain-filling period. One of the exciting challenges of the coming year, using that system, is to test the hypothesis, which could not be tested before, that maintaining water uptake during the grain-filling period would also indirectly drive N uptake during the same period. As seen above, several stay-green genotypes in different species have been shown to enhance N uptake during the post-anthesis period. Since N uptake requires that this nutrient be dissolved in water to be taken up, what remains to be establish is whether the higher N uptake could be a consequence of a higher water uptake.

Water in the soil profile can become available during the grain-filling period through several possible mechanisms. The most immediate one is the capacity to extract water. This has been shown in wheat (Manshadi et al. 2006), where a stay-green wheat genotype extracted more water from deeper layers of the soil profile than did a senescent line. Recently, the stay-green QTL *Stg1* in sorghum has also shown its capacity to enhance water uptake in senescent S 35 background (Vadez et al. 2011b). However, the effect of *Stg1* was not visible in the R 16 background. The likely explanation for this is the higher "baseline" capacity for extracting water in R 16 than in S 35. This highlights the importance for future research on stay-green to precisely decipher the mechanisms involved, and to determine whether any of these mechanisms are already available in intended target recurrent parent genotypes.

The case of pearl millet described above is interesting because the materials that differed in stay-green (Vadez et al. in preparation) did *not* differ in the total water extracted from the soil profile. In other words, stay-green in this case was not related to an effect on rooting. By contrast, other studies have showed that these materials vary in constitutive water-saving traits, that is, through a lower leaf conductance (Kholova

et al. 2010a) and a further restriction of leaf conductance under high vapor-pressure deficit conditions (Kholova et al. 2010b). The expression of these traits, at the vegetative stage in the absence of water stress, leaves water available in the soil profile eventually leading to stay-green expression differences (Vadez et al. in preparation). Similar findings have been reported in stay-green *Miscanthus* genotypes, where the stay-green genotype Sin-H6 appeared to have a lower leaf conductance (Clifton-Brown et al. 2002). In the case of pearl millet, several QTLs have been identified for these water-saving traits (Kholova et al. 2012). Interestingly, water-saving traits, measured in pots, and stay-green expression and yield measured under field conditions co-map to the same genomic regions (Sehgal et al. in preparation).

Other possibilities for saving water during the vegetative growth stage, before any stress occurs, involve the development of smaller leaf area. One recent report shows that having a faster leaf-appearance rate reduced tillering and then decreased the overall plant leaf area at anthesis. The effect was to decrease water use prior to anthesis, leading to higher grain yield under terminal drought conditions (van Oosterom et al. 2011). In our current work at ICRISAT-Patancheru, we have also demonstrated the capacity of certain stay-green QTLs from donor parent B35 to reduce the leaf size in S 35 background (Kholova et al. unpublished). However, it was also shown that faster leaf-appearance rate was sensitive to temperature and that the beneficial effects were reduced in higher temperature environments (van Oosterom et al. 2011). Similarly, leaf expansion is highly dependent on both the evaporative demand and soil moisture in maize (Reymond et al. 2003). Therefore, future challenges with the use of stay-green expression will also be to better understand how some of the explanatory mechanisms of stay-green, like the leaf-area development addressed here, respond to the environment. Unless this is precisely known, the prediction of the effect of stay-green mechanisms will be inaccurate and the use of stay-green in breeding will be a blind exercise at best. Therefore, the use of stay-green in the future will very likely evolve to introgressing genomic elements involved in its key mechanisms rather than introgressing QTLs for stay-green per se. This implies that a more thorough undrestanding is needed that can decipher the mechanisms underlying stay-green expression in sorghum (and other crops in which this trait might be found useful), and the interaction of these mechanisms with the environment. This work is on-going at ICRISAT in India and in Niger.

Advances in Sorghum Genomics and Applications for Stay-Green Research

Among the available marker systems, simple sequence repeat (SSR) markers gained breeders' interest for mapping and introgression of different traits in crop species because these markers are amenable to simple assays, multiplexing, and reproducibility, and more importantly are co-dominantly inherited. SSR markers have been greatly exploited for the mapping of different traits in sorghum, and the stay-green trait is no exception (Haussmann et al. 2002, Harris et al. 2007, Habyarimana et al. 2010, and Sabadin et al. 2012). The major limiting factor for utilization of SSR markers is their resolution power. Recent advances in sorghum genomics, including availability of an aligned sorghum genome sequence (Paterson et al. 2009), access to larger numbers of markers including both SSRs (e.g., Ramu et al. 2010) and DArTs (Mace et al. 2010), with very large numbers of GbS-SNPs on the way (Elshire et al. 2011, Nelson et al. 2011). Alignment of major trait genes and QTLs to integrated linkage and physical maps (Mace et al. 2011) has strengthened the foundation for better integration of molecular marker technologies in applied sorghum breeding programs.

With the invention of next generation sequencing (NGS) technologies, identification of a large number of markers, especially single

nucleotide polymorphism (SNPs), has become very inexpensive compared to other marker systems. Utilizing an Illumina NGS platform, Ed Buckler's lab at Cornell University has developed a technically very simple and highly multiplexed (96-plex/384-plex) method for rapidly and inexpensively sequencing large numbers of DNA samples, and subsequently analysing the sequencer output with an associated bioinformatics pipeline for genotyping germplasm of any species. This protocol is referred as genotyping-by-sequencing (GbS) (Elshire et al. 2011). For this procedure, genome complexity is reduced by digestion of each DNA sample with restriction enzymes, and the resulting restricted fragments are then ligated with sample-specific "barcodes," called "restriction site-associated DNA tags" (RAD tags), and the restricted, barcoded DNA samples are then multiplexed (at 48-, 96- or 384-plex) and subjected to "skim" sequencing to a depth of 0.1X. The resulting 66-base pair sequence reads (after sorting by barcode) are aligned to the reference genome sequence of BTx623 (Paterson et al. 2009) to identify SNPs with the help of customized bioinformatics pipelines. This analytical pipeline can readily be adapted for species lacking a reference-aligned genome sequence.

By employing GbS, ~265,000 SNPs have been identified for stay-green donor parents E 36-1 and B35 by aligning their skim sequence reads against the sorghum reference-genome sequence. The primary challenges involved in handling these large data sets are the need for substantial computational power. Analysis of combined field, pot, and lysimeter phenotype data sets for QTL introgression line sets and RIL populations is underway at present, and in the near future we expect to be able to identify genomic regions (major and minor effect QTLs) associated with putative components of the stay-green phenotype, with or without terminal-drought tolerance, in sorghum. The high marker-density and genome-wide coverage that is possible with this GbS-SNP platform will help us to identify SNPs closest to or inside the individual genes and/or regulatory elements associated with variation in stay-green phenotype. Once genomic regions associated with underlying mechanisms of the stay-green trait are identified, tagged SNPs (reduced representations of SNPs based on their linkage) can be identified and converted to a customized SNP assay using the BeadXpress platform (currently available in ICRISAT's Genomics Service Laboratory at Patancheru) or CAPS (cleaved amplified polymorphic sequences) markers. Such SNP markers can be used individually or in small multiplexes at a much lower cost than that required for genome-wide genotyping with the GbS platform and will be appropriate for use in foreground genotyping and identification of recombination events occurring in QTL-flanking regions during the transfer of this trait to desired sorghum recurrent-parent backgrounds. This will greatly improve the efficiency of introgression of the stay-green trait and its components, by reducing the number of breeding cycles required (for recurrent-parent background genotype recovery) and facilitating stacking of complementary stay-green alleles at various loci as may be needed for improved variety development.

Application of NGS tools such as GbS for dissecting complex traits such as stay-green at the DNA-sequence level will capture most of the functional factors of the genome related to trait expression. However, another application of NGS tools in RNA-sequencing (commonly referred as RNA-seq) will help to capture the regulatory elements (Ozsolak and Milos 2011). For a complex development trait such as stay-green, many plant growth and development pathways are involved, probably throughout the life cycle of the plant, for trait expression. Application of RNA-seq can help us understand the role of regulatory and transcription factors (including small RNA, micro RNA) and their interactions with other pathways. We hope to utilize recent advances in RNA-seq technologies with the recombinants identified from the ongoing fine-mapping exercise to move toward a better understanding of stay-green expression

associated with one or more stay-green QTLs in sorghum.

Use of Modeling to Manipulate Mechanisms Associated with Stay-Green

Among research areas attempting to address the food and feed demand of growing human and livestock populations living under conditions of harsh climate and erratic rainfalls across the semi-arid tropics, crop improvement efforts are not only particularly challenging but are also particularly promising. Despite the progress made in the field of crop breeding strategies, for example quantitative genetics, marker-assisted selection processes, improvement of trait-screening techniques, investigation of stress tolerance differences, and so forth, the progress made in the development of improved cultivars has been slow because of complex interactions of genoype and environmental factors, including management practices (i.e., the G × E × M problem). This slowed progress is partially because investigating these interactions in vivo requires years of precisely managed multi-locational field trials, which are extremely time- and cost-intensive, and often simply impossible to do properly covering all possible relevant environments.

This as yet unresolved G × E × M problem suggests that the existence of crop genotypes that can adapt to a broad range of stress environments is very unlikely and that breeding strategies for stressful environments should probably instead focus on the development of crop genotypes suited to particular environments. In recent years a pragmatic way appeared for at least beginning to decipher the complexity of G × E × M interactions by crop simulation modeling. This approach interlinks mechanistic knowledge of crop growth characteristics and allows estimates of crop productivity across the region(s) of interest. The crop model sensibility is highly dependent on knowledge of a given production environment (weather, soil) as well as knowledge of the crop, making it extremely useful as a guideline for "precision breeding," that is, analysis of the environments from the crops' perspective and development of genotypes possessing specific features that permit maximum utilization of environment potential. In other words, modeling allows reasonable diagnostics of environment-restricting factors, such as type of drought stress and probability that a crop will face a particular stress type at a specific location (Chenu et al. 2011, Hammer and Jordan 2007, Chapman et al. 2008).

Such knowledge can be used further for (1) in vivo selection and screening for crop traits providing putative adaptation in the well-defined target environmental conditions and management practices (2) in silico designing of virtual genotypes possessing hypothetical/existing traits and estimation of their benefits across time in the location of interest with a given suite of management practices. This approach has already been used for the characterization of environments for wheat and sorghum in Australia (Chenu et al. 2011, Hammer and Jordan 2007, Chapman et al. 2008) and there is an on-going effort to diagnose the sorghum production constraints using this methodology for winter cropping seasons (post-rainy season) of the semi-arid tropics in peninsular India. Here the modeling tool allowed, for the first time, the differentiation between various water-stress types, quantification of stress types' frequencies, and their effects on sorghum production across heterogeneous parts of major production regions (Kholova et al. in preparation). At the same time, substantial progress has been made in understanding the mechanisms contributing to drought adaptation (e.g., water utilization dynamics and efficiency, plant developmental dynamics, and N utilization – many of which may result in so called "stay-green" phenotypes; see section above on "Mechanisms Explaining Stay-Green"). The well-defined physiological basis of any genotypes' specific machinery can be simulated using such models and tested across a range of specific environments. In this way, modeling can help approximate the

benefits or possible negative trade-offs of any given trait/mechanism and thereby estimate its potential value for region-specific breeding programs.

As an example, the genotype-specific stomatal closure at high evaporative demand, one of the "water-saving" traits reviewed above, is a feasible proposition. The idea of intra-specific variability in VPD(vapor pressure deficit)-driven stomatal closure was proposed long ago (e.g., Squire 1979), although it was not until recently that this mechanism was explored by modeling (Sinclair et al. 2005; 2010). In the meantime, variability for VPD response was found across other crop species (e.g., groundnut: Devi et al. 2010; pearl millet: Kholova et al. 2010; and chickpea: Zamman-Allah et al. 2011). There is evidence that similar variability exists in sorghum (Gholipoor et al. 2010), and, indeed the model suggests that this mechanism will lead to desired improvement of the *post-rainy season* sorghum cultivation in terms of absolute grain yield as well as yield stability (our work in progress).

Conclusions

Much progress has been achieved in the deciphering of genomic regions responsible for the expression of the stay-green phenotype in sorghum. This information has been used in a number of breeding programs worldwide, mostly through marker-assisted backcrossing to move donor parent alleles for this trait into otherwise locally-adapted agronomically elite open-pollinated varieties and/or hybrid parental lines. However, the physiological mechanisms underlying the expression of stay-green are less clear. It appears now that the availability of water during the grain-filling period, when stay-green is scored, is the most likely candidate, and may likely be additive to N absorption after anthesis. The sources of water availability could be several, including water-saving traits but also possibly deeper rooting capacity. Clearly, the more recent progress pointing at clear mechanisms explaining the stay-green phenotype will likely reorient the breeding of stay-green towards the breeding for its most important components. This will require a "re-mapping" of these explanatory traits and possibly the identification of new/better donor sources for these traits than the donors for stay-green that are currently in use. The recent dramatic progress achieved in terms of density of marker coverage across the full nuclear genome will be extremely useful for precisely mapping these explanatory traits. Several of the putative traits leading to stay-green expression closely interact with the environment. Therefore their manipulation will also require a thorough understanding of these interaction effects, and the use of crop simulation modeling will then become increasingly important to help the breeding program navigate the complexity of plant-trait interactions with the environment (including crop management practices), in order to better target the type of trait combinations needed for each specific target environment.

Acknowledgement

The authors are grateful for a grant from ACIAR (CIM/2007/120) (Improving postrainy sorghum varieties to meet the growing grain and fodder demand in India), which has supported a substantial portion of the work described here.

References

Allard RW. 1999. Principles of Plant Breeding, 2nd ed. New York, NY: Wiley.

Armstead I, Donnison I, Aubry S, Harper J, Hörtensteiner S, James C, Mani J, Moffet M, Ougham H, Roberts L, Thomas A, Weeden N, Thomas H, King I. 2006. From crop to model to crop: Identifying the genetic basis of the staygreen mutation in the *Lolium* /*Festuca* forage and amenity grasses. New Phytologist 172, 592–597.

Borrell A, Hammer G, van Oosterom E. 2001. Stay-green: A consequence of the balance between supply and demand for nitrogen during grain filling. Annals of Applied Biology 138, 91–95.

Borrell AK, Hammer GL, Henzell RG. 2000b. Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. Crop Science 40, 1037–1048.

Borrell AK, Hammer GL. 2000. Nitrogen dynamics and the physiological basis of stay-green in sorghum. Crop Science 40, 1295–1307.

Borrell AK, Henzell RG, Douglas ACL. 1996. Visual rating of green leaf retention is highly correlated with measured green leaf area in sorghum. In: Foale MA et al. (eds.). Proceedings of the Third Australian Sorghum Conference, Tamworth, 20 to 22 Feb. 1996. Occasional Publication No. 93. Australian Institute of Agricultural Science, Melbourne, pp. 323–326.

Burke JJ, Franks CD, Burow G, Xin Z. 2010. Selection system for the stay-green drought tolerance trait in sorghum germplasm. Agronomy Journal 102(4). 1118–1122.

Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF, Chapman SC. 2011. Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. *Journal of Experimental Botany* 62(6), 1743–1755.

Clifton-Brown JC, Lewandowski I, Bangerth F, Jones MB. 2002. Comparative response to water stress in stay-green, rapid- and slow senescing genotypes of the biomass crop *Miscanthus*. New Phytologist 154, 335–345. doi:10.1046/j.1469-8137.2002.00381.x

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT. 1999. Mapping of post-flowering drought resistance traits in grain sorghum: Association between QTLs influencing premature senescence and maturity. Molecular and General Genetics 262, 579–588.

Devi MJ, Sinclair TR, Vadez V. 2010. Genotypic variation in peanut for transpiration response to vapor pressure deficit. Crop Science 50, 191–196.

Duncan RR. 1984 The association of plant senescence with root and stalk diseases in sorghum. In: Mughogho LK (ed.) Sorghum root and stalk rots, a critical review. ICRISAT, Patancheru, AP, India, pp. 99–110.

Duncan RR, Bockholt AJ, Miller FR. 1981. Descriptive comparison of senescent and non-senescent sorghum genotypes. *Agronomy Journal* 73, 849–853.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5), e19379. doi:10.1371/journal.pone.0019379.

Gan S and Amasino RM. 1995. Inhibition of leaf senescence by autoregulation of leaf production of cytokinin. Science 270, 1986–1988.

Gholipoor M, Vara Prasad PV, Mutava RN, Sinclair TR. 2010. Genetic variability of transpiration response to vapour pressure deficit among sorghum genotypes. Field Crops Research 119(1), 85–90.

Gupta PK, Varshney RK. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113, 163–185.

Gupta PK, Varshney RK, Sharma PC, Ramesh B. 1999. Molecular markers and their applications in wheat breeding. Plant Breeding 118, 369–390.

Hamidou F, Ratnakumar P, Halilou O, Mponda O, Kapewa T, Monyo E, Faye I, Ntare BR, Nigam SN, Upadhyaya HD, Vadez V. 2012. Selection of intermittent drought tolerant lines across years and locations in the reference collection of groundnut (*Arachis hypogaea* L.). *Field Crops Research* 126, 189–199 doi:10.1016/j.fcr.2011.10.009.

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, Mullet J. 2007. Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. Journal of Experimental Botany 58, 327–338.

Haryarimana E, Lorenzoni C, Busconi M. 2010. Search for new stay-green sources *in Sorghum bicolor* (L.) Moench. Maydica 55, 187–194.

Hammer G, Chapman S, van Oosterom E, Podlich D 2005. Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. Australian Journal of Agricultural Research 56, 947–960.

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, Mullet J. 2007. Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. Journal of Experimental Botany 58, 327–338.

Hash CT, Bhasker Raj AG, Lindup S, Sharma A, Beniwal CR, Folkertsma RT, Mahalakshmi V, Zerbini E, Blümmel M. 2003. Opportunities for marker-assisted selection (MAS) to improve the feed quality of crop residues in pearl millet and sorghum. Field Crops Research 84, 79–88.

Hatlitligil MB, Olson RA, Compton WA. 1984. Yield, water use, and nutrient uptake of corn hybrids under varied irrigation and nitrogen regimes. Fertilizer. Research. 5, 321–333.

Haussmann BIG, Hess DE, Seetharama N, Welz HG, Geiger HH. 2002a. Construction of a combined sorghum linkage map from two recombinant inbred populations using AFLP, SSR, RFLP, and RAPD markers, and comparison with other sorghum maps. Theoretical and Applied Genetics 105, 629–637.

Haussmann BIG, Mahalakshmi V, Reddy BVS, Seetharama N, Hash CT, Geiger HH. 2002b. QTL mapping of stay-green in two sorghum recombinant inbred populations. Theoretical and Applied Genetics 106, 133–142.

Kassahun B, Bidinger FR, Hash CT, Kuruvinashetti, MS. 2010. Stay-green expression in early generation sorghum [*Sorghum bicolor* (L.) Moench] QTL introgression lines. Euphytica 172(3), 351–362. DOI: 10.1007/s10681-009-0108-0.

Kebede H, Subudhi PK, Rosenow DT, Nguyen HT. 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theoretical and Applied Genetics 103, 266–276.

Kholová J, Hash CT, Kocóvá M, Vadez V. 2010b. Constitutive water conserving mechanisms are correlated with the terminal drought tolerance of pearl millet (*Pennisetum americanum* L.). Journal of Experimental Botany 61, 369–377.

Kholová J, Hash CT, Lava Kumar P, Yadav SR, Kočová M, Vadez V. 2010a. Terminal drought-tolerant pearl millet [*Pennisetum glaucum* (L.) R. Br.] have high leaf ABA and limit transpiration at high vapor pressure deficit. Journal of Exerimental Botany 61(5), 1431–1440.

Kholová J, Nepolean T, Hash CT, Supriya A, Rajaram V, Senthilvel S, Kakkera A, Yadav RS, Vadez V. 2012. Water saving traits co-map with a major terminal drought tolerance quantitative trait locus in pearl millet (*Pennisetum glaucum* (L.) R. Br.). Molecular Breeding DOI: 10.1007/s11032-012-9720-0.

Kim JS, Klein PE, Klein RR, Price HJ, Mullet JE and Stelly DM. 2005. Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169–1173.

Ludlow MM, Muchow RC. 1990. A critical evaluation of traits for improving crop yields in water-limited environments. Advances in Agronomy 43, 107–153.

Mace ES and Jordan DR. 2010. Location of major effect genes in sorghum (*Sorghum bicolor* (L.) Moench). Theoretical and Applied Genetics, 121: 1339–1356.

Mackay AD, Barber SA 1986. Effect of nitrogen on root growth of two genotypes in the field. Agronomy Journal 78, 699–703.

Mackill DJ, Ni J. 2000. Molecular mapping and marker assisted selection for major-gene traits in rice. In: Khush GS, Brar DS, Hardy B (eds.) Proc. Fourth International Rice Genetics Symposium. International Rice Research Institute, Los Baños, The Philippines, pp. 137–151.

Mahalakshmi V, Bidinger FR. 2002. Evaluation of stay-green sorghum germplasm lines at ICRISAT. Crop Science 42(3), 965–974.

Manschadi AM, Christopher JT, Peter de Voil P, Hammer GL. 2006. The role of root architectural traits in adaptation of wheat to water-limited environments. Functional Plant Biology 33, 823–837.

McBee BB, Waskom RM, Miller FR, Creelman RA. 1983. Effect of senescence and non-senescence on carbohydrates in sorghum during late kernel maturity states. Crop Science 23, 372–376.

McBee GG, Miller FR. 1982. Carbohydrates in sorghum culms as influenced by cultivars, spacing, and maturity over a diurnal period. Crop Science 22, 381–385.

Mohler V, Singrun C. 2004. General considerations: Marker-assisted selection. In: Lorz H, Wenzel G (eds.) Biotechnology in Agriculture and Forestry, Vol. 55: Molecular Marker Systems. Springer, Berlin, pp. 305–317.

Mutava RN, Prasad PVV, Tuinstra MR, Kofoid KD, Yu J. 2011. Characterization of sorghum genotypes for traits related to drought tolerance. Field Crops Research 123(1), 10–18. doi: 10.1016/j.fcr.2011.04.006.

Nakamura T, Adu-Gyamfi JJ, Yamamoto A, Ishikawa S, Nakano H, Ito O. 2002. Varietal differences in root growth as related to nitrogen uptake by sorghum plants in low-nitrogen environment. Plant and Soil 245, 17–24.

Nelson JC, Wang S, Wu Y, Li X, Antohy G, White FF, Yu J. 2011. Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. BMC Genomics 2011, 12:352. http://www.biomedcentral.com/1471-2164/12/352

Ougham H, Hörtensteiner S, Armstead I, Donnison I, King I, Thomas H, Mur L. 2008. The control of chlorophyll catabolism and the status of yellowing as a biomarker of leaf senescence. Plant Biology 10(Suppl. 1), 4–14.

Ozsolak F, Milos MM. 2011. RNA sequencing: Advances, challenges and opportunities. Nature Reviews Genetics 12, 87–98.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS. 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457, 551–556.

Peleg ZV, Reguera M, Tumimbang E, Walia H, Blumwald E. 2011. Cytokinin-mediated source/sink modifications improve drought tolerance and increase grain yield in rice under water-stress. Plant Biotechnology Journal, 1–12. doi: 10.1111/j.1467-7652.2010.00584.x.

Rajcan I, Tollenaar M. 1999a. Source:sink ratio and leaf senescence in maize. I. Dry matter accumulation and partitioning during kernel filling. Field Crops Research 60, 245–253.

Rajcan I, Tollenaar M. 1999b. Source:sink ratio and leaf senescence in maize. II. Nitrogen metabolism during kernel filling. Field Crops Research 60, 255–265.

Ramu P, Deshpande SP, Senthilvel S, Jayashree B, Billot C, Deu M, Reddy LA, Hash CT. 2010. *In silico* mapping of important genes and markers available in the public domain for efficient sorghum breeding. Molecular Breeding 26, 409–418.

Ratnakumar P, Vadez V. 2011. Groundnut (*Arachis hypogaea* L.) genotypes tolerant to intermittent drought maintain a high harvest index and have small leaf canopy under stress. Functional Plant Biology 38, 1016–1023. http://dx.doi.org/10.1071/FP11145.

Reddy BVS, Ramesh S, Sanjana Reddy P, Ashok Kumar A. 2009. Genetic options for drought management in sorghum. Plant Breeding Reviews 31:189–222.

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F. 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiology 131, 664-675.

Rivero RM, Kojima M, Gepstein A, Sakakibara H, Mittler R, Gepstein S, Blumwald E. 2007. Delayed leaf senescence induces extreme drought tolerance in a flowering plant. Proceedings of the National Academy of Sciences (USA) 104, 19631–19636.

Roitsch T, Ehneß R. 2000. Regulation of source/sink relations by cytokinins. Plant Growth Regulation, 32, 359–367.

Rosenow DT, Clark LE. 1981. Drought tolerance in sorghum. In: Loden HD, Wilkinson D (eds.) Proceedings of the 36th Annual Corn and Sorghum Industry Research Conference. Chicago, IL. American Seed Trade Association, Washington, DC, pp. 18–30.

Rosenow DT, Quisenberry JE, Wendt CW, Clark LE. 1983. Drought tolerant sorghum and cotton germplasm. Agriculture and Water Management 7, 207–222. doi:10.1016/0378-3774(83)90084-7.

Sabadin P, Malosetti M, Boer M, Tardin F, Santos F, Guimarães C, Gomide R, Andrade C, Albuquerque P, Caniato F, Mollinari M, Margarido G, Oliveira B, Schaffert R, Garcia A, van Eeuwijk F, Magalhaes J. 2012. Studying the genetic basis of drought tolerance in sorghum by managed stress trials and adjustments for phenological and plant height differences. Theoretical and Applied Genetics doi: 10.1007/s00122-012-1795-9.

Sanchez AC, Brar DS, Huang N, Li Z, Khush GS. 2000. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Science 40, 792–797.

Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT. 2002. Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). Plant Molecular Biology 48, 713–726.

Shan X, Blake TK, Talbert LE. 1999. Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. Theoretical and Applied Genetics 98, 1072–1078.

Sharp PJ, Johnston S, Brown G, McIntosh RA, Pallotta M, Carter M, Bariana HS, Khatkar S, Lagudah ES, Singh RP, Khairallah M, Potter R, Jones MGK. 2001. Validation of molecular markers for wheat breeding. Australian Journal of Agricultural Research. 52, 1357–1366.

Sinclair TR, Messina CD, Beatty A, Samples M. 2010. Assessment across the United States of the benefits of altered soybean drought traits. Agronomy Journal 102(2), 475–482.

Sinclair TR, Vadez V. 2002. Physiological traits for crop yield improvement in low N and P environment. Plant and Soil 245(1), 1–15.

Sinclair TR, Hammer GL, van Oosterom EJ. 2005. Potential yield and water-use efficiency benefits in sorghum from limited transpiration rate. Functional Plant Biology 32, 945–952.

Squire GR. 1979. The response of stomata of pearl millet (*Pennisetum typhoides* S. and H.) to atmospheric humidity. Journal of Experimental Botany 118, 925–933.

Subedi KD, Ma BL 2005 Nitrogen uptake and partitioning in staygreen and leafy maize hybrids. Crop Science 45, 740–747.

Subudhi PK, Rosenow DT, Nguyen HT. 2000. Quantitative trait loci for the stay green trait in sorghum (*Sorghum bicolor* L. Moench): Consistency across genetic backgrounds and environments. Theoretical and Applied Genetics 101, 733–741.

Tao YZ, Henzell RG, Jordan DR, Butler DG, Kelly AM, McIntyre CL. 2000. Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments. Theoretical and Applied Genetics 100, 1225–1232.

Thomas H, Howarth CJ. 2000. Five ways to stay green. Journal of Experimental Botany 51, 329–337.

Thomas H, Smart CM. 1993. Crops that stay green. Annals of Applied Biology 123, 193–219.

Tuinstra MR, Ejeta G, Goldsbrough P. 1998. Evaluation of near isogenic sorghum lines contrasting for QTL markers associated with drought tolerance. Crop Science 38, 835–842.

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G. 1996. Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum. Crop Science 36, 1337–1344.

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G. 1997. Genetic analysis of post-flowering drought tolerance and components of grain development of *Sorghum bicolor* (L.) Moench. Molecular Breeding 3, 439–448.

Vadez V, Deshpande SP, Kholova J, Hammer GL, Borrell AK, Talwar HS, Hash CT. 2011a. Staygreen QTL effects on water extraction and transpiration efficiency in a lysimetric system: Influence of genetic background. Functional Plant Biology 38, 553–566.

Vadez V, Krishnamurthy L, Hash CT, Upadhyaya HD, Borrell AK. 2011b. Yield, transpiration efficiency, and water use variations and their relationships in the sorghum reference collection. Crop & Pasture Science 62(8), 1–11. DOI: 10.1071/CP11007.

Vadez V, Rao S, Kholova J, Krishnamurthy L, Kashiwagi J, Ratnakumar P, Sharma KK, Bhatnagar-Mathur P, Basu PS. 2008. Roots research for legume tolerance to drought: Quo vadis? Journal of Food Legumes 21(2), 77–85.

Van Oosterom E., Borrell AK, Deifel KS, Hammer GL. 2011. Does increased leaf appearance rate enhance adaptation to postanthesis drought stress in sorghum. Crop Science 51, 2728–2740.

Van Oosterom EJ, Jayachandran R, Bidinger FR. 1996. Diallel analysis of the stay-green trait and its components in sorghum. Crop Science 36, 549–555.

Witcombe JR, Virk DS. 2001. Number of crosses and population size for participatory and classical plant breeding. Euphytica 122, 451–462.

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

Zaman-Allah M, Jenkinson DM, Vadez V. 2011. A conservative pattern of water use, rather than deep or profuse rooting, is critical for the terminal drought tolerance of chickpea. Journal of Experimental Botany 62(12), 4239–4252.

Chapter 9

Genetic Improvement of Grain Quality in *Japonica* Rice

Kiyosumi Hori and Masahiro Yano

Abstract

Rice grain quality largely determines market price and consumer acceptance. Rice grain quality is determined by a set of complex traits encompassing a wide range of physical and cooking characteristics. In general, *japonica* rice consumers prefer short, translucent grains that are sticky and soft when cooked. To improve the grain quality of *japonica* rice cultivars, many genetic studies of grain-quality traits have been conducted and have revealed the involvement of multiple genes and quantitative trait loci (QTLs). Grain size QTLs have been identified, and some responsible genes have been isolated by means of map-based cloning. A number of QTLs were found for grain chalkiness traits induced by high temperatures during the grain ripening stage. Several of these QTLs have been fine-mapped, but no genes have been cloned yet. The pasting characteristics and texture of cooked rice are largely influenced by grain components of amylose, amylopectin, and proteins. Genetic analysis of rice endosperm mutants has revealed many genes related to starch synthesis, such as *GBSSI* (*waxy*) and *SSIIa* (*alk*). Recently we detected QTLs for eating quality, as evaluated by sensory testing of a cooked rice experimental population derived from a cross between *japonica* rice cultivars. Fine-mapping of these QTLs and gene isolation could reveal novel aspects of eating quality, since the QTLs were not associated with differences in amylose or protein content. Further progress in research, such as improvement of evaluation instruments and genetic dissection by means of novel reverse genetic approaches, is still needed to explain and improve grain-quality traits in *japonica* rice cultivars.

Introduction

Rice (*Oryza sativa* L.) is the most important food crop in the world, as it is a staple food for more than half of the world's population. Rice cultivars are divided into two major subspecies: *indica* and *japonica*. The two subspecies have been historically classified based on morphological traits including seed shape, grain shattering, and plant architecture (Kato et al. 1928; Matsuo 1952), and were recently classified by using whole-genome DNA polymorphism analysis (Kovach et al. 2009; Ebana et al. 2010; Famoso et al. 2011). *Indica* rice is produced worldwide and is more

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

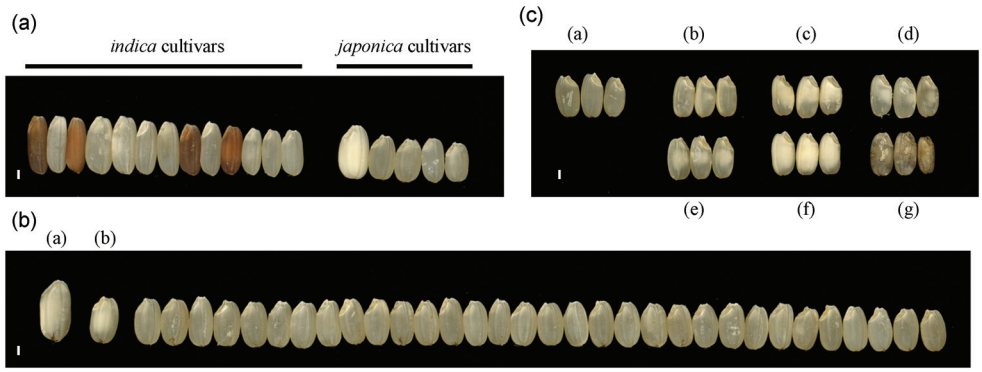


Fig. 9.1. (a) Husked rice grains of *indica* (long-grain) and *japonica* (short-grain) rice cultivars. (b) Grain-size variation in Japanese *japonica* rice cultivars of non-glutinous rice: (a) a large-grain cultivar, Oochikara; (b) a Japanese brewing cultivar, Gohyakumangoku. Rice grains were supplied by Dr. Ebana of the National Institute of Agrobiological Sciences Genebank. (c) Appearance of chalkiness in grains of (a) normal type (no chalkiness), (b) white-back type, (c) white-based type, (d) white-belly type, (e) white-core type, (f) milky-white type, and (g) abortive type. *White scale bars* represent 1 mm. For a color version of this figure, please refer to the color plate.

prevalent than *japonica* rice, but *japonica* rice is preferentially grown and consumed in the northern parts of the Asian rice cultivation region, which generally lie in the temperate region. *Japonica* rice cultivars are predominantly grown in Japan, Korea, some parts of China, southern Europe, Australia, and California (USA) (Mackill 1995). The area of *japonica* rice grown in China has expanded over the past two decades, from 11% of the total rice cultivation area in 1980 to 29% in 2000. Consumer preferences and demand have increased *japonica* rice production in both northern and southern China (Wei et al. 2009). In Japan, South Korea, and Taiwan, most of the rice grown is *japonica* rice. Southern Europe is the largest area outside Asia where *japonica* rice is consumed.

Rice cultivars are required to show high yield and strong resistance to biotic and abiotic stresses. Of late, grain quality has become the primary consideration of consumers, the food industry, and seed producers (Champagne et al. 1999; Fitzgerald et al. 2009). Grain quality in rice largely determines market price and consumer acceptance, so it is a major target in rice breeding programs (Peng and Khush 2003; Liu et al. 2008). An understanding of the genetic factors that influence rice grain quality is necessary to efficiently develop new cultivars with the high quality demanded by consumers. In fact, "grain quality" represents a set of complex traits, each controlled by multiple genes (or quantitative trait loci, QTLs) and largely influenced by environmental factors. Since grain quality in *japonica* rice shows varietal differences (Figure 9.1) (Nakagahra et al. 1997; Kim 2009), many researchers have been trying to identify the associated genetic factors. Before the development of molecular marker technology in the 1990s, most genetic studies focused on the characterization of mutant lines that differed in grain appearance (Takeda and Saito 1980; Satoh and Omura 1981). Even now, many researchers study mutants with clear phenotypic differences in particular traits (Ryoo et al. 2007; Satoh et al. 2008; Fujita et al. 2009; She et al. 2010). During the last two decades, genome-sequencing projects have provided new and powerful genetic tools such as DNA markers (McCouch et al. 2002; IRGSP 2005; Yamamoto et al. 2010; Nagasaki et al. 2010; Arai-Kichise et al. 2011). The use of DNA markers has facilitated the identification and molecular cloning of rice genes involved in morphological and physiological traits, including grain quality (Yano 2001). In addition, progress

in plant breeding, molecular biology, and plant physiology is enhancing our understanding of the gene pathways that control grain-quality traits. These new research strategies will not only advance our understanding of the molecular mechanisms, but will also lead to high-quality rice through efficient and targeted grain improvement.

In this chapter, we review (1) component traits of grain and eating quality, such as grain shape, grain chalkiness, and contents of seed storage components; and (2) genetic analyses related to grain quality (QTL detection and gene cloning) in *japonica* rice. We also discuss new approaches and prospects for the future of grain-quality improvement, including advances in phenotyping methods and genomics research.

Component Traits of Rice Quality

Grain-quality traits in rice encompass a wide range of characteristics that can be classified as pertaining to physical appearance and cooking characteristics (Table 9.1). The optimum score for each grain-quality trait varies according to the preferences of local consumers. The physical appearance of the rice grain includes its shape (length, width, and thickness), surface whiteness, and grain chalkiness (degree of translucence). These physical traits are immediately obvious to consumers and are among the major factors defining market value. Cooking characteristics typically include hydration and temperature during cooking time, textural properties of cooked rice, human evaluation for sensory qualities of cooked rice, the ability to remain soft for several hours after cooking, and aroma retention after cooking. Textural properties of cooked rice grain are important for consumers of *japonica* rice, who prefer strong stickiness and softness of cooked rice (Okabe 1979; Kim 2009; Sun et al. 2011). The types and amounts of each grain component, such as starch, protein, lipids, oligosaccharides, amino acids, vitamins, minerals, and secondary metabolites, affect the various cooking characteristics, sensory qualities, and nutritional value.

Physical Appearance of Rice Grain

Grain Shape

Grain shape is a critical aspect of rice grain quality. Grain shape affects not only grain yield, but also cooking and processing qualities (Fan et al. 2006; Song et al. 2007; Shomura et al. 2008). In the United States, rice cultivars are classified into several categories according to grain shape. Long-grain rice is approximately three times as long as it is wide. Short-grain rice is less than two times as long as it is wide. Medium-grain rice lies in between (Bergman et al. 2004).

Table 9.1. Components of grain quality in rice. Each component is intimately interrelated with the others to form overall grain quality

Major classification	Small classification	Specific example
Physical appearance	Grain shape	Seed size, length, width, thickness and weight
	Grain chalkiness	Degree of translucence endosperm
	Husking and milling quality	Amount of cracked grain after husking or milling
Cooking characteristics	Cooking property	Gelatinization temperature, gel consistency and texture of cooked rice grain
	Sensory quality	Glossiness, stickiness, hardiness and taste of cooked rice evaluated by persons
	Component of endosperm	Starch (amylose and amylopectin), oligosaccharides, protein, amino acids, lipids, vitamins, minerals and second metabolite products
	Aroma	Amount of 2-acetyl-1-pyrroline (2-AP)
	Nutritional Value	Amount of amino acid, vitamin and mineral

Medium- and short-grain rice cultivars are generally preferred by people from Japan, northern China, and North and South Korea, where *japonica* rice is typically grown. Slender grains are preferred in many areas where *indica* rice is grown (Sundaram et al. 2008).

Many genetic studies of grain shape in rice have confirmed that seed size, shape, and weight are all under polygenic control. Numerous QTLs associated with each aspect of grain size have been identified in diverse populations derived from crosses between *indica* and *japonica* cultivars over the past 20 years (Yonemaru et al. 2010; Youens-Clark et al. 2011). Recently, five grain-shape genes (*GS3*, *GW2*, *qSW5*, *GIF1*, and *GS5*) have been cloned and characterized (Fan et al. 2006; Song et al. 2007; Shomura et al. 2008; Wang et al. 2008; Li et al. 2011). Cultivars carrying the recessive alleles of *GS3*, *qSW5*, *GW2*, and *GIF1* have longer, wider, and heavier seeds than those carrying the wild-type alleles. In contrast, *GS5* functions as a positive and dominant regulator of grain size by increasing cell number and cell size and leading to enhanced length of the grain. Endosperm size in rice is largely determined by the dimensions of the lemma and palea (Li et al. 2004). Each of the five subpopulations of rice cultivars (*indica*, *aus*, *tropical japonica*, *temperate japonica*, and *basmati*) carry different combinations of alleles of grain-size genes, and the *indica* rice population contains significantly greater genetic variation for grain size than the others (Takano-Kai et al. 2009). The large number of possible allele combinations makes it possible to obtain a range of grain sizes to satisfy the diverse quality requirements of consumers. By creating novel allele combinations or introducing *indica* rice alleles into *japonica* rice cultivars, we could make changes and improvements to rice grain shape within the range acceptable to *japonica* rice consumers. To this end, Japanese breeders have developed the large-grain cultivars Oochikara (Figure 9.1b) and BG1 from crosses between *indica* and *japonica* cultivars (Kobayashi et al. 1990; Song et al. 2007).

Husking and Milling Quality

Assessing the appearance of husked and milled (polished) rice grains is necessary to determine the grade and price of harvested grain (Bergman et al. 2004). The grade is reduced by the presence of cracked, discolored, or immature grains. The majority of consumers prefer well-milled rice that has little to no bran remaining, despite the fact that brown rice contains more protein, lipids, vitamins, minerals, and phytochemicals with potential health benefits than milled rice (Bergman et al. 2004). Previous studies have suggested that grain shape, grain chalkiness, hull and bran diffusivity, and endosperm chemical components may affect the frequency of grain breakage during husking and milling (Bhattacharya 1980; Juliano et al. 1993; Sarker et al. 1996). Environmental factors are known to have a considerable impact on the frequency of grain breakage during husking and milling (Jodari and Linscombe 1996). For example, rain just before harvest increases the moisture content of grain and promotes grain fissuring (Lan and Kunze 1996). Takita (1992) reported varietal differences in cracked grain ratio. Rice breeders have tried to select cultivars showing a low frequency of cracked, discolored, and immature grains as breeding materials to develop novel rice cultivars. However, little is known about the genetic contribution to the number of cracked grains produced during the husking and milling processes.

Grain Chalkiness

Chalkiness is an important aspect of grain appearance that affects consumer acceptance. Chalky grains have opaque spots in various regions of the endosperm (Lisle et al. 2000; Fitzgerald et al. 2009). Chalkiness is classified into five types based on the positions of the opaque spots: white-belly, white-back, white-based, white-core, and milky-white (Figure 9.1c). In many cases, the presence of grain chalkiness leads to a downgrade in grain quality

and retail price. However, some consumers prefer white-core chalkiness and whole-grain chalkiness in certain *japonica* rice cultivars (Tashiro and Wardlaw 1991; Zakaria et al. 2002). For example, white-core chalkiness is preferred in Arborio-style cultivars and Japanese brewing-rice cultivars; and glutinous rice, which is very soft and sticky after cooking, has a completely opaque endosperm.

A wide range of variation in the level of grain chalkiness among *japonica* cultivars has been observed (Wan et al. 2005; Cheng et al. 2005; Kobayashi et al. 2007; Tabata et al. 2007). The frequencies of white-belly and milky-white types are increased by low solar radiation and large numbers of glumes per panicle. The frequencies of white-back and white-based types are increased by high humidity and low nutrient availability at the ripening stage (Yamakawa et al. 2007). Grain chalkiness is also increased by high temperature and large numbers of glumes per panicle (Yamakawa et al. 2007; Ishimaru et al. 2009): it is currently being studied by many researchers because of the potential for an increase in chalkiness caused by higher temperatures during the ripening stage resulting from global warming. Zhang et al. (2008) reported that alternate wetting and severe soil drying during grain maturity significantly reduced grain chalkiness and improved grain quality. Chalkiness is also negatively correlated with grain size: slender and small rice grains tend to have a lower frequency of chalkiness (Wan et al. 2005).

Starch granules are smaller and more loosely packed in the chalky areas of the rice endosperm than in the translucent areas (Lisle et al. 2000; Fitzgerald et al. 2009). Scanning electron microscopy has revealed incomplete grain-filling or progressive decomposition on the surface of starch granules in chalky endosperm (Zakaria et al. 2002). Therefore, many researchers have focused on the processes of starch synthesis and degradation (Bhattacharya 1979; Tashiro and Wardlaw 1991; Lisle et al. 2000; Patindol and Wang 2003; Wan et al. 2005; Cheng et al. 2005; Woo et al. 2008). High temperature decreases the expression of starch synthesis genes and the enzymatic activities of the resulting proteins (Umemoto and Terashima 2002; Jiang et al. 2003); consequently, amylose content is reduced and the branching structure of amylopectin is changed (Asaoka et al. 1984; Umemoto et al. 1999; Umemoto and Terashima 2002; Yamakawa et al. 2007). Several mutants of the starch synthesis genes *GBSSI*, *BEIIb*, and *PPDKB* show abnormal features such as floury or opaque endosperm (Satoh and Omura 1981; Nishi et al. 2001; Kang et al. 2005). Overexpression of the starch-degrading enzymes encoded by *Amy1A* and *Amy3D* results in increasing milky endosperm, unless the grains are heat-treated (Asatsuma et al. 2006). Moreover, *japonica* cultivars showing low levels of chalkiness contain high levels of nonstructural carbohydrate (NSC) in the vegetative tissues of the plants (Morita and Nakano 2011). High NSC content at the grain-filling stage would help to minimize grain chalkiness. These observations support the hypothesis that source–sink interactions are important in the development of grain chalkiness.

To perform comprehensive comparisons of storage components associated with chalkiness in grain, several researchers have used proteome and metabolome analyses (Yamakawa et al. 2007; Lin et al. 2010; Yamakawa and Hakata 2010). In these analyses, the levels of allergen-related proteins, translation elongation factors, and heat shock proteins increased under high temperature, whereas the level of 13-kDa prolamin decreased. High temperature increased the accumulation of pyruvate/oxaloacetate-derived amino acids and decreased the levels of sugar phosphates and organic acids involved in glycolysis/gluconeogenesis and the TCA cycle. She et al. (2010) reported that ATP contents decreased during the grain-filling stage under high temperature.

Several QTLs for grain chalkiness have been detected in *japonica* rice cultivars and found to be stably expressed across multiple environments (Wan et al. 2004; Wan et al. 2005; Kobayashi et al. 2007; Tabata et al. 2007). Three

QTLs have been fine-mapped on chromosomes 1, 7, and 9 (Wan et al. 2004; Zhou et al. 2009), but no genes associated with chalkiness have been cloned yet. Some of the QTLs for grain chalkiness are located near known starch synthesis genes (Yamakawa et al. 2008), which may represent candidate genes for these QTLs. Additional studies will be required to fully uncover the genetic and biochemical processes underlying chalkiness.

Cooking Characteristics and Sensory Qualities

Cooking Properties of Milled Rice

Rice cultivars show various cooking, sensory, and processing qualities. Researchers have recognized differences in cooking qualities among cultivars since the early 20th century (Warth and Darabsett 1914). Research on cooking and eating qualities has been performed to discover the factors determining these qualities.

In most places, rice is cooked in boiling water. The hydration properties of the grain are evaluated by the absorption of water and the expansion of the grain after cooking (Bergman et al. 2004; Liu et al. 2008). During cooking, starch granules absorb water, swell, and melt at a critical temperature called the gelatinization temperature. A study including both *japonica* and *indica* cultivars showed a range of gelatinization temperatures from 55 to 79° C (Bhattacharya 1979). The gelatinization temperature influences the cooking time: rice starch with a low gelatinization temperature needs less cooking time, which represents a massive potential for both global energy and cost savings for rice consumers, especially the poor (Fitzgerald et al. 2009). The alkali spreading (digestion) assay is used to predict the gelatinization temperature from the degree of disintegration of polished rice soaked in potassium hydroxide. Gel consistency is measured as the length of the gel that can be made from a sample of rice flour treated with potassium hydroxide (Cagampang et al. 1973). In the gel consistency test, hard rice grain makes a short gel, but soft rice grain makes a long and viscous gel. These differences in gel formation are associated with rheological properties (Tran et al. 2011).

Cooked *japonica* rice grain is stickier and softer than cooked *indica* grain. The pasting characteristics and texture of cooked rice grain have been widely evaluated in *japonica* rice cultivars and breeding lines. To assess the pasting characteristics of cooked rice, the Rapid Visco Analyzer (RVA) is often used to measure the gelatinization and viscosity changes that occur during heating, holding, and cooling of a starch–water slurry. In several countries, including China, Japan, Australia, and the USA, the RVA has become the standard method for the evaluation of pasting properties by processing companies and in breeding programs (Bergman et al. 2004). Because of the small sample size required for RVA evaluation, this method is suitable for screening the pasting characteristics of materials in breeding programs. One of the important pasting properties evaluated by the RVA is the "breakdown score" (calculated as final viscosity minus trough viscosity), which is highly correlated with the eating quality of cooked rice as assessed by sensory testing. The breakdown score has also been reported to be a good predictor of the baking expansion of Japanese rice crackers (Yamada et al. 1993). Final viscosity is used to define the deterioration rate after cooking and cooling. RVA scores reveal a wide range of phenotypic variation in *japonica* rice cultivars (Kim 2009; Yan et al. 2011).

Texture is the primary sensory property of cooked rice that determines consumer acceptance (Okabe 1979). The major components of cooked rice texture relevant to *japonica* rice cultivars are hardness and stickiness (adhesiveness). Instrument-based methods for evaluating cooked rice texture have been developed by rice-processing industries and breeding programs (Mohandoss and Pillaiyar 1980; Lee and Peleg 1988; Ohtsubo et al. 1998; Okadome et al. 1999). Okabe (1979) evaluated the texture of cooked rice grains in *japonica* cultivars and reported

that strong stickiness and softness are preferred among Japanese consumers.

A wide range of variation in physicochemical properties related to cooking characteristics between *indica* and *japonica* cultivars and between glutinous and non-glutinous cultivars has been observed (Juliano et al. 1964; Champagne et al. 1999; Lisle et al. 2000). Several QTLs for physicochemical properties have been reported in populations derived from crosses between *indica* and *japonica* rice cultivars (Bao et al. 1999; Bao et al. 2000). Many QTLs were detected near genes related to starch synthesis, such as *Waxy* (*Wx*, which encodes granule-bound starch synthase) and *alkali disintegration* (*Alk*, which encodes starch synthase IIa), both on the short arm of chromosome 6. These genes are responsible for the main differences in the physicochemical properties of cooked rice between *indica* and *japonica* cultivars, and may conceal other genetic factors with smaller effects. There have been a few reports of phenotypic variation in these cooking traits among non-glutinous *japonica* rice cultivars (Lin et al. 2005; Kang et al. 2006), but the genetic basis of this variation could not be determined, possibly because of low levels of DNA polymorphism in the materials. In general, *japonica* cultivars show less phenotypic variation in grain physicochemical properties than do *indica* cultivars (Bao et al. 2004). Several instruments for testing the pasting properties of rice grain can clearly detect differences between *indica* and *japonica* cultivars, but not between *japonica* cultivars. This gap is an impediment to an in-depth genetic analysis of the traits in *japonica* rice cultivars.

Components of Rice Endosperm

The major components of polished rice grain are starch (up to 95% of dry weight), protein (5–7%), and lipids (0.5–1%). The chemical components of rice grain (endosperm) affect many cooking characteristics and sensory qualities (Han and Hamaker 2001; Martin and Fitzgerald 2002). Because starch is the main component of the grain, many studies of rice cooking characteristics and grain appearance have focused on the starch components, primarily amylose and amylopectin. Amylose is considered the most important factor affecting cooking and sensory quality in rice. The granule-bound starch synthase I enzyme (GBSSI) is required for amylose synthesis in rice, and the gene encoding GBSSI was named *Waxy* (*Wx*) when it was first discovered through classical genetic studies (Ikeno 1914). The amylose content of rice cultivars is classified as waxy (0–2%), very low (3–9%), low (10–19%), intermediate (20–25%), or high (>25%) (Fitzgerald et al. 2003). Glutinous rice cultivars and *waxy* mutant lines carry a deletion in the *Wx* gene that is fatal to the activity of GBSSI (Mikami et al. 1999), and consequently they contain no amylose. In non-glutinous rice cultivars, a single-nucleotide polymorphism (SNP) at the splice site of intron 1 defines two alleles, Wx^a and Wx^b, and differentiates intermediate- and high-amylose cultivars from low-amylose cultivars (Isshiki et al. 1998). Wx^a is distributed predominantly in *indica* cultivars and Wx^b in *japonica* cultivars. A recent association study identified an SNP in exon 6 that differentiates high- and intermediate-amylose cultivars (Chen et al. 2008; Mikami et al. 2008). By using near-isogenic lines, it was shown that this SNP decreases the level of GBSSI protein, lowering the amylose content from high to intermediate (Mikami et al. 2008). An SNP in exon 4 is associated with an opaque phenotype and defines the *Wx* allele of the very-low-amylose cultivars (Mikami et al. 2008). Cultivars with very low amylose show even less GBSSI activity than low-amylose cultivars (Mikami et al. 1999).

Other genes besides *Wx* also affect the amylose content. For example, *dull* (*du*) mutants have low amylose content (Satoh and Omura 1981; Okuno and Yano 1984), and several *du* mutant genes have been mapped and isolated (Yano et al. 1988; Zeng et al. 2007; Isshiki et al. 2008). The mutant gene *du1* is a member of a family of pre-mRNA processing genes, and *du3*

encodes the cap-binding protein 20-kD subunit (Zeng et al. 2007; Isshiki et al. 2008). Despite these advances, we do not yet fully understand the genetics of amylose content variation in rice cultivars. In particular, it is not yet possible to explain the genetic basis of variation among non-glutinous *japonica* rice cultivars that have the same Wx^b (*GBSSI*) allele. Recently, another genetic factor (QTL) for amylose content has been detected on chromosome 9 in *japonica* cultivars (Ando et al. 2010); a small reduction (2–3%) in amylose content occurred in a cultivar with a particular allele at this new QTL.

Amylose content is also affected by environmental factors such as temperature during seed ripening. The amylose content of the same cultivar grown in different environments may vary by up to 6%. For example, low-amylose cultivars, which typically have 12% to 15% amylose when grown at higher temperatures, have up to 18% amylose at lower temperatures (Larkin and Park 1999; Bao et al. 2000). Consequently, breeders need to evaluate amylose content across multiple years and locations to obtain more accurate estimates.

Amylopectin structure also affects cooking characteristics and sensory qualities. Abundant long-chain amylopectin creates strong and resilient starch granules, resulting in low stickiness and high hardness of cooked rice. Several studies have reported differences in the chain-length distribution of amylopectin between *indica* and *japonica* cultivars: *indica* cultivars have more long-chain amylopectin and less short-chain amylopectin, whereas *japonica* cultivars have the converse (Takeda et al. 1987; Hizukuri et al. 1989; Umemoto et al. 2002). However, variation in the amount of long-chain amylopectin among both *indica* and *japonica* cultivars has been reported (Kang et al. 1995; Yoshio et al. 1995).

Amylopectin is synthesized by enzymes with multiple subunits or isoforms, which are encoded by many genes (Comparot-Moss and Denyer 2009; Sun et al. 2011). Several mutant genes for starch synthesis have been isolated from *japonica* cultivars, and their roles in amylopectin formation have been characterized (Satoh and Omura 1981). These mutant genes include *waxy* (*wx*; mutation in the gene encoding GBSSI), *alkali disintegration* (*alk*; SSIIa), *shrunken 1* (*sh1*; SSI), *floury 5* (*flo5*; SSIII), *starch-branching enzyme mutant 1* (*sbe1*; BEI), *amylose extender* (*ae*; BEIIb), and *sugary 1* (*su1*; ISAI) (Yano et al. 1985; Yamanouchi and Nakamura 1992; Nishi et al. 2001; Nakamura 2002; Ball and Morell 2003; Dian et al. 2003; Satoh et al. 2003; Tanaka et al. 2004; Dian et al. 2005; Ryoo et al. 2007). Satoh and colleagues (2008) detected *phol* (plastidial alpha-glucan phosphorylase) mutants, Fujita and colleagues (2009) detected *pul* (pullulanase-type of debranching enzyme) mutants, and She and colleagues (2010) detected *flo2* (protein harboring a tetratricopeptide repeat motif) mutants. The proteins encoded by these genes also influence amylopectin structure. GBSSI is not only responsible for amylose synthesis, but is also involved in amylopectin synthesis, especially in forming the extra-long chain of amylopectin (Hanashiro et al. 2008). In *japonica* cultivars, mutations in the gene encoding SSI (*sh1*) had no effect on the size and shape of seeds and starch granules or on the crystallinity of endosperm (Fujita et al. 2006), but did show changes in amylopectin chain-length distribution. Mutants defective in SSIIIa (*flo5*) have seeds with a white core, small starch granules, and relatively loosely packed endosperm cells (Fujita et al. 2007); they show decreased amounts of long-chain amylopectin and increased amylose content.

Moreover, several mutants for starch precursor synthesis have been identified and the responsible genes have been isolated. These include *flo4* (mutant gene for PPDK), *bt1* (mutant gene for an adenylate translocator), and *sh2* and *brittle 2* (*bt2*) (mutant genes for the large and small subunits of AGP, respectively) (Kang et al. 2005; Ohdan et al. 2005; Comparot-Moss and Denyer 2009). The relationship between eating quality and mutations in starch synthesis genes is not yet clear. However, some studies indicate that it

might be possible to develop high-quality (i.e., high-palatability) cultivars through modification of starch synthesis genes in *japonica* cultivars (Isshiki et al. 2008; Sun et al. 2011).

There is evidence that grain components other than starch contribute to cooking characteristics and sensory properties in rice. For example, the amounts of both proteins and lipids affect sensory properties (Martin and Fitzgerald 2002; Philpot et al. 2006). Protein content plays an important role in determining pasting characteristics, including grain texture and the surface hardness of cooked rice (Yanase et al. 1984; Okadome et al. 1999). The protein content of *japonica* rice is negatively correlated with palatability (Kim et al. 1997). Yang and Chang (1999) and Fitzgerald and colleagues (2003) demonstrated that lipids affect the pasting properties of rice flour: high lipid content raises gelatinization and most processing and cooking temperatures.

Sensory Test

In practical breeding programs, the eating quality of cooked rice is usually measured via sensory analysis by well-trained panels of individuals (Kobayashi and Tomita 2008; Takeuchi et al. 2008; Wada et al. 2008; Kwon et al. 2011). Sensory tests evaluate the glossiness, stickiness, hardness, and taste of cooked rice and provide a score of overall eating quality (Yamamoto and Ogawa 1992). Sensory tests are difficult to perform with early-generation breeding populations because of the requirement for large amounts of grain (e.g., several hundred grams of polished rice), considerable labor to perform the polishing and cooking, and a well-trained panel of at least 20 people. Physicochemical properties such as amylose content, protein content, gelatinization temperature, and pasting characteristics have been used to predict the eating quality of cooked rice. However, these properties are still supplementary indices of eating quality, in that they do not completely agree with the results of eating quality evaluations by sensory tests (Kobayashi and Tomita 2008; Wada et al. 2008). Thus, the sensory test remains the most effective method to determine eating quality.

Little information is currently available on the genetic control of eating quality in *japonica* rice cultivars as evaluated by sensory tests (Yamamoto and Ogawa 1992; Kwon et al. 2011). In a QTL analysis using populations derived from crosses between *indica* and *japonica* cultivars, major-effect QTLs for eating quality were detected on the short arm of chromosome 6 (Wan et al. 2004; Ebitani et al. 2005; Takeuchi et al. 2007). This chromosome region includes the *Wx* and *Alk* loci; therefore, allelic differences in these two genes control some of the main differences in eating quality between *indica* and *japonica* cultivars. Recently, some researchers have identified eating-quality QTLs by using doubled-haploid lines, recombinant inbred lines, and backcross inbred lines derived from crosses between *japonica* cultivars (Suh et al. 2006; Kobayashi and Tomita 2008; Takeuchi et al. 2008; Wada et al. 2008; Kwon et al. 2011). In these studies, QTLs for eating quality were also found in genomic regions other than the short arm of chromosome 6.

In Japan, QTL studies of eating quality have used the Japanese *japonica* cultivar Koshihikari. Koshihikari has been a top cultivar in Japan since 1979 (Yokoo et al. 2005) and possesses superior eating quality among Japanese *japonica* cultivars. Several studies have detected eating-quality QTLs on the short arm of chromosome 3 (Figure 9.2) (Takeuchi et al. 2007, 2008; Kobayashi and Tomita 2008; Wada et al. 2008). At each of the QTLs, the Koshihikari allele was associated with superior eating quality scores in sensory tests. In the same region on the short arm of chromosome 3, QTLs for other eating-quality-related traits were also detected, namely for contents of glutamine and asparagine, hardness/stickiness ratio, and RVA profiles of hot-paste viscosity, cool-paste viscosity, and consistency viscosity (Kobayashi and Tomita 2008; Wada et al. 2008). No QTLs were detected for amylose or protein content in this region.

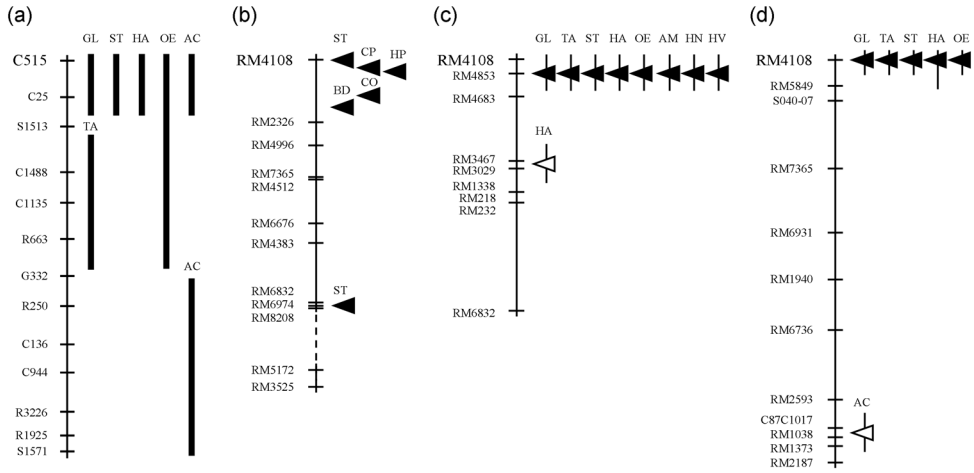


Fig. 9.2. Eating quality QTLs on rice chromosome 3 in genetic populations derived from Japanese *japonica* rice cultivar Koshihikari. *Boxes* and *triangles* indicate significant marker intervals and QTL peaks, respectively, in (a) chromosome substitution lines derived from a cross of Koshihikari × Kasalath (Takeuchi et al. 2007), (b) recombinant inbred lines of Sakihikari (a progeny cultivar of Koshihikari) × Nipponbare (Kobayashi and Tomita 2008), (c) recombinant inbred lines of Moritawase × Koshihikari (Wada et al. 2008), and (d) backcross inbred lines of Nipponbare × Koshihikari (Takeuchi et al. 2008). *Black bars* and *triangles* indicate that the Koshihikari allele of that region/QTL improves eating quality; while *white triangles* indicate that the Koshihikari allele decreases eating quality. Abbreviations are as follows: GL, glossiness; TA, taste; ST, stickiness; HA, hardness; OE, overall evaluation of cooked rice by sensory test; AC, amylose content; HP, hot-paste viscosity; CP, cool-paste viscosity; CO, consistency viscosity; BD, breakdown viscosity by Rapid Visco Analyzer; AM, (glutamine + asparagine)/total amino acid; HN, hardness/adhesion; HV, hardness/adhesiveness by texturometer.

Takeuchi et al. (2008) confirmed the genetic effect of these Koshihikari QTL alleles by analyzing a chromosome segment substitution line (CSSL) containing a Koshihikari segment of the short arm of chromosome 3 in the genetic background of Nipponbare, a *japonica* rice cultivar with inferior eating quality. The CSSL grain was glossier, tasted better, was stickier, was softer, and had a superior overall evaluation score compared to Nipponbare (Figure 9.3). It also had increased contents of the starch precursors glucose-6-phosphate and fructose-6-phosphate (Hori et al. unpublished data). Follow-up studies such as fine-mapping, gene cloning, and phenotypic characterizations could reveal genetic factors underlying aspects of eating quality other than amylose or protein content, and biochemical measures of good eating quality conferred by this chromosome region from Koshihikari. Thus, the QTL localized on the short arm of chromosome 3 is an important target region for the development of new *japonica* cultivars showing superior eating quality in practical breeding programs.

Other Traits Affecting Grain Quality

The aroma of fragrant rice is of particular importance as it has a clear local and national identity. Jasmine- and Basmati-style cultivars are widely known types of *indica* fragrant rice from Thailand and India, respectively. More than 100 compounds that contribute to rice aroma have been identified; of these, 2-acetyl-1-pyrroline (2-AP) is the primary component (Tsugita 1986; Bradbury et al. 2005). However, strong aroma traits are not demanded of most *japonica* cultivars, because aroma is not a preferred trait in *japonica* cultivation areas.

Opportunities for improving the nutritional value of rice grains have become increasingly evident. Nutritional components such as minerals and vitamins are either absent or present

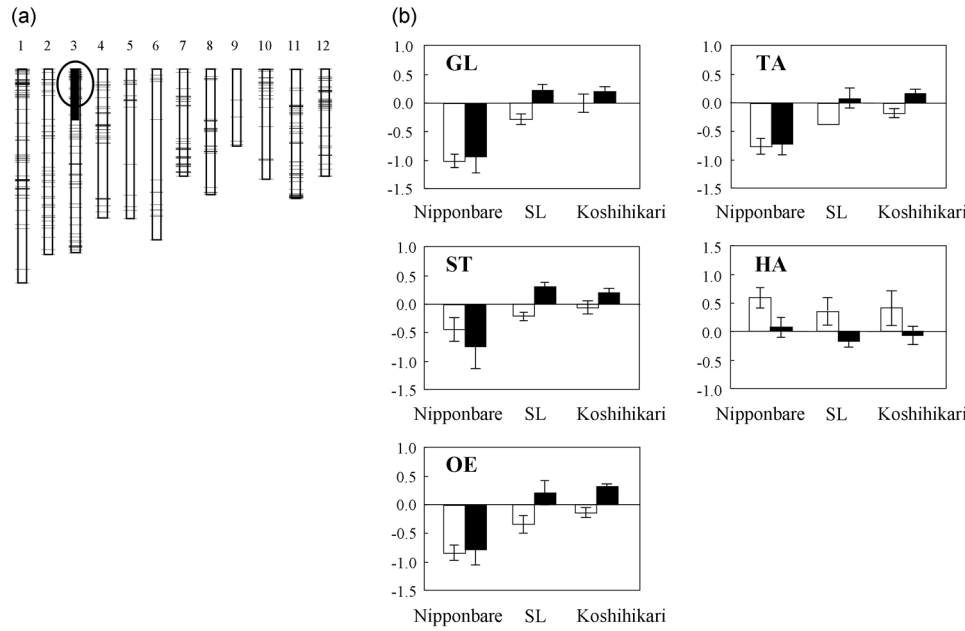


Fig. 9.3. (a) Graphical genotype of a chromosome segment substitution line (CSSL) containing a Koshihikari segment of the short arm of chromosome 3 in the genetic background of Nipponbare. The 12 *boxes* represent the 12 rice chromosomes, as numbered at the top. *Black* and *white boxes* denote regions derived from Koshihikari and Nipponbare, respectively. *Circle* indicates an eating quality QTL on the short arm of chromosome 3 detected in backcross inbred lines of Nipponbare/Koshihikari. *Horizontal bars* show the positions of 718 DNA markers used for genotyping. (b) Mean scores evaluated by sensory tests of cooked rice in CSSL, Nipponbare, and Koshihikari. *White* and *black boxes* denote scores in 2007 and 2008, respectively. *Error bars* indicate standard deviations. Abbreviations are as follows: GL, glossiness; TA, taste; ST, stickiness; HA, hardness; OE, overall evaluation of cooked rice by sensory test. (Revised version of figures from Takeuchi et al. 2008)

at low levels in polished grains (Lucca et al. 2006). Varietal differences in grain components have been investigated by metabolome analysis, mainly in *indica* rice cultivars (Kusano et al. 2007; Mochida et al. 2009). In recent years, transgenic techniques have been used successfully in several studies: elevating the Fe content of rice endosperm by introducing ferritin genes (Vasconcelos et al. 2003), increasing the amino acid content of rice endosperm by introducing tryptophan synthesis genes (Dubouzet et al. 2007; Saika et al. 2011), and producing nutritionally valuable amounts of β-carotene in the endosperm (Golden Rice) by introducing the genes encoding carotene desaturase and phytoene synthase (Hoa et al. 2003; Paine et al. 2005). These transgenic approaches can enhance breeding efforts leading to the development of locally adapted *japonica* cultivars. Although these types of nutritionally enhanced rice are still in the experimental stages, transgenic techniques for rice improvement are certain to have a major impact on human health in both developing countries (ingestion of insufficient nutrients) and advanced countries (enhancement of food functionality).

Perspectives

There will certainly be future demand for *japonica* rice cultivars with improved grain quality and yield, necessitating the use of the latest technologies for rice improvement. Advances in technology are currently being seen in two main areas: application of novel phenotyping tools and genetic dissection procedures.

The currently available instruments for measuring physicochemical properties in rice grain can detect differences between *indica* and *japonica* cultivars but do not differentiate well among *japonica* cultivars. On the other hand, rice breeders and consumers can recognize the slight differences in eating-quality traits among *japonica* cultivars. For example, two cultivars with the same amylose content can be differentiated by sensory test panels and consumers (Fitzgerald et al. 2003). Improvements in measuring instruments and techniques will be needed to provide wider measurement ranges and less measurement error. Metabolome analysis is a notable example: advances in analytical technologies, combined with dedicated data analysis tools, are already beginning to show advantages in the measurement of grain components in rice (Fitzgerald et al. 2009; Kusano et al. 2007; Oikawa et al. 2008; Mochida et al. 2009). Comprehensive detection of grain components by metabolome analysis may detect grain-quality differences not distinguishable by conventional measurement methods. These novel applications could enable QTL detection and gene cloning of novel eating-quality components beyond amylose content, amylopectin structure, and protein content.

Until recently, the extremely low frequency of DNA polymorphisms that could be detected among *japonica* rice cultivars prevented the molecular genetic analysis of many agronomic traits. Recently the genome sequences of several *japonica* cultivars, including Nipponbare, Koshihikari, Rikuu 132, Eiko, and Omachi, have made it possible to detect polymorphisms within this subspecies (IRGSP 2005; Yamamoto et al. 2010; Nagasaki et al. 2010; Arai-Kichise et al. 2011). SNPs have been detected among these cultivars. The large number of SNPs distributed throughout the *japonica* rice genome can overcome the limitations once caused by the extremely low frequency of DNA markers available. These SNPs have enhanced the genetic dissection of phenotypic differences among *japonica* cultivars by facilitating QTL analysis (Shibaya et al. 2011). High-density SNP genotyping will be a powerful tool for the detection and pyramiding of QTLs for grain-quality traits in *japonica* cultivars.

QTL analysis of naturally occurring phenotypic variation has contributed greatly to our understanding of the genetic control of grain quality (Kobayashi et al. 2007; Tabata et al. 2007; Kobayashi and Tomita 2008; Wada et al. 2008; Takeuchi et al. 2008; Kwon et al. 2011). Genetic populations used in these studies have played a crucial role in this progress. Recently a novel analytical method for QTL detection has been proposed: the genome-wide association study (GWAS) (Yu et al. 2006; Iwata et al. 2007) makes it possible to detect QTLs directly in germplasm accessions without the construction of genetic populations and has already been applied successfully in analyses of heading date and grain yield, two agronomically important traits (Zhao et al. 2011; Huang et al. 2012). Although the sensitivity of GWAS is sometimes insufficient to detect QTLs that are strongly correlated with population structure (Iwata et al. 2007; Zhao et al. 2011), its application may be suitable for *japonica* cultivars because of the relatively simple population structure within this subspecies. GWAS could reduce many of the labor-intensive and time-consuming aspects of QTL mapping and facilitate genetic analysis of naturally occurring variation in grain-quality traits.

Reverse-genetic approaches are facilitated by the extensive genomic sequence information available for rice, and such approaches are already providing useful results. For example, introduction of artificial microRNAs has been successfully used to induce post-transcriptional gene silencing in rice with unprecedented specificity, resulting in the modulation of agronomically important traits such as plant height and tillering (Warthmann et al. 2008). In particular, extensive knowledge of regulatory and biochemical pathways that are involved in trait expression has facilitated successful biofortification of rice grain (Bekaert et al. 2008; Storozhenko et al. 2007). TILLING (Targeting Induced Local Lesions in Genomes) is a non-transgenic reverse

genetics approach that can be used to decipher plant genome sequences (Till et al. 2007; Tsai et al. 2011). This approach can detect an allelic series of silent, missense, and splicing alterations in a target gene from mutant panels induced by exposure to radiation and chemical mutagens. Application of TILLING in *japonica* rice cultivars could prove the functional effects of grain-quality genes. Successes in reverse-genetic approaches will unlock real opportunities for targeting grain components through the design of rice plants that express novel genes and pathways in their grains.

Recent studies of grain quality in *japonica* rice cultivars have identified many genes and QTLs involved in grain quality. However, these results have been insufficient for the directed improvement of cultivars showing both high yield and high grain quality. Advances in research efforts are still needed to improve grain-quality traits in *japonica* rice cultivars.

Acknowledgements

Our own research described in this review has been supported by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan (Integrated Research Project for Plant, Insect and Animal using Genome Technology QT2011 and Genomics for Agricultural Innovation NVR0001 and QTL4010). The authors thank Dr. Yoshinobu Takeuchi and Dr. Keitaro Suzuki, National Agriculture and Food Research Organization (NARO) Institute of Crop Sciences, and Dr. Yozo Okazaki, RIKEN Plant Science Center, for continuous support of our genetic studies of eating quality in *japonica* rice cultivars.

References

Ando, I., H. Sato, N. Aoki, Y. Suzuki, H. Hirabayashi, M. Kuroki, H. Shimizu, T. Ando, and Y. Takeuchi. 2010. Genetic analysis of the low-amylose characteristics of rice cultivars Oborozuki and Hokkai-PL9. *Breed Sci* 60:187–194.

Arai-Kichise, Y., Y. Shiwa, H. Nagasaki, K. Ebana, H. Yoshikawa, M. Yano, and K. Wakasa. 2011. Discovery of genome-wide DNA polymorphisms in a landrace cultivar of *japonica* rice by whole-genome sequencing. *Plant Cell Physiol* 52:274–282.

Asaoka, M., K. Okuno, Y. Sugimoto, J. Kawakami, and H. Fuwa. 1984. Effect of environmental temperature during development of rice plants on some properties of endosperm starch. *Starch* 36:189–193.

Asatsuma, S., C. Sawada, A. Kitajima, T. Asakura, and T. Mitsui. 2006. α-amylase affects starch accumulation in rice grains. *J Appl Glycosci* 53:187–192.

Ball, S. G., and M. K. Morell. 2003. From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* 54:207–233.

Bao, J. S., and Y. W. Xia. 1999. Genetic control of paste viscosity characteristics in *indica* rice (*Oryza sativa* L.). *Theor Appl Genet* 98:1120–1124.

Bao, J. S., X. W. Zheng, Y. W. Xia, P. He, Q. Y. Shu, X. Lu, Y. Chen, and L. H. Zhu. 2000. QTL mapping for the paste viscosity characteristics in rice (*Oryza sativa* L.). *Theor Appl Genet* 100:280–284.

Bao, J., H. Corke, and M. Sun. 2004. Genetic diversity in the physicochemical properties of waxy rice (*Oryza sativa* L) starch. *J Sci Food Agric* 84:1299–1306.

Bekaert, S., S. Storozhenko, P. Mehrshahi, M. J. Bennett, W. Lambert, J. F. Gregory, 3rd, K. Schubert, J. Hugenholtz, D. Van Der Straeten, and A. D. Hanson. 2008. Folate biofortification in food plants. *Trends Plant Sci* 13:28–35.

Bergman, C. J., K. R. Bhattacharya, and K. Ohtsubo. 2004. Rice end-use quality analysis. In Champagne, E. T. (ed.), *Rice: Chemistry and technology*, 3rd ed. American Association of Cereal Chemists Inc, St. Paul, MN, pp. 415–472.

Bhattacharya, K. R. 1979. Gelatinization temperature of rice starch and its determination. In *Chemical aspects of rice grain quality*, International Rice Research Institute, Los Baños, Philippines, pp. 231–249.

Bhattacharya, K. R. 1980. Breakage of rice during milling: A review. *Trop Sci* 22:255–276.

Bradbury, L. M., T. L. Fitzgerald, R. J. Henry, Q. Jin, and D. L. Waters. 2005. The gene for fragrance in rice. *Plant Biotechnol J* 3:363–370.

Cagampang, G. B., C. M. Perez, and B. O. Juliano. 1973. A gel consistency test for eating quality of rice. *J Sci Food Agric* 24:1589–1594.

Champagne, E. T., K. L. Bett, B. T. Vinyard, A. M. McClung, F. E. Barton, K. Moldenhauer, S. Linscombe, and K. McKenzie. 1999. Correlation between cooked rice texture and Rapid Visco Analyser measurements. *Cereal Chem* 76:764–771.

Chen, M. H., C. J. Bergman, S. R. M. Pinson, and R. G. Fjellstrom. 2008. *Waxy* gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. *J Cereal Sci* 47:536–545.

Cheng, F. M., L. J. Zhong, F. Wang, and G. P. Zhang. 2005. Differences in cooking and eating properties between

chalky and translucent parts in rice grains. *Food Chemistry* 90:39–46.

Comparot-Moss, S., and K. Denyer 2009. The evolution of the starch biosynthetic pathway in cereals and other grasses. *J Exp Bot* 60:2481–2492.

Dian, W., H. Jiang, Q. Chen, F. Liu, and P. Wu. 2003. Cloning and characterization of the granule-bound starch synthase II gene in rice: Gene expression is regulated by the nitrogen level, sugar and circadian rhythm. *Planta* 218:261–268.

Dian, W., H. Jiang, and P. Wu. 2005. Evolution and expression analysis of starch synthase III and IV in rice. *J Exp Bot* 56:623–632.

Dubouzet, J. G., A. Ishihara, F. Matsuda, H. Miyagawa, H. Iwata, and K. Wakasa. 2007. Integrated metabolomic and transcriptomic analyses of high-tryptophan rice expressing a mutant anthranilate synthase alpha subunit. *J Exp Bot* 58:3309–3321.

Ebana, K., J. Yonemaru, S. Fukuoka, H. Iwata, H. Kanamori, N. Namiki, H. Nagasaki, and M. Yano. 2010. Genetic structure revealed by a whole-genome single-nucleotide polymorphism survey of diverse accessions of cultivated Asian rice (*Oryza sativa* L.). *Breed Sci* 60:390–397.

Ebitani, T., Y. Takeuchi, Y. Nonoue, T. Yamamoto, K. Takeuchi, and M. Yano. 2005. Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of *indica* rice cultivar Kasalath in a genetic background of *japonica* elite cultivar Koshihikari. *Breed Sci* 55:65–73.

Famoso, A. N., K. Zhao, R. T. Clark, C. W. Tung, M. H. Wright, C. Bustamante, L. V. Kochian, and S. R. McCouch. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet* 7:e1002221.

Fan, C., Y. Xing, H. Mao, T. Lu, B. Han, C. Xu, X. Li, and Q. Zhang. 2006. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet* 112:1164–1171.

Fitzgerald, M. A., M. Martin, R. M. Ward, W. D. Park, and H. J. Shead. 2003. Viscosity of rice flour: A rheological and biological study. *J Agr Food Chem* 51:2295-2299.

Fitzgerald, M. A., S. R. McCouch, and R. D. Hall. 2009. Not just a grain of rice: The quest for quality. *Trends Plant Sci* 14:133–139.

Fujita, N., Y. Toyosawa, Y. Utsumi, T. Higuchi, I. Hanashiro, A. Ikegami, S. Akuzawa, M. Yoshida, A. Mori, K. Inomata, et al. 2009. Characterization of pullulanase (*PUL*)-deficient mutants of rice (*Oryza sativa* L.) and the function of *PUL* on starch biosynthesis in the developing rice endosperm. *J Exp Bot* 60:1009–1023.

Fujita, N., M. Yoshida, N. Asakura, T. Ohdan, A. Miyao, H. Hirochika, and Y. Nakamura. 2006. Function and characterization of starch synthase I using mutants in rice. *Plant Physiol* 140:1070–1084.

Fujita, N., M. Yoshida, T. Kondo, K. Saito, Y. Utsumi, T. Tokunaga, A. Nishi, H. Satoh, J. H. Park, J. L. Jane, et al. 2007. Characterization of SSIIIa-deficient mutants of rice: The function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the rice endosperm. *Plant Physiol* 144:2009–2023.

Han, X. Z., and B. R. Hamaker. 2001. Amylopectin fine structure and rice starch paste breakdown. *Cereal Sci* 34:279–284.

Hanashiro, I., K. Itoh, Y. Kuratomi, M. Yamazaki, T. Igarashi, J. Matsugasako, and Y. Takeda. 2008. Granule-bound starch synthase I is responsible for biosynthesis of extra-long unit chains of amylopectin in rice. *Plant Cell Physiol* 49:925–933.

Hizukuri, S., Y. Takeda, N. Maruta, and B. O. Juliano. 1989. Molecular structure of rice starch. *Carbohydr Res* 189:227–235.

Hoa, T. T., S. Al-Babili, P. Schaub, I. Potrykus, and P. Beyer. 2003. Golden *Indica* and *Japonica* rice lines amenable to deregulation. *Plant Physiol* 133:161–169.

Huang, X., Y. Zhao, X. Wei, C. Li, A. Wang, Q. Zhao, W. Li, Y. Guo, L. Deng, C. Zhu, et al. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet* 44:32–39.

Ikeno, S. 1914. Über die Bestäubung und die Bastardierung von Ries. *Z Pflanzenzücht* 2:495–503.

IRGSP. 2005. The map-based sequence of the rice genome. *Nature* 436:793–800.

Ishimaru, Y., K. Bashir, M. Fujimoto, G. An, R. N. Itai, N. Tsutsumi, H. Nakanishi, and N. K. Nishizawa. 2009. Rice-specific mitochondrial iron-regulated gene (MIR) plays an important role in iron homeostasis. *Mol Plant* 2:1059–1066.

Isshiki, M., Y. Matsuda, A. Takasaki, H. L. Wong, H. Satoh, and K. Shimamoto. 2008. *Du3*, a mRNA cap-binding protein gene, regulates amylose content in *Japonica* rice seeds. *Plant Biotechnol J* 25:483–487.

Isshiki, M., K. Morino, M. Nakajima, R. J. Okagaki, S. R. Wessler, T. Izawa, and K. Shimamoto. 1998. A naturally occurring functional allele of the rice *waxy* locus has a GT to TT mutation at the 5' splice site of the first intron. *Plant J* 15:133–138.

Iwata, H., Y. Uga, Y. Yoshioka, K. Ebana, and T. Hayashi. 2007. Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms. *Theor Appl Genet* 114:1437–1449.

Jiang, H., W. Dian, and P. Wu. 2003. Effect of high temperature on fine structure of amylopectin in rice endosperm by reducing the activity of the starch branching enzyme. *Phytochemistry* 63:53–59.

Jodari, F., and S. D. Linscombe. 1996. Grain fissuring and milling yields of rice cultivars as influenced by environmental conditions. *Crop Sci* 36:1496–1502.

Juliano, B. O., G. M. Bautista, J. C. Lugay, and A. C. Reyes. 1964. Rice quality, studies on physicochemical properties of rice. *J Agr Food Chem* 12:131–138.

Juliano, B. O., C. M. Perez, and F. Cuevas-Perez. 1993. Screening for stable high head rice yields in rough rice. *Cereal Chem* 70:650–655.

Kang, H. G., S. Park, M. Matsuoka, and G. An. 2005. White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C-type pyruvate orthophosphate dikinase gene (*OsPPDKB*). *Plant J* 42:901–911.

Kang, H. J., I. K. Hwang, K. S. Kim, and H. C. Choi. 2006. Comparison of the physicochemical properties and ultrastructure of *japonica* and *indica* rice grains. *J Agr Food Chem* 54:4833–4838.

Kang, K., K. Kim, and S. K. Kim. 1995. Relationship between molecular structure of rice amylopectin and texture of cooked rice. *Korean J Food Sci Technol* 27:105–111.

Kato, S., H. Hosaka, and S. Hara. 1928. On the affinity of rice varieties as shown by fertility of hybrid plants. *Bull Sci Fac Agric Kyushu Univ* 3:132–147.

Kim, H. J., S. P. Hong, K. H. Park, I. H. Kim, and S. H. Lee. 1997. Comparative study of *japonica* rice protein with different eating quality. *Foods Biotechnol* 6:149–153.

Kim, K. H. 2009. Grain quality improvement in *japonica* rice: Achievements and prospects. *J Agro Indonesia* 37:183–187.

Kobayashi, A., Y. Koga, H. Uchiyamada, S. Samoto, H. Horiuchi, K. Miura, K. Okuno, Y. Fujita, Y. Uehara, S. Ishizaka, M. Nakagahara, T. Yamada, and K. Maruyama. 1990. Breeding a new rice variety, Oochikara. *Bull Hokuriku Natl Agric Exp Stn* 32:85–104.

Kobayashi, A., B. Genliang, Y. Shenghai, and K. Tomita. 2007. Detection of quantitative trait loci for white-back and basal-white kernels under high temperature stress in *japonica* rice varieties. *Breed Sci* 57:107–116.

Kobayashi, A., and K. Tomita. 2008. QTL detection for stickiness of cooked rice using recombinant inbred lines derived from crosses between *japonica* rice cultivars. *Breed Sci* 58:419–426.

Kovach, M. J., M. N. Calingacion, M. A. Fitzgerald, and S. R. McCouch. 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc Natl Acad Sci USA* 106:14444–14449.

Kusano, M., A. Fukushima, M. Kobayashi, N. Hayashi, P. Jonsson, T. Moritz, K. Ebana, and K. Saito. 2007. Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. *J Chromatogr A* 855: 71–79.

Kwon, S. W., Y. C. Cho, J. H. Lee, J. P. Suh, J. J. Kim, M. K. Kim, I. S. Choi, H. G. Hwang, H. J. Koh, and Y. G. Kim. 2011. Identification of quantitative trait loci associated with rice eating quality traits using a population of recombinant inbred lines derived from a cross between two temperate *japonica* cultivars. *Mol Cells* 31:437–445.

Lan, Y., and O. R. Kunze. 1996. Fissure characteristics related to moisture absorption stresses in rice. *Trans ASAE* 39:2169–2174.

Larkin, P. D., and W. D. Park. 1999. Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. *Plant Mol Biol* 40:719–727.

Lee, S. J., and M. Peleg. 1988. Direct measurement of the attractive force between individual cooked rice grains of sticky and flaky cultivars. *J Food Sci* 53:1113–1115.

Li, J., J. Xiao, S. Grandillo, L. Jiang, Y. Wan, Q. Deng, L. Yuan, and S. R. McCouch. 2004. QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. *Genome* 47:697–704.

Li, Y., C. Fan, Y. Xing, Y. Jiang, L. Luo, L. Sun, D. Shao, C. Xu, X. Li, J. Xiao, et al. 2011. Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat Genet* 43:1266–1269.

Lin, C. J., C. Y. Li, S. K. Lin, F. H. Yang, J. J. Huang, Y. H. Liu, and H. S. Lu. 2010. Influence of high temperature during grain filling on the accumulation of storage proteins and grain quality in rice (*Oryza sativa* L.). *J Agr Food Chem* 58:10545–10552.

Lin, J., C. Shi, M. Wu, and J. Wu. 2005. Analysis of genetic effects for cooking quality traits of *japonica* rice across environments. *Plant Sci* 168:1501–1506.

Lisle, A. J., M. Martin, and M. A. Fitzgerald. 2000. Chalky and translucent rice grains differ in starch composition and structure and cooking properties. *Cereal Chem* 77:627–632.

Liu, L. L., X. Y. Yan, L. Jiang, W. W. Zhang, M. Q. Wang, S. R. Zhou, Y. Shen, Y. Y. Shen, S. J. Liu, L. M. Chen, et al. 2008. Identification of stably expressed quantitative trait loci for cooked rice elongation in non-Basmati varieties. *Genome* 51:104–112.

Lucca, P., S. Poletti, and C. Sautter. 2006. Genetic engineering approaches to enrich rice with iron and vitamin A. *Physiol Plantarum* 126:291–303.

McCouch, S. R., L. Teytelman, Y. Xu, K. B. Lobos, K. Clare, M. Walton, B. Fu, R. Maghirang, Z. Li, Y. Xing, et al. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res* 9:199–207.

Mackill, D. J. 1995. Classifying *japonica* rice cultivars with RAPD markers. *Crop Sci* 35:889–894.

Martin, M., and M. A. Fitzgerald. 2002. Proteins in rice grains influence cooking properties. *Cereal Sci* 36:285–294.

Matsuo, T. 1952. Genecological studies on cultivated rice. *Bull Natl Inst Agric Sci* 3:1–111.

Mikami, I., M. Aikawa, H.-Y. Hirano, and Y. Sano. 1999. Altered tissue-specific expression at the *Wx* gene of the opaque mutants in rice. *Euphytica* 105:91–97.

Mikami, I., N. Uwatoko, Y. Ikeda, J. Yamaguchi, H. Y. Hirano, Y. Suzuki, and Y. Sano. 2008. Allelic diversification at the *wx* locus in landraces of Asian rice. *Theor Appl Genet* 116:979–989.

Mochida, K., T. Furuta, K. Ebana, K. Shinozaki, and J. Kikuchi. 2009. Correlation exploration of metabolic and genomic diversity in rice. *BMC Genomics* 10:568.

Mohandoss, R., and P. Pillaiyar. 1980. An extrusion test for determining the palatability of parboiled rice. *J Food Sci Technol* 17:244–246.

Morita, S., and H. Nakano. 2011. Nonstructural carbohydrate content in the stem at full heading contributes to high performance of ripening in heat-tolerant rice cultivar Nikomaru. *Crop Sci* 51:818–828.

Nagasaki, H., K. Ebana, T. Shibaya, J. I. Yonemaru, and M. Yano. 2010. Core single-nucleotide polymorphisms-a tool for genetic analysis of the Japanese rice population. *Breed Sci* 60:648–655.

Nakagahra, M., K. Okuno, and D. Vaughan. 1997. Rice genetic resources: history, conservation, investigative characterization and use in Japan. *Plant Mol Biol* 35:69–77.

Nakamura, Y. 2002. Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: Rice endosperm as a model tissue. *Plant Cell Physiol* 43:718–725.

Nishi, A., Y. Nakamura, N. Tanaka, and H. Satoh. 2001. Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. *Plant Physiol* 127:459–472.

Ohdan, T., P. B. Francisco, Jr., T. Sawada, T. Hirose, T. Terao, H. Satoh, and Y. Nakamura. 2005. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *J Exp Bot* 56:3229–3244.

Ohtsubo, K., H. Toyoshima, and H. Okadome. 1998. Quality assay of rice using traditional and novel tools. *Cereal Food World* 43:203–206.

Oikawa, A., F. Matsuda, M. Kusano, Y. Okazaki, and K. Saito. 2008. Rice Metabolomics. *Rice* 1:63–71.

Okabe, M. 1979. Texture measurement of cooked rice and its relationship to the eating quality. *J Texture Stud* 10:131–152.

Okadome, H., H. Toyoshima, and K. Ohtsubo. 1999. Multiple measurements of physical properties of individual cooked rice grains with a single apparatus. *Cereal Chem* 76:855–860.

Okuno, K., and M. Yano. 1984. New endosperm mutants modifying starch characteristics of rice, *Oryza sativa L. JARQ* 18:73–78.

Paine, J. A., C. A. Shipton, S. Chaggar, R. M. Howells, M. J. Kennedy, G. Vernon, S. Y. Wright, E. Hinchliffe, J. L. Adams, A. L. Silverstone, et al. 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotechnol* 23:482–487.

Patindol, J., and Y. J. Wang. 2003. Fine structures and physicochemical properties of starches from chalky and translucent rice kernels. *J Agr Food Chem* 51:2777–2784.

Peng, S., and G. S. Khush. 2003. Four decades of breeding for varietal improvement of irrigated lowland rice in the International Rice Research Institute. *Plant Prod Sci* 6:157–164.

Philpot, K., M. Martin, V. Butardo, Jr., D. Willoughby, and M. Fitzgerald. 2006. Environmental factors that affect the ability of amylose to contribute to retrogradation in gels made from rice flour. *J Agr Food Chem* 54:5182–5190.

Ryoo, N., C. Yu, C. S. Park, M. Y. Baik, I. M. Park, M. H. Cho, S. H. Bhoo, G. An, T. R. Hahn, and J. S. Jeon. 2007. Knockout of a starch synthase gene *OsSSIIIa*/*Flo5* causes white-core floury endosperm in rice (*Oryza sativa* L.). *Plant Cell Rep* 26:1083–1095.

Saika, H., A. Oikawa, F. Matsuda, H. Onodera, K. Saito, and S. Toki. 2011. Application of gene targeting to designed mutation breeding of high-tryptophan rice. *Plant Physiol* 156:1269–1277.

Sarker, N. N., O. R. Kunze, and T. Strouboulis. 1996. Transient moisture gradients in rough rice mapped with finite element model and related to fissures after heated air drying. *Trans ASAE* 39:625–631.

Satoh, H., A. Nishi, K. Yamashita, Y. Takemoto, Y. Tanaka, Y. Hosaka, A. Sakurai, N. Fujita, and Y. Nakamura. 2003. Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. *Plant Physiol* 133:1111–1121.

Satoh, H., and T. Omura. 1981. New endosperm mutations induced by chemical mutagens in rice (*Oryza sativa* L.). *Jpn J Breed* 31:316–326.

Satoh, H., K. Shibahara, T. Tokunaga, A. Nishi, M. Tasaki, S. K. Hwang, T. W. Okita, N. Kaneko, N. Fujita, M. Yoshida, et al. 2008. Mutation of the plastidial alpha-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. *Plant Cell* 20:1833–1849.

She, K. C., H. Kusano, K. Koizumi, H. Yamakawa, M. Hakata, T. Imamura, M. Fukuda, N. Naito, Y. Tsurumaki, M. Yaeshima, et al. 2010. A novel factor *FLOURY ENDOSPERM2* is involved in regulation of rice grain size and starch quality. *Plant Cell* 22:3280–3294.

Shibaya, T., Y. Nonoue, N. Ono, U. Yamanouchi, K. Hori, and M. Yano. 2011. Genetic interactions involved in the inhibition of heading by heading date QTL, *Hd2* in rice under long-day conditions. *Theor Appl Genet* 123:1133–1143.

Shomura, A., T. Izawa, K. Ebana, T. Ebitani, H. Kanegae, S. Konishi, and M. Yano. 2008. Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet* 40:1023–1028.

Song, X. J., W. Huang, M. Shi, M. Z. Zhu, and H. X. Lin. 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630.

Storozhenko, S., V. De Brouwer, M. Volckaert, O. Navarrete, D. Blancquaert, G. F. Zhang, W. Lambert, and D. Van Der Straeten. 2007. Folate fortification of rice by metabolic engineering. *Nat Biotechnol* 25:1277–1279.

Suh, J. P., Y. C. Cho, S. J. Kwon, I. S. Choi, H. C. Hong, Y. G. Kim, S. N. Ahn, and H. G. Hwang. 2006. Identification of QTLs for grain quality on a RIL population derived from a cross between *japonica* rices. *Korean J Breed* 38:105–112.

Sun, M. M., S. E. Abdula, H. J. Lee, Y. C. Cho, L. Z. Han, H. J. Koh, and Y. G. Cho. 2011. Molecular aspect of good eating quality formation in *Japonica* rice. *PLoS One* 6:e18385.

Sundaram, R., M. Vishnupriya, S. Biradar, G. Laha, G. Reddy, N. Rani, N. Sarma, and R. Sonti. 2008. Marker assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite *indica* rice variety. *Euphytica* 160:411–422.

Tabata, M., H. Hirabayashi, Y. Takeuchi, I. Ando, Y. Iida, and R. Ohsawa. 2007. Mapping of quantitative trait loci for the occurrence of white-back kernels associated with high temperatures during the ripening period of rice (*Oryza sativa* L.). *Breed Sci* 57:47–52.

Takano-Kai, N., H. Jiang, T. Kubo, M. Sweeney, T. Matsumoto, H. Kanamori, B. Padhukasahasram, C. Bustamante, A. Yoshimura, K. Doi, et al. 2009. Evolutionary history of *GS3*, a gene conferring grain length in rice. *Genetics* 182:1323–1334.

Takeda, K., and K. Saito. 1980. Major genes controlling grain size of rice. *Jpn J Breed* 30:280–282.

Takeda, Y., S. Hizukuri, and B. O. Juliano. 1987. Structure of rice amylopectins with low and high affinities for indine. *Carbohydr Res* 187:287–294.

Takeuchi, Y., K. Hori, K. Suzuki, Y. Nonoue, Y. Takemoto-Kuno, H. Maeda, H. Sato, H. Hirabayashi, H. Ohta, T. Ishii, et al. 2008. Major QTLs for eating quality of an elite Japanese rice cultivar, Koshihikari, on the short arm of chromosome 3. *Breed Sci* 58:437–445.

Takeuchi, Y., Y. Nonoue, T. Ebitani, K. Suzuki, N. Aoki, H. Sato, O. Ideta, H. Hirabayashi, M. Hirayama, H. Ohta, et al. 2007. QTL detection for eating quality including glossiness, stickiness, taste and hardness of cooked rice. *Breed Sci* 57:231–242.

Takita, T. 1992. Varietal differences in crack formation in *japonica* and *indica* rice. *Jpn J Breed* 42:397–402.

Tanaka, N., N. Fujita, A. Nishi, H. Satoh, Y. Hosaka, M. Ugaki, S. Kawasaki, and Y. Nakamura. 2004. The structure of starch can be manipulated by changing the expression levels of starch branching enzyme IIb in rice endosperm. *Plant Biotechnol J* 2:507–516.

Tashiro, T., and I. F. Wardlaw. 1991. The effect of high temperature on kernel dimensions and the type and occurrence of kernel damage in rice. *Aust J Agric Res* 42:485–496.

Till, B. J., J. Cooper, T. H. Tai, P. Colowit, E. A. Greene, S. Henikoff, and L. Comai. 2007. Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol* 7:19.

Tran, N. A., V. D. Daygon, A. P. Resurreccion, R. P. Cuevas, H. M. Corpuz, and M. A. Fitzgerald. 2011. A single nucleotide polymorphism in the *Waxy* gene explains a significant component of gel consistency. *Theor Appl Genet* 123:519–525.

Tsai, H., T. Howell, R. Nitcher, V. Missirian, B. Watson, K. J. Ngo, M. Lieberman, J. Fass, C. Uauy, R. K. Tran, A. A. Khan, V. Filkov, T. H. Tai, J. Dubcovsky, and L. Comai. 2011. Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiol* 156:1257–1268.

Tsugita, T. 1986. Aroma of cooked rice. *Food Rev Int* 1:497–520.

Umemoto, T., Y. Nakamura, H. Satoh, and K. Terashima. 1999. Differences in amylopectin structure between two rice varieties in relation to the effects of temperature during grain-filling. *Starch* 51:58–62.

Umemoto, T., and K. Terashima. 2002. Activity of granule-bound starch synthase is an important determinant of amylose content in rice endosperm. *Funct Plant Biol* 29:1121–1124.

Umemoto, T., M. Yano, H. Satoh, A. Shomura, and Y. Nakamura. 2002. Mapping of a gene responsible for the difference in amylopectin structure between *japonica*-type and *indica*-type rice varieties. *Theor Appl Genet* 104:1–8.

Vasconcelos, M., K. Datta, N. Oliva, M. Khalekuzzaman, L. Torrizo, S. Krishnan, M. Oliveira, F. Goto, and S. K. Datta. 2003. Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. *Plant Sci* 164:371–378.

Wada, T., T. Ogata, M. Tsubone, Y. Uchimura, and Y. Matsue. 2008. Mapping of QTLs for eating quality and physicochemical properties of the *japonica* rice Koshihikari. *Breed Sci* 58:427–435.

Wan, X. Y., J. M. Wan, C. C. Su, C. M. Wang, W. B. Shen, J. M. Li, H. L. Wang, L. Jiang, S. J. Liu, L. M. Chen, et al. 2004. QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. *Theor Appl Genet* 110:71–79.

Wan, X. Y., J. M. Wan, J. F. Weng, L. Jiang, J. C. Bi, C. M. Wang, and H. Q. Zhai. 2005. Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. *Theor Appl Genet* 110:1334–1346.

Wang, E., J. Wang, X. Zhu, W. Hao, L. Wang, Q. Li, L. Zhang, W. He, B. Lu, H. Lin, et al. 2008. Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat Genet* 40:1370–1374.

Warth, F. J., and D. B. Darabsett. 1914. *Disintegration of rice grains by means of alkali*. Bull Agric Res Inst no. 38.

Warthmann, N., H. Chen, S. Ossowski, D. Weigel, and P. Herve. 2008. Highly specific gene silencing by artificial miRNAs in rice. *PLoS One* 3:e1829.

Wei, X. J., L. Jiang, J. F. Xu, X. Liu, S. J. Liu, H. Q. Zhai, and J. M. Wan. 2009. The distribution of *japonica* rice cultivars in the lower region of the Yangtze River valley is determined by its photoperiod-sensitivity and heading date genotypes. *J Integr Plant Biol* 51:922–932.

Woo, M. O., T. H. Ham, H. S. Ji, M. S. Choi, W. Jiang, S. H. Chu, R. Piao, J. H. Chin, J. A. Kim, B. S. Park, et al. 2008. Inactivation of the *UGPase1* gene causes genic male sterility and endosperm chalkiness in rice (*Oryza sativa* L.). *Plant J* 54:190–204.

Yamada, H., N. Sasagawa, and K. Kitazawa. 1993. Suitabilities of improved rice varieties and lines for rice crackers. *J Jpn Soc Food Sci Technol* 40:278–286.

Yamakawa, H., T. Ebitani, and T. Terao. 2008. Comparison between locations of QTLs for grain chalkiness and genes responsive to high temperature during grain filling on the rice chromosome map. *Breed Sci* 58:337–343.

Yamakawa, H., and M. Hakata. 2010. Atlas of rice grain filling-related metabolism under high temperature: Joint analysis of metabolome and transcriptome demonstrated inhibition of starch accumulation and induction of amino acid accumulation. *Plant Cell Physiol* 51:795–809.

Yamakawa, H., T. Hirose, M. Kuroda, and T. Yamaguchi. 2007. Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. *Plant Physiol* 144:258–277.

Yamamoto, T., H. Nagasaki, J. Yonemaru, K. Ebana, M. Nakajima, T. Shibaya, and M. Yano. 2010. Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics* 11: 267.

Yamamoto, Y., and T. Ogawa. 1992. Eating quality in Japanese rice cultivars. *Jpn J Breed* 42:177–183.

Yamanouchi, H., and Y. Nakamura. 1992. Organ specificity of isoforms of starch branching enzyme (Q-enzyme) in rice. *Plant Cell Physiol* 33:985–991.

Yan, C. J., Z. X. Tian, Y. W. Fang, Y. C. Yang, J. Li, S. Y. Zeng, S. L. Gu, C. W. Xu, S. Z. Tang, and M. H. Gu. 2011. Genetic analysis of starch paste viscosity parameters in glutinous rice (*Oryza sativa* L.). *Theor Appl Genet* 122:63–76.

Yanase, H., K. Ohtsubo, K. Hashimoto, H. Sato, and T. Teranishi. 1984. Correlation between protein contents of brown rice and textural parameters of cooked rice and cooking quality of rice. *Rep Natl Food Res Inst* 45:118–122.

Yang, C. H., and W. H. Chang. 1999. Effects of protein and lipid binding to starch on the physicochemical and pasting properties of rice flour. *Food Sci Agric Chem* 1:277–285.

Yano, M., K. Okuno, J. Kawakami, H. Satoh, and T. Omura. 1985. High amylose mutants of rice, *Oryza sativa* L. *Theor Appl Genet* 69:253–257.

Yano, M., K. Okuno, H. Satoh, and T. Omura. 1988. Chromosomal location of genes conditioning low amylose content of endosperm starches in rice, *Oryza sativa* L. *Theor Appl Genet* 76:183–189.

Yano, M. 2001. Genetic and molecular dissection of naturally occurring variation. *Curr Opin Plant Biol* 4:130–135.

Yokoo, M., M. Hirao, and T. Imai. 2005. Annual change in leading rice varieties between 1956 and 2000 in Japan. *Bull Natl Inst Crop Sci* 7:19–125.

Yonemaru, J., T. Yamamoto, S. Fukuoka, Y. Uga, K. Hori, and M. Yano. 2010. Q-TARO: QTL annotation rice online database. *Rice* 3:194–203.

Yoshio, N., I. Maeda, K. Teranishi, M. Hisamatsu, and T. Yamada. 1995. Molecular structures and properties of starches from different cultivars of rice (*Oryza sativa* L. *japonica*) produced in Japan. *J Appl Glycosci* 42:365–374.

Youens-Clark, K., E. Buckler, T. Casstevens, C. Chen, G. Declerck, P. Derwent, P. Dharmawardhana, P. Jaiswal, P. Kersey, A. S. Karthikeyan, et al. 2011. Gramene database in 2010: Updates and extensions. *Nucleic Acids Res* 39:1085–1094.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208.

Zakaria, S., T. Matsuda, S. Tajima, and Y. Nitta. 2002. Effect of high temperature at ripening stage on the reserve accumulation in seed in some rice cultivars. *Plant Prod Sci* 5:160–168.

Zhang, H., S. Zhang, J. Zhang, J. Yang, and Z. Wang 2008. Postanthesis moderate wetting drying improves both quality and quantity of rice yield. *Agron J* 100:726–734.

Zeng, D., M. Yan, Y. Wang, X. Liu, Q. Qian, and J. Li. 2007. *Du1*, encoding a novel Prp1 protein, regulates starch biosynthesis through affecting the splicing of *Wxb* pre-mRNAs in rice (*Oryza sativa* L.). *Plant Mol Biol* 65:501–509.

Zhao, K., C. W. Tung, G. C. Eizenga, M. H. Wright, M. L. Ali, A. H. Price, G. J. Norton, M. R. Islam, A. Reynolds, J. Mezey, et al. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2:467.

Zhou, L., L. Chen, L. Jiang, W. Zhang, L. Liu, X. Liu, Z. Zhao, S. Liu, L. Zhang, J. Wang, et al. 2009. Fine mapping of the grain chalkiness QTL *qPGWC-7* in rice (*Oryza sativa* L.). *Theor Appl Genet* 118:581–590.

Chapter 10

Biofortified Maize – A Genetic Avenue for Nutritional Security

Raman Babu, Natalia Palacios and BM Prasanna

Abstract

Over the past four decades significant progress has been made in advancing maize production as well as productivity, resulting in increased per capita availability in many developing nations. However, overdependence on cereal-based diets coupled with inadequate grain legume production and availability will likely lead to serious nutritional deficiencies, especially in developing nations, where animal products are either expensive or insufficiently produced. It is therefore imperative to pay attention to improving the nutritional qualities of staple cereals such as maize, which is extensively consumed as human food in many parts of sub-Saharan Africa, Latin America, and South Asia. Here we briefly review the biochemical characteristics of normal grain maize, its nutritional deficiencies, with special reference to imbalanced amino acid composition in the endosperm, low carotenoid and high phytic acid content, and various scientific approaches to "biofortify" maize grain. Quality protein maize (QPM) represents an excellent technological intervention, which contains twice the amount of lysine and tryptophan as normal maize. The recent discovery of two key genes influencing grain carotenoid composition has enabled increasing the beta carotene content of yellow maize two- to tenfold in the tropical germplasm. A number of scientific advances have been made in understanding the phytate metabolic pathway in maize, which appears promising for product development in the future. The biofortification objectives in maize are likely to evolve further as new targets such as antioxidant and prebiotic content and bioavailability emerge relevant to the target sub-populations.

Introduction

Maize (*Zea mays* L.), which literally means "that which sustains life," is an important cereal grain providing nutrients for humans and animals worldwide. In addition, it is a basic raw material for the production of starch, oil, protein, alcoholic beverages, food sweeteners, and, more recently, fuel. Maize is a major source of calories in the diets of 230 million inhabitants of developing countries – 81 million in sub-Saharan Africa, 141 million in South Asia, and 8 million in Latin America. Annual per capita maize consumption averages 36, 10, and 23 kilograms,

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

respectively, in these regions, but this masks significant variation and per capita food consumption of maize. In Mesoamerica, annual maize consumption exceeds 80 kg per capita in Guatemala, Honduras, and El Salvador, rising to 125 kg in Mexico. Maize is also the most important cereal food crop in sub-Saharan Africa, where consumption levels exceed 130 kg per capita per year in Lesotho, Malawi, and Zambia. The highest amounts of maize consumed are found in southern Africa, at 85 kg/capita/year as compared to 27 in East Africa and 25 in West and Central Africa (Shiferaw et al. 2011). In South and Southeast Asia, where direct maize consumption on an annual average is estimated to be only 6 and 16 kg per capita, respectively, there are several areas (especially in the highlands and tribal regions) where maize is consumed directly at much higher rates. Maize is also used as animal feed and raw material for industrial use. In industrialized countries, a larger proportion of the grain is used as livestock feed and as industrial raw material for food and non-food uses. On the other hand, the bulk of maize produced in developing countries is used as human food, although its use as animal feed is increasing.

Maize kernels provide many macronutrients and can also provide small amounts of micronutrients such as vitamins (provitamin A, vitamin E) and minerals (iron, zinc). The typical composition of the maize kernel is 8–10% protein, 3.5–4.5% oil, 1.5–2.0% ash, 1.5–2.1% crude fiber, 1.4 –2% soluble sugars, 10–15% water, and 65–70% starch. The endosperm contains all of the starch and about 70% of the protein. The remaining protein and high levels of oils are found in the germ. Fiber and ash are predominantly in the pericarp (Nuss and Tanumihardjo 2010). Maize can also be considered one of the cereals with a large and diverse set of nutraceutical compounds. Nutraceutical food, or functional food, is defined as containing chemical compounds that exert a positive effect on human health (Serna-Saldivar 2011). These chemicals are not considered nutrients that are normally associated with deficiencies, but they do play an essential role combating oxidative stress, chronic diseases (obesity, diabetes), and cancer (Wildman 2000). The main nutraceuticals found in maize are phenolics, such as anthocyanins and flavonoids (blue maize), carotenoids (yellow maize), along with phytosterols, fiber, unsaturated lipids, and folic acid. Even so, overdependence on maize-based diets without other complementary food sources may lead to nutritional deficiency-related diseases, such as kwashiorkor, anemia, and corneal blindness.

Micronutrient malnutrition alone affects more than 2 billion people, mostly in resource-poor families in developing countries. For example, more than 300 million people in India suffer micronutrient deficiencies, and 35% of the world's malnourished children live in that country. Protein-energy malnutrition (PEM) is a potentially fatal body-depletion disorder, which is the leading cause of death in children in developing countries. Maize cultivars that combine high grain yield with balanced amino acid composition, enhanced levels of provitamin A, and kernel zinc concentration will have a positive impact on nutrition, health, and the quality of life, especially in areas where poverty and low incomes limit access to diversified diets, dietary supplements, or fortified foods (Ortiz-Monasterio et al. 2007; Pfeiffer and McClafferty 2007).

"Biofortification" is the breeding of staple food crops to increase micronutrient density. Graham and colleagues (2001) suggested that because of the widespread consumption of staple crops, biofortification may be an effective and sustainable way of addressing micronutrient malnutrition. An important advantage with respect to biofortified crops is that the recurrent costs are low and the benefits can be made available to all developing countries around the world. The biofortification approach generally involves a set of one-time fixed costs in developing breeding methodologies, introgressing nutritional quality traits into elite germplasm, and adapting these varieties to diverse environments. Some researchers (e.g., Bouis et al.

2011) have highlighted the complex challenge of demonstrating the potential of biofortified crops to improve consumer health. Nevertheless, several initiatives to take advantage of the natural diversity of nutrients and micronutrients found in staple food crops are under way (www.harvestplus.org). Some of the important biofortification targets in maize are increased amino acid content, enhanced mineral and vitamin content, and/or bioavailability, and reduced antinutrient content such as low phytic acid.

Enhanced and Balanced Amino Acid Content

Human beings as well as a number of farm animals are incapable of synthesizing certain amino acids; this fact has stimulated research on improving the levels of some "essential amino acids" in staple food crops. While cereals are primarily deficient in lysine (Lys) and tryptophan (Trp), legumes are found to be significantly short of methionine (Met). Consequently, these three essential amino acids have frequently been the targets of manipulation in maize as well as other food crops. In developed countries, the focus is generally on feed quality, as meat consumption provides a sufficient supply of essential amino acids for humans. In contrast, in developing countries where maize is directly consumed as food, both humanitarian and economic interests prevail (Ufaz and Galili 2008; Atlin et al. 2011). Here, we highlight two specific cases of genetic improvement in maize that resulted in a high nutritional value addition, particularly with respect to essential amino acid content in the endosperm.

Quality Protein Maize (QPM)

In normal maize endosperm, the average proportions of various fractions of protein are albumins 3%, globulin 3%, zein (prolamine) 60%, and glutelin 34%. While the embryo protein is dominated by albumins (+60%), which is superior in terms of nutritional quality, the zein in maize endosperm is low in lysine content (0.1 g/100 g of protein), which negatively affects the growth of animals (Osborne and Mendel 1914). In the maize kernel, the endosperm and the germ (embryo) constitute approximately 80% and 10% of the mature kernel dry weight, respectively. In the 1920s, in a Connecticut (USA) maize field, a natural spontaneous mutation of maize with soft, opaque grains was discovered, which was eventually named as *opaque2 (o2)* (Singleton 1939). In 1964, Dr. Oliver Nelson's team at Purdue University, also in the USA, discovered that the homozygous recessive *o2* allele had substantially higher lysine (+69%) in grain endosperm compared to normal maize (Mertz et al. 1964). In *o2* maize, the zein fraction is markedly reduced by roughly 50% with a concomitant increase in the relative amounts of nutritionally superior fractions such as albumins, globulins, and glutelins. The lysine value of *o2* maize is 3.3 to 4.0 g/100 g of protein, which is more than twice that of endosperm from the normal maize (1.3 g lysine/100 g protein). The protein quality of *o2* maize is 43% higher than that of common maize and 95% of the value of casein. The decreased level of zein (5–27%) in *o2* maize along with reduced leucine, leads to more tryptophan for niacin synthesis and thus helps to combat pellagra and significantly improves its nutritional quality.

Genes and gene combinations that bring about drastic alterations in either plant or kernel characteristics also produce several secondary or undesirable effects. The low prolamine or high lysine mutants are no exception. In addition to influencing several biochemical traits, they adversely affected a whole array of agronomic and kernel characteristics. The *o2* and other mutants adversely affect dry matter accumulation, resulting in lower grain yield due to increased endosperm size. The kernels dry slowly following physiological maturity of the grain and have a higher incidence of ear rot. Other changes generally associated with high lysine mutants include thicker pericarp, larger germ size, reduced cob weight, increased color

intensity in yellow maize grains, and reduction in kernel weight and density. Thus, despite the nutritional superiority of *o2* maize, it did not become popular with farmers or consumers mainly because of reduced grain yield, chalky and dull kernel appearance, and susceptibility to ear rot and stored grain pests. Accordingly, CIMMYT, the International Maize and Wheat Improvement Center, undertook to improve the phenotype of *o2* kernels in order to facilitate greater acceptability, by developing hard endosperm grain types with the protein quality of chalky *o2* strains. CIMMYT received funding support beginning in 1965 from the United Nations Development Program and introduced gene modifiers that changed the soft, starchy endosperm to a vitreous type preferred by farmers and consumers whilst retaining the elevated levels of lysine and tryptophan. CIMMYT has subsequently developed a range of hard endosperm *o2* genotypes with better protein quality through genetic selection, and these are popularly known as quality protein maize (QPM). Today's QPM is essentially interchangeable with normal maize in both cultivation and agronomic characteristics and is competitive in terms of yield, lodging, disease and pest resistance, and moisture level, while retaining the superior lysine and tryptophan content (Vasal 2001). In 2005, QPM was planted on 695,200 hectares across 24 developing countries.

There are various breeding options for developing QPM that is competitive in agronomic performance and market acceptance. Among the several approaches tested in CIMMYT, the most successful and rewarding option exploited combined use of *o2* with the associated endosperm and amino acid modifier genes. Intrapopulation selection for genetic modifiers in *o2* backgrounds exhibiting a higher frequency of modified *o2* kernels and recombination of superior hard endosperm *o2* families resulted in the development of good quality QPM donor stocks with a high degree of endosperm modification. This was followed by the large-scale development of QPM germplasm with a wide range of genetic backgrounds, representing tropical, subtropical, and highland maize germplasm and involving different maturities, grain color, and texture. An innovative breeding procedure designated as "modified backcross cum recurrent selection" was designed to enable rapid and efficient conversion programs (Vasal 1980; Prasanna et al. 2001). More recently, pedigree backcrossing schemes have been used to convert elite QPM lines to maize streak virus (MSV) resistant versions for deployment in Africa, as well as for conversion of elite African lines to QPM versions (Krivanek et al. 2007).

QPM hybrid breeding efforts were initiated at CIMMYT in 1985, as the QPM hybrid product has several advantages over open-pollinated QPM cultivars: (1) higher yield potential comparable to the best normal hybrids, (2) assured seed purity, (3) more uniform and stable endosperm modification, and (4) less monitoring of protein quality required during seed production. Several QPM hybrid combinations were derived and tested through international trial series at multiple CIMMYT and NARS locations in Asia, Africa, and Latin America. Current QPM breeding strategies at CIMMYT focus on pedigree breeding wherein the best performing inbred lines with complementary traits are crossed to establish new segregating families. Both QPM × QPM and QPM × non-QPM crosses are made, depending on the specific requirements of the breeding project. In addition, backcross conversion is also followed to develop QPM versions of parental lines of popular hybrid cultivars that are widely grown in CIMMYT's target regions. Inbred lines developed through this process are then used in the formation of QPM hybrids and QPM synthetics (Krivanek et al. 2007; Atlin et al. 2011).

The breeding of QPM involves manipulation of three distinct genetic systems (Bjarnason and Vasal 1992; Krivanek et al. 2007): (1) the recessive mutant allele of the *O2* gene, (2) the endosperm hardness modifier genes; and (3) the amino acid modifiers/genes influencing free amino acid content in the endosperm. The

O2 gene was cloned using a transposon tagging strategy with the maize mobile genetic elements, *Spm* (Schmidt et al. 1987) and *Ac* (Motto et al. 1988). The *O2* gene encodes a leucine-zipper class transcription factor required mainly for the expression of 22 kDa α-zein-coding genes and a gene encoding a ribosomal inactivating protein (Lohmer et al. 1991; Bass et al. 1992). Genotypes with the homozygous recessive allele (*o2/o2*) have a significant decrease in production of α-zeins and a corresponding increase in non-zein proteins that are rich in lysine and tryptophan (Gibbon and Larkins 2005). Additionally, the recessive allele of the *O2* transcription factor also reduces the production of the enzyme lysine keto-glutarate reductase, involved in free lysine degradation, resulting in enhanced free lysine in the endosperm of *o2* maize. In the segregating generations, this recessive allele is selected either visually (identifying mosaic ears on F_2 harvests) or using molecular markers. The endosperm hardness modifier genes, which convert the soft/opaque endosperm to a hard/vitreous endosperm without much loss of protein quality, are selected through a low cost but effective method of light-box screening, where light is projected through the vitreous grains or blocked by the opaque grains. Endosperm modification is polygenically controlled. However, genetic and molecular analyses revealed some major loci involved in *o2* modification; for example, one locus maps near the centromere of chromosome 7 and the second maps near the telomere on the long arm of chromosome 7 (Lopes et al. 1995). Despite the presence of *o2* and associated endosperm hardness modifier genes, the lysine and tryptophan levels in segregating families vary widely, indicating the existence of third set of genes that modify the amino acid content, which necessitates systematic biochemical evaluation of lysine and/or tryptophan levels in each breeding generation (Nurit et al. 2007). The lysine content of normal maize is around 2%, whereas it is approximately 4% (of the total protein) in QPM, with a range 1.6–2.6% in normal maize and 2.7–4.5% in QPM. Three genes associated with lysine level have been mapped to locations on chromosomes 2, 4, and 7, besides several major *o2* modifier-QTLs on chromosomes 1, 7, and 9 (Gibbon and Larkins 2005). Therefore, it is possible to get favorable responses to selection for endosperm texture modification as well as relative content of the essential amino acids, if they are monitored efficiently, during the QPM breeding programs.

Molecular Marker-Assisted QPM Breeding

The transfer of the QPM trait into elite maize lines is not straightforward, in that the *o2* allele has to be in homozygous recessive state along with the polygenic endosperm modifiers. Therefore, this process is influenced by three major factors: (1) each conventional backcross generation needs to be selfed to identify the *o2* recessive gene and a minimum of six backcross generations are required to recover satisfactory levels of the recurrent parent genome, (2) in addition to maintaining the homozygous *o2* gene, multiple endosperm modifiers must also be selected, and (3) rigorous biochemical tests to ensure enhanced lysine and tryptophan levels in the selected materials in each breeding generation require enormous labor, time, and financial resources. Although conventional breeding procedures have been used to convert commercial lines to QPM forms, these procedures are tedious and time-consuming. Rapid advances in genomics research and technologies have led to the use of marker-assisted selection (MAS), which holds promise in enhancing selection efficiency and expediting the development of new cultivars with higher yield potential (Ribaut and Hoisington 1998; Xu and Crouch 2008). While marker-assisted foreground selection (Tanksley 1983; Melchinger 1990) helps in identifying the gene of interest without extensive phenotypic assays, marker-assisted background selection (Young and Tanksley 1989; Frisch et al. 1999a, Frisch et al. 1999b) significantly expedites the rate of genetic gain/recovery

of recurrent parent genome in a backcross breeding program. With the development and access to reliable PCR-based allele-specific markers such as simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), MAS is becoming an attractive option, particularly for oligogenic traits such as QPM (Babu et al. 2004).

A rapid line conversion strategy for QPM has been developed (Babu et al. 2005) consisting of a two-generation backcross program that employs foreground selection for the *o2* gene in both backcross (BC) generations, background selection at non-target loci in the BC_2 generation, and phenotypic selection for kernel modification and other desirable agronomic traits in two subsequent selfed generations. This brings together the salient features of both marker-aided and phenotypic-based selection approaches, such as fixing the large segregating generation for the target allele (*o2*), reducing the linkage drag by selection of flanking markers for recipient allele type, recovering the maximum amount of the recurrent parent genome within two BC generations, and providing scope for precise phenotypic selection for desirable agronomic and biochemical traits on a reduced number of progeny.

There are a few successful examples of MAS for maize improvement using *o2*-specific molecular markers (Babu et al. 2005; Gupta et al. 2009; Prasanna et al. 2010). The parental lines of 'Vivek Hybrid 9' (CM145 and CM212) developed at Vivekananda Institute of Hill Agriculture (VPKAS), Almora, India, were converted into QPM versions through transfer of the *o2* gene using MAS and phenotypic screening for endosperm modifiers. The MAS-derived QPM hybrid 'Vivek QPM 9' was released in the year 2008 for commercial cultivation in zones I and IV in India. Vivek QPM 9 shows a 41% increase in tryptophan, 30% in lysine, 23% in histidine, and 3.4% in methionine, coupled with a 12% reduction in leucine, as compared to Vivek Hybrid 9. Domestic consumption of such QPM grains will help in reducing protein malnutrition in the hills and mountains. In view of this, F_1 hybrid seeds of Vivek QPM 9 are being produced on a large scale for distribution in Uttarakhand and other parts of the country. A few villages in Uttarakhand have also been identified for conversion to QPM villages. Vivek QPM 9 can potentially replace Vivek Hybrid 9 as well as composites without any yield penalty in these areas. The approach outlined above was also used to develop QPM versions of several elite, early maturing inbred lines adapted to the hill regions of India (Gupta et al. 2009). QPM versions of six elite inbred lines, which are the parents of three single-cross hybrids – PEHM2, Parkash, and PEEHM5 – have been developed recently by the Maize Genetics Unit, Indian Agricultural Research Institute (IARI), New Delhi, through the ICAR Network Project on Molecular Breeding (Prasanna et al. 2010). A Network Project on molecular breeding for QPM is also being implemented in India, funded by the Department of Biotechnology (DBT), Govt. of India, for conversion of several important Indian maize inbred lines into QPM versions.

QPM as an Animal Feed

Another important use of QPM is as a less expensive source of high protein animal feed for monogastric animals, such as pigs and poultry. Unlike ruminant animals (i.e., cattle, sheep, goats), monogastric animals require a more complete protein than cereals alone can generally provide. Conventional maize, insofar as it lacks sufficient lysine and tryptophan, presents nutritional protein limitations for monogastric animals. These amino acids are usually supplemented in animal feeds with soybeans, pulses, or commercially produced synthetic amino acids. QPM presents another option. Studies have documented improved growth in pigs and poultry when QPM is substituted for conventional maize, thereby increasing the bioavailable protein (Sullivan et al. 1989).

Although using QPM in commercial feed could possibly have a significant impact (Lauderdale 2000), more frequently, the poultry industry supplements maize-based feed with

inexpensive synthetic Lys and Trp additives. In contrast, methionine (Met) is the first limiting amino acid in maize-based poultry feed that is available only in synthetic form and at expensive rates. The poultry industry, especially in Asia, where the demand is significantly rising, is keenly looking for new maize types – maize for poultry feed – with an improved level of methionine and high oil content (Hellin and Erenstein 2009). Efforts are being made by several institutions, including CIMMYT, to characterize maize germplasm for Met content (e.g., Scott and Blanco 2009) as well as for enhancing Met content in elite maize germplasm through recurrent selection (Scott et al. 2008).

Impact and Reach of QPM in the Developing World

In Ethiopia, Tanzania, and Uganda, randomized trials have shown significantly improved height and weight of children consuming such varieties, particularly in Southern Ethiopia, where the population relies heavily on maize (Gunaratna et al. 2008). Despite demonstrated significant food and feed benefits of QPM, large-scale cultivation in farmers' fields is yet to be realized in many developing countries that have released QPM cultivars. There is an array of reasons for this. As with most nutritional traits, enhanced amino acid content is a "hidden trait," which generally does not command a premium price in the market; hence, any QPM cultivar recommended for commercial cultivation should be as high yielding or better than the normal cultivars already present in the market.

An additional challenge with respect to widespread adoption of QPM is the recessive nature of the *o2* gene. If the QPM cultivar is pollinated by normal maize (NM) pollen, there may be a loss of high protein quality resulting in erosion of the trait in farmer-saved seed systems. However, one of the factors contributing to low levels of outcrossing is the copious amounts of desirable pollen from the QPM that compete for the silks with the foreign pollen, thereby reducing the chances of foreign pollen landing on QPM silks (Burris 2001). Using geostatistical models ("kriging") under different environmental conditions, Machida and colleagues (2012) recently demonstrated that the weighted average levels of outcrossing (less than 20%) is far less than previously thought; hence, a greater part of the QPM grain will not be outcrossed when grown near a NM crop. Also this study suggested that the nutritional value of the QPM trait can be optimally sustained if farmers plant relatively large areas in a square rather than a rectangular field. Where absolute or near-absolute isolation from NM is difficult to achieve, the proponents of QPM varieties can confidently advocate for the coexistence of QPM and NM crops.

Possibilities for Genetically Engineered High-Lysine Maize

Genetic engineering efforts targeted towards enhancement of lysine content in maize kernels can be broadly classified into two categories: (1) anti-sense suppression (RNAi) of alpha-zein production in transgenic maize, leading to at least twice the amount of lysine; and (2) deregulation of the aspartate metabolic pathway through either feedback-insensitive mutant enzymes or suppressing the catabolic reactions downstream to lysine synthesis.

RNA interference (RNAi) technology is particularly aimed at developing a dominant *o2* trait in maize and has been used specifically to reduce 22-kDa (Segal et al. 2003) and 19-kDa alpha zeins (Huang et al. 2004 and Huang et al. 2005), which resulted in moderate increases (15–20%) in lysine content. In a subsequent study, using an improved double strand RNA (dsRNA) suppression construct, Huang and colleagues (2006) reported lysine and tryptophan levels similar to conventionally bred QPM genotypes. While the dominant nature of the anti-sense transgene is a definite advantage as compared to the recessive allele of *o2*, the opaque endosperm still needs to be modified by endosperm modifier genes whose epistasis with the transgene has not

yet been tested. Very recently, Wu and Messing (2011) reported a potential accelerated QPM selection scheme, which is based on an RNAi construct that is directed against 22 and 19 kda zeins, fused with visible green fluorescent protein (GFP) marker gene. When such RNAi lines were crossed with QPM lines, carrying *o2* kernel hardness modifier genes, green and vitreous progenies could be selected in the segregating generations, thereby demonstrating that high lysine content and hard endosperm traits could be selected in dominant fashion.

An alternative option for enhancing the Lys content is to manipulate the critical rate limiting steps in the aspartate pathway, which plays an essential role in the regulation of the biosynthesis of several essential amino acids, including lysine in higher plants. Increases in seed lysine content have been achieved through engineering of both lysine anabolism and lysine catabolism. The first key step of this pathway is regulated by aspartate kinase (AK), which catalyses the conversion of aspartate to b-aspartyl phosphate and is feedback-regulated by several end products, including lysine and threonine (Galili 1995; Wang et al. 2001). At least two isoforms of AK exist in plants – lysine-sensitive and threonine-sensitive enzymes (Dotson et al. 1989; Muehlbauer et al. 1994a, b; Azevedo et al. 1997). In maize, two mutant loci encoding monofunctional AK, *ask1* and *ask2*, were identified by genetic screening (Diedrick et al. 1990; Dotson et al. 1990b; Muehlbauer et al. 1994a), and the enzymes they encode are feedback-inhibited by lysine (Dotson et al. 1989). Mutations at these loci result in AKs that are less sensitive to lysine feedback inhibition and result in the overproduction of free lysine, threonine, methionine, and isoleucine (Dotson et al. 1990b; Muehlbauer et al. 1994a). While *ask1* is located on short arm of chr. 7, *ask2* is mapped and cloned on long arm of chr. 2 in maize (Muehlbauer et al. 1994a; Wang et al. 2007). DHDPS (dihydrodipicolinate synthase), a key regulatory enzyme in Lys biosynthesis, catalyzes the formation of dihydrodipicolinic acid by condensing pyruvate and ASA (aspartate beta semialdehide). DHDPS is highly sensitive to Lys feedback regulation. Plants with a mutant DHDPS are less sensitive to Lys feedback inhibition and overproduce the amino acid (Ghislain et al. 1995). Because bacterial DHDPS is less sensitive than plant DHDPS to Lys, genes encoding bacterial DHDPS have been used to genetically engineer plants that overproduce Lys (Falco et al. 1995). Another possible intervention point in the Asp metabolic pathway is the Lys degradation reaction, where LKR (Lysine ketoglutaric acid reductase) catalyses the first step of Lys degradation. LKR activity is dramatically reduced in *o2* maize and hence it is believed that depressed LKR activity is one of the significant factors resulting in enhanced Lys in *o2* maize endosperm. A two-pronged approach (Frizzi et al. 2008) of introducing a feedback insensitive DHDPS (CordapA) and suppressing the LKR by an intron-embedded suppression construct through a single transgene has yielded impressive increase in seed lysine content (4000 ppm of Lys), representing ~100-fold enhancement, the highest ever reported in maize kernels.

Although RNAi technology has emerged as a powerful tool for overcoming the pleiotropic and secondary effects of the desirable mutant genes, social acceptance and biosafety concerns regarding GM (genetically modified) food crops for large scale adoption still exist in some countries, and high-lysine GM maize may be no exception.

High Provitamin A Maize

Vitamin A deficiency in a majority of the chronically hungry nations is responsible for a number of disorders, such as impaired iron mobilization, severe growth retardation, blindness, depressed immunological responses, increased susceptibility to infections, and increased childhood mortality (FAO 2000; Sommer and Davidson 2002; WHO 2009; Wurtzel et al. 2012). Interestingly, maize happens to be the predominant staple food in those areas such as sub-Saharan Africa and Latin America, where vitamin A-deficiency symptoms abound. While carotenoids are absent

in white maize kernels, yellow maize is known to accumulate carotenoids in the endosperm, and is thus a good target for biofortification efforts. Maize exhibits considerable natural variation for ratios and concentrations of kernel carotenoids, with some genotypes accumulating as much as 80 μg/g of total carotenoids in grain on a dry-weight basis. Several of the carotenoids present in maize have important roles in human health. Provitamin A carotenoids (β-cryptoxanthin, and α- and β-carotene) are the precursors of vitamin A, which is essential in different systems in the human body and for the prevention of diet-related chronic diseases. Lutein and zeaxanthin, on the other hand, have been associated with lowering the risk of cataracts, age-related macular degeneration, and other degenerative diseases. The fraction of provitamin A carotenoids is typically only 10–20%, whereas zeaxanthin and lutein each commonly represent 30–50% of total carotenoids in maize (Ortiz-Monasterio et al. 2007). Most yellow maize grown and consumed throughout the world, however, has only 2 μg/g or less of provitamin A carotenoids. Based on the analysis of a wide range of temperate and limited tropical germplam, it appears that tropical maize contains more β-cryptoxanthin and less β-carotene than temperate maize, and because the emphasis was on enhancing β-carotene concentration, most of the initial breeding sources of high provitamin A germplasm were selected from temperate regions in the CIMMYT provitamin A breeding program. However, evaluation of a wider range of tropical germplasm in the future may identify potential donors for enhanced β-carotene content.

The carotenoid metabolic pathway has been well researched in model species, and key genes governing critical steps have been identified. In maize, of the many genes implicated in carotenoid metabolism, roles of the following three in the final accumulation of useful carotenoids in the grain are worth mentioning. Phytoene synthase1 (*Y1/Psy1*) catalyses the first committed step in the pathway leading to formation of phytoene from geranylgeranyl diphosphate and is primarily responsible for the shift from white to yellow maize. Once the carotenoid pathway is activated, two other genes, *LcyE* and *CrtRB1,* have been shown to regulate the accumulation of provitamin A-related compounds. Lycopene epsilon cyclase converts lycopene into zeta-carotene and eventually to alpha-carotene through the action of other associated genes. Naturally existing mutant alleles of *LcyE* with reduced functionality have been identified that proportions more lycopene into the beta-carotene branch of the pathway, thereby enhancing the flux towards provitamin A-related compounds (Harjes et al. 2008). *CrtRB1* is a hydroxylase gene that converts beta-carotene into beta-cryptoxanthin, whose provitamin A activity is only half that of beta-carotene. Natural genetic variation for *CrtRB1* has recently been discovered that results in the retention of more beta-carotene in the maize endosperm (Yan et al. 2010). Molecular markers have been developed based on the functional polymorphisms within the abovementioned three genes, which hold great potential for accelerated development of high carotenoid lines in a time- and resource-efficient manner. As these molecular markers are located within the target genes and are truly diagnostic of the allelic constitution, they offer an efficient means of tracking the favorable alleles in backcross or pedigree breeding programs.

Carotenoid degradation also plays an equally important role in determining the total carotenoid accumulation as well as its composition. A number of maize carotenoid cleavage genes have now been identified (Vallabhaneni et al. 2010). *ZmCCD1* has been found on chr. 9 (bin 9.07), which effectively cleaves carotenoids, thereby depleting the pool. *ZmCCD1* is linked to dominant *white cap1* (*wc1*) locus. Dominant *wc1* alleles and a higher copy number of ZmCCD1 result in low endosperm carotenoid content (Vallabhaneni et al. 2010). Identification of favorable alleles of CCD genes will likely add significantly to the enhanced retention of endosperm carotenoids.

As outlined by Wurtzel and colleagues (2012), efforts to increase the provitamin A concentration in maize need to target three potential intervention points in the carotenoid metabolic pathway: (1) maximizing the pathway flux through manipulation of flux determinants such as PSY, CrtISO, DXS3, DXR, HDR, and GGPPS1 (Vallbhaneni and Wurtzel 2009); (2) Optimising the provitamin A composition through manipulation of *LcyE* and/or *CrtRB1* (hydroxylase); and (3) minimizing the carotenoid degradation/sequestration by manipulating CCD, CrtISO, ZEP1, and ZEP2.

HarvestPlus, a multi-institutional Program on Agriculture for Improved Nutrition and Health, leads a global effort to develop and deliver biofortified staple food crops with one or more of the three most limiting nutrients in the diets of the poor: vitamin A, zinc, and iron (Brown 1991; Bouis 2010). CIMMYT leads the HarvestPlus-Maize Program, where the primary target is improving provitamin A concentration in the endosperm beyond 15 μg/g. Provitamin A biofortification of maize started seven years ago in CIMMYT and considerable progress has been achieved to date at CIMMYT and IITA in active collaboration with several institutions/universities worldwide. One of the major objectives of HarvestPlus-Mazie is to tropicalize the high provitamin A temperate maize sources so as to convert the popular African OPVs (open pollinated varieties) and inbred lines into high Provitamin A versions.

As one of the key target countries for deployment of high provitamin A maize, the HarvestPlus-Maize program is focusing on Zambia, where the average per capita consumption of maize is more than 130 kg per year (356 g per day), and the vitamin A deficiency is as high as 54% in children under the age of five, and 13% in women aged 15–49 (WHO 2009). The first-generation experimental hybrids developed at CIMMYT have about 6 to 9 μg/g of ProA, and five of these hybrids were submitted to the Zambian National Performance Trials (NPT) during 2010–2011. In addition to provitamin A-enriched elite germplasm development, several ongoing activities assess and validate farmer and consumer acceptance of these promising hybrids, and have begun creating interest, demand, and supply for seed of the best hybrids. It is expected that the best one or two hybrids will be formally released for commercialization and planting in 2012/2013. With the discovery of useful allelic diversity for *LcyE* and *CrtRB1* and development of molecular markers, source lines with >15 μg/g of provitamin A carotenoids have been identified and are now routinely used as parents for new crosses at CIMMYT. This has led to the selection of lines with 40–250% higher provitamin A carotenoid concentrations than lines without the favorable allele.

Another improvement in the breeding program was achieved by using UPLC (Ultra-Performance Liquid Chromatography) in place of HPLC (High-Performance Liquid Chromotography) for carotenoid screening, which greatly increased the number of samples of provitamin A that could be evaluated and allowed selection for this trait earlier in the breeding pipeline. The big challenge is to ensure that new biofortified varieties are competitive for all the crucial traits, including seed production, food processing quality, taste, and other characteristics that determine acceptability to farmers and consumers.

Kernel Fe and Zn-rich Maize for Alleviating "Hidden Hunger"

Micronutrient malnutrition, popularly called "hidden hunger," is one of the alarming problems in the developing world, afflicting an estimated 3 billion people (UNSCN 2004). Fe-related deficiencies affect cognitive development, growth, reproduction, and productivity (Bouis 2002). Zn deficiency leads to anorexia, depression, impaired growth and development, altered reproductive biology, gastrointestinal problems, and impaired immunity, and affects 49% of world population (Solomons 2003). Accordingly, understanding the genetic

variability for kernel micronutrients (like Fe and Zn) in maize germplasm and breeding for kernel Fe- and/or Zn-enriched elite maize germplasm assume considerable significance.

Studies undertaken by CIMMYT and IITA in collaboration with national partners have led to identification of some elite maize lines with high levels of kernel Fe and Zn in both normal maize (e.g., Banziger and Long 2000; Menkir 2008; Prasanna et al. 2011; Chakraborti et al. 2011a) as well as QPM genetic backgrounds (Chakraborti et al. 2011b). A study undertaken in India (Chakraborti et al. 2011b) demonstrated that despite dilution effect, QPM genotypes have considerable potential, particularly in Zn biofortification programs, as compared to the non-QPM. The apparent homeostasis for kernel Zn concentration of the QPM inbreds and hybrids may be possibly attributed to the pleiotropic effect of *opaque-2* allele or its close linkage with genes responsible for accumulation of higher Zn (Arnold et al. 1977; Gupta et al. 1980). Currently, the CIMMYT-Harvest Plus-maize program is pursuing a large-scale GWAS (genome-wide association study) for identifying genomic regions influencing kernel Zn and Fe in the tropical germplasm background.

Low Phytate Maize

The nutritional quality of cereals and legumes depends on the major seed phosphorus storage compound, phytic acid (*myo*-inositol-hexa*kis*phosphate). Phytic acid forms one to several percent of seed dry weight and typically is deposited in seeds as mixed phytate or phytin salts of potassium and magnesium. Phytic acid P represents from 65 to 85% of seed total P. Phytic acid is an effective chelator of positively charged cations. When consumed in feeds and foods, phytic acid will bind to nutritionally important mineral cations in the intestinal tract, such as calcium, iron, and zinc, and to proteins as well, making them unavailable biologically. Humans and non-ruminants, such as poultry, swine, and fish, excrete a large fraction of these salts. This phenomenon can contribute to human mineral deficiency, particularly with respect to iron and zinc. It has been demonstrated that substantial reductions in feed phytic acid result in improvements in calcium and zinc utilization. Efforts to increase Fe and its bioavailability in maize kernels have been limited due to the low Fe genetic diversity found in maize. Bioavailabity screening was also researched as an option for developing biofortified Fe maize, but the screening procedures and low diversity are not highly encouraging (Pixley et al. 2011). Bioavailable Fe has been achieved through transgenic approaches reducing phytate (Drakakai et al. 2001).

The first use of normal-phytate and low-phytate isolines in a human nutrition study evaluated fractional absorption of iron from tortillas (Raboy 2002), demonstrating an improvement of 49% for low phytate over normal lines. A similar study found that fractional absorption of zinc from an *lpa1-1* maize food was 30%, whereas it was 17% from normal maize (Adams et al. 2000). Dietary phytic acid may also have beneficial health roles, for example as an antioxidant or anticancer agent. The relative merits or demerits of dietary phytic acid, therefore, depends on the characteristics of the target populations. For example, children and pregnant woman in developing countries who are at the greatest risk for mineral deficiencies and dependent on cereals and legumes as staple foods may benefit from a diet that is significantly reduced in phytic acid content. On the other hand, a subpopulation that might benefit from dietary phytic acid may be aging adults in the developed world (Raboy 2002), where the anti-cancer/anti-oxidant effect of phytic acid is notable.

In the context of poultry and swine production, feeds are based largely on cereal grains and oilseed meals. Approximately two-thirds of the P in cereal grains and oilseed meals is present in the form of P bound to phytic acid (phytate P), which is not digested by poultry and largely excreted in the manure, which contributes to water pollution. Phytic acid-derived P in animal waste can contribute to environmental pollution, a

significant problem in the developed and developing countries.

A number of nutritional approaches could be explored to deal with the poor availability of phytate P in maize and the resultant potential for P pollution. Besides adding microbial phytase to poultry diets to increase phytate P availability, genetically lowering the phytic acid content of maize, thereby improving plant P availability, appears to be a long-term, sustainable biofortification approach. In 1996, USDA scientists developed two non-lethal low-phytate mutants, *lpa1-1* and *lpa2-1,* which are phenotypically identical to wild-type maize hybrids. These mutants show 60% *(lpa1-1)* and 50% *(lpa2-1)* reductions in phytic acid P, with no reduction in total P in the seed. Chick studies have demonstrated that the relative bioavailability of P in *lpa1-1* ranged from 45 to 52%, compared with 10% for wild-type maize. Phosphorus excretion was also found to be significantly reduced in chicks fed with the *lpa1-1* maize. The *lpa1* mutant does not accumulate myo-inositol monophosphate or polyphosphate intermediates. It was proposed that *lpa1* is a mutation in myo-inositol supply, the first part of the phytic acid biosynthesis pathway (Raboy et al. 2000). However, recent study by Shi and colleagues (2007) showed that *lpa1* mutants are defective in a multidrug resistance-associated protein (MRP) ATP-binding cassette (ABC) transporter that is expressed most highly in embryos, immature endosperm, germinating seed, and vegetative tissues. The *lpa2* mutant has reduced phytic acid content in seeds and accumulates myo-inositol phosphate intermediates. Maize *lpa2* gene encodes a myo-inositol phosphate kinase (ZmIPK) that belongs to the Ins(1,3,4)P3 5/6-kinase gene family (Shi et al. 2003). In 2005, maize *lpa3* mutant was identified and characterized (Shi et al. 2005). The lpa3 mutant seeds have reduced phytic acid content and accumulate myo-inositol, but not myo-inositol phosphate intermediates. Maize *lpa3* gene was found to encode a myo-inositol kinase (MIK).

Based on the phytic acid metabolic pathway, Raboy (2009) proposed four potential target intervention points for developing the low-phytate trait: (1) inositol (Ins) and Ins(3)P1 synthesis, starting with the conversion of glucose 6-P (top left) to Ins(3)P1; (2) phytic acid synthesis through either a "lipid-independent" pathway that proceeds via sequential phosphorylation of Ins and soluble Ins phosphates, or a "lipid-dependent" pathway that uses precursors that include phosphatidylinositol (PtdIns) and PtdIns phosphates; (3) transport and storage of phytin salts in globoids; and (4) phytase-encoding transgenes.

Although *lpa* mutations-based approaches significantly reduce phytic acid levels in maize seeds, considerable difficulties are often encountered because of negative unintended effects on seed and plant performance, such as compromised germination, emergence, stress tolerance, and grain filling. Although field trials of the first set of isogenic maize hybrid pairs obtained following four generation of backcrosses to a normal parent indicated only 6% yield reduction (Ertl et al. 1998), later observations on fully isogenic pairs, derived after six back crosses clearly indicated large negative effects on flowering date, germination, and stress tolerance (Raboy 2009). A recent effort, which identified *lpa1* locus coding for an ABC transporter gene, engineered embryo-specific suppression of inositol phosphate metabolism, which effectively avoids systemic disruption of phytic acid accumulation, thereby enabling targeted reduction of phytic acid levels in seeds with no undesirable consequences on whole plant performance (Shi et al. 2007). Another option is to utilize recently identified mutants such as *lpa3*, which encodes an Ins kinase whose expression is highly embryo-specific. Although *lpa3* may not affect housekeeping vegetative Ins phosphate metabolism, seed chemistry changes may still lead to unforeseen consequences that impact crop quality or performance.

Genetic manipulation of maize kernels to accumulate high levels of "phytase" is an alternative and potentially very powerful approach to problems associated with seed phytate. In an attempt to enhance the bioavailability

of iron in maize grains, Drakakaki and colleagues (2005) co-expressed a fungal phytase gene and a ferritin gene (for enhanced iron storage) in the maize endosperm so as to enhance the total pool of iron as well as bioavailability. However, it should be emphasized that the phytase has to be heat-resistant so as to remain active when maize flour is subjected to cooking processes. A successful example of thermostable phytase engineering was reported for endosperm-targeted expression in wheat (Brinch-Pederson et al. 2006). The advantage of such an approach is that mature seed chemistry is largely unaltered, thereby reducing significantly the risk for unanticipated negative impacts on plant or seed function, chemistry, or quality.

Very impressive progress has been made with respect to genetics and pathway characterization of low phytate mutations and encouraging leads have been obtained in successful engineering of the "high-phytase" trait in maize seeds. However, an important fact to remember is that phytic acid and its anabolic pathways are central to a number of metabolic, developmental, and signaling pathways vital to plant function and productivity, and low phytate may often inadvertently lead to low yield or stress susceptibility (Raboy 2009). Looking to the future, efforts to develop low-phytate maize cultivars will have to ensure high yield as well tolerance to abiotic and biotic stresses, unaccompanied by undesirable agronomic characteristics. Thermostable "high-phytase" engineering strategies promise to overcome a number of undesirable whole plant consequences that are often encountered with "low phytate"-based approaches.

Conclusions

Biofortification is an interdisciplinary science that requires coordinated efforts from breeders, geneticists, nutritionists, economists, seed system specialists, and agricultural extension specialists. As emphasized by Bouis and Welch (2010), agricultural science and nutrition have to be integrated in such a way that plant breeders incorporate nutrition as a routine target trait in their breeding programs, while nutritionists and health professionals accommodate agriculture-based approaches in their toolbox along with clinical interventions. The current collaborative biofortification efforts on maize in CIMMYT, funded substantially by HarvestPlus, target the following on a continuing and routine basis: select and breed nutritionally improved maize varieties with superior agronomic properties; carefully test promising varieties under development to ensure that the nutrients are sufficiently retained and bioavailable as consumed; develop efficient and accelerated mechanisms for testing promising materials with farmers, consumers, and other end users; and measure the nutritional impacts of these improved varieties in community-based studies where these varieties have been adopted.

Looking ahead, a number of new potential biofortification targets may be discovered in maize. These include increased concentration of prebiotics such as inulin, rafinose, and stachyose, for improved absorption and utilization of nutrients; elevated levels of ascorbic acid in maize endosperm that may result in enhanced bioavailaibility of micronutrients; and higher concentrations of tocopherols/VitaminE, which has number of beneficial antioxidant properties. The decision as to whether a breeding program is to be built on these new novel traits will depend on the nutritional value to disadvantaged and marginal communities, the ease of screening, the extent of germplasm variability, the heritability of the trait, and the trait's correlation with grain yield and other agronomic characteristics. Many important lessons could be learned from the extensive experience of institutions such as CIMMYT in developing and disseminating nutritionally enriched maize germplasm, especially QPM. These include the need for: (1) assurance of competitive agronomic performance of the nutritionally enhanced germplasm (vis-à-vis normal maize); (2) high throughput, low-cost and easily accessible phenotyping/screening tools; (3) generating awareness and capacity building of national partners on the strengths and constraints (if any) of

nutritionally enriched maize germplasm; (4) effective seed production systems; (5) strong partnerships with national research programs, and health and agricultural ministries for complementing the technologies with proper policy support and institutional innovations.

Biofortification strategies must include both breeding and improved agronomy practices, as micronutrients such as Zn and Fe are highly dependent on soil quality and farming practices. When consumed regularly, biofortified maize can contribute significantly to elevated levels of micronutrient reserves in the human body throughout the life cycle and hence should result in overall reduction of micronutrient deficiencies in a population. However as cautioned by Bouis and Welch (2010), biofortified crops alone should not be considered a panacea for eliminating micronutrtient deficiencies in all population groups, but they certainly are a powerful and sustainable tool for achieving nutritional security.

References

Arnold JM, Bauman LF, Aycock HS (1977) Interrelations among protein, lysine, oil, certain mineral element concentrations, and physical kernel characteristics in two maize populations. *Crop Sci.* 17:421–425.

Azevedo RA, Arruda P, Turner WL Lea PJ (1997) The biosynthesis and metabolism of the aspartate derived amino acids in higher plants. *Phytochemistry* 46:395–419.

Atlin GN, Palacios N, Babu R, Das B, Twumasi-Afriyie S, Friesen DK, DeGroote H, Vivek B, Pixley KV (2011) Quality protein maize: Progress and prospects. *Plant Breed. Rev.* 34:83–130.

Babu R, Nair SK, Kumar A, Venkatesh S, Sekhar JC, Singh NN, Srinivasan G, Gupta HS (2005) Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to quality protein maize (QPM). *Theor. Appl. Genet.* 111:888–897.

Babu R, Nair SK, Prasanna BM, Gupta HS (2004) Integrating marker assisted selection in crop breeding-prospects and challenges. *Curr. Sci.* 87:607–619.

Banziger M, Long J (2000) The potential of increasing the iron and Zn density of maize through plant breeding. *Food Nutr. Bulletin* 21:397–400.

Bass HW, Webster C, Obrian GR, Roberts JKM, Boston RS (1992) A maize ribosome-inactiviating protein is controlled by the transcriptional activator *Opaque-2*. *Plant Cell* 4:225–234.

Bjarnason M, Vasal SK (1992) Breeding of quality protein maize (QPM). *Plant Breed. Rev.* 9:181–216.

Bouis H (2002) Plant breeding: A new tool for fighting micronutrient malnutrition. *J. Nutr.* 132:491S–494S.

Bouis HE, Welch RM (2010) Biofortification – A sustainable agricultural strategy for reducing micronutrient malnutrition in the global South. *Crop Sci.* 50:S20–S32.

Bouis HE, Hotz C, McClafferty B, Meenakshi JV, Pfeiffer WH (2011) Biofortification: A new tool to reduce micronutrient Malnutrition. *Food Nutr. Bulletin* 32, no. 1 (supplement).

Brinch-Pedersen H, Hatzack F, Stoger E, Arcalis E, Pontopidan K, Holm PB (2006) Heat-stable phytases in transgenic wheat (Triticum aestivum L.): Deposition pattern, thermostability, and phytate hydrolysis, *J. Agric. Food Chem.* 54 4624–4632.

Brown, KH (1991) The importance of dietary quality versus quantity for weanlings in less developed countries: A framework of discussion. *Food Nutr. Bulletin* 13:86–92.

Chakraborti M, Prasanna BM, Hossain F, Majumdar S, Singh AM, Guleria S, Gupta HS (2011) Identification of kernel iron- and zinc-rich maize inbred lines and analysis of genetic diversity using microsatellite markers. *J. Pl. Biochem. Biotech.* 20:224–233.

Chakraborti M, Prasanna BM, Hossain F, Singh AM (2011b) Evaluation of single cross Quality Protein Maize (QPM) hybrids for kernel iron and zinc concentrations. *Indian J. Genet.* 71:312–319.

Diedrick TL, Frisch DA, Gengenbach BG (1990) Tissue culture isolation of a second mutant locus for increased threonine accumulation in maize. *Theor. Appl. Genet.* 79:209–215.

Dotson SB, Somers DA, Gengenbach BG (1989) Purification and characterization of lysine-sensitive aspartate kinase from maize cell cultures. *Plant Physiol.* 91:1602–1608.

Dotson SB, Somers DA, Gengenbach BG (1990) Kinetic studies of lysine-sensitive aspartate kinase purified from maize suspension cultures. *Plant Physiol.* 93:98–104.

Drakakaki G, Marcel S, Glahn RP, Lund EK, Pariagh S, Fischer R, Christou P, Stoger E (2005) Endosperm-specific co-expression of recombinant soybean ferritin and Aspergillus phytase in maize results in significant increases in the levels of bioavailable iron. *Plant Mol. Biol.* 59:869–880.

Ertl DS, Young KA, Raboy V (1998) Plant genetic approaches to phosphorus management in agricultural production. *J Environ. Qual*. 27:299–304.

Falco SC, Guida T. Locke M, Mauvais J, Sandres C, Ward RT, Webber P (1995) Transgenic canola and soybean seeds with increased lysine. *Bio/Technology* 13:577–582.

FAO (2000) Food insecurity: When people live with hunger and fear starvation. Rome: FAO.

Frisch M, Bohn M, Melchinger AE (1999a) Comparison of selection strategies for marker assisted back crossing of a gene. *Crop Sci.* 39:1295–1301.

Frisch M, Bohn M, Melchinger AE (1999b) Minimum sample size and optimum positioning of flanking markers

in marker assisted back crossing for transfer of a target gene. *Crop Sci.* 39:967–975.

Frizzi A, Huang S, Gilbertson LA, Armstrong TA, Luethy MH, Malvar TM (2008) Modifying lysine biosynthesis and catabolism in corn with a single bifunctional expression/silencing transgene cassette. *Plant Biotechnol. J.* 6:13–21.

Galili G (1995) Regulation of lysine and threonine synthesis. *Plant Cell* 7:899–906.

Ghislain M, Frankard V, Jacobs M (1995) A dinucleotide mutation in dihydrodipicolinate synthase of Nicotiana sylvestris leads to lysine overproduction. *Plant J* 8:733–743.

Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. *Trends Genet*. 21:227–233.

Graham RD, Welch RM and Bouis HE (2001) Addressing micronutrient malnutrition through enhancing the nutritional quality of staple foods: Principles, perspectives and knowledge gaps. *Adv. Agron.* 70:77–142.

Gunaratna N, De Groote H, Nestel P, Pixley KV, McCabe GP (2010) A meta-analysis of community-based studies on quality protein maize. *Food Policy* 35:202–210.

Gupta HO, Lodha ML, Mehta SL, Rastogi DK, Singh J (1980) Changes in minerals, proteins and amino acids in hard endosperm *opaque2 Zea mays* during development. *Indian J. Exp. Biol.* 18:1419–1422.

Gupta HS, Agarwal PK, Mahajan V, Bisht GS, Kumar A, Verma P, Srivastava A, Saha S, Babu R, Pant MC, Mani VP (2009) Quality protein maize for nutritional security: Rapid development of short duration hybrids through molecular marker assisted breeding. *Curr. Sci.* 96:230–237.

Harjes C, Rocherford T, Bai L, Brutnell T, Bermudez-Kandiannis C, Sowinski S, Stapleton A, Vallabhaneni R, Williams M, Wurtzel E, Yan J, Buckler E (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319:330–333.

Hellin J, Erenstein O (2009) Maize-poultry value chains in India: Implications for research and development. *Journal of New Seeds* 10:245–263.

Huang SS, Adams WR, Zhou Q, Malloy KP, Voyles DA, Anthony J, Kriz AL, Luethy MH (2004) Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. *J. Agric. Food Chem.* 52:1958–1964.

Huang SS, Frizzi A, Florida CA, Kruger DE, Luethy MH (2006) High lysine and high tryptophan transgenic maize resulting from the reduction of both 19- and 22-kD alpha-zeins. *Plant Mol. Biol.* 61:525–535.

Huang SS, Kruger DE, Frizzi A, D'Ordine RL, Florida CA, Adams WR, Brown WE, Luethy MH (2005) High-lysine corn produced by the combination of enhanced lysine biosynthesis and reduced zein accumulation. *Plant Biotechnol. J.* 3:555–569.

Krivanek AF, De Groote H, Gunaratna NS, Diallo AO, Friesen D (2007) Breeding and disseminating quality protein maize (QPM) for Africa. *African J. Biotech.* 6:312–324.

Lauderdale J (2000) Issues Regarding Targeting and Adoption of Quality Protein Maize (QPM). CIMMYT Economics Working Paper 00-02. Mexico D.F.: CIMMYT.

Lohmer S, Maddaloni M, Motto M, Difonzo N, Hartings H, Salamini F, Thompson RD (1991) The maize regulatory locus Opaque-2 encodes a DNA-binding protein which activates the transcription of the b-32 gene. *EMBO J.* 10:617–624.

Lopes MA, Takasaki K, Botswick DE, Helentjaris T, Larkins BA (1995) Identifications of two *o2* modifier loci in Quality Protein Maize. *Mol. Gen. Genet.* 247:603–613.

Machida L, Derrera J, Tongoona P, Mutanga O, MacRobert J (2012) Geostatistical analysis of Quality Protein Maize outcrossing with pollen from adjacent normal endosperm maize varieties. *Crop Sci.* 52:1235–1245.

Melchinger AE (1990) Use of molecular markers in plant breeding. *Plant Breed. Rev.* 104:1–19.

Menkir A (2008) Genetic variation for grain mineral content in tropical-adapted maize inbred lines. *Food Chem.* 110:454–464.

Mertz ET, Bates LS and Nelson OE (1964) Mutant genes that change protein composition and increase lysine content of maize endosperm. *Science* 145:279–280.

Muehlbauer GJ, Gengenbach BG, Somers DA, Donovan CM (1994a) Genetic and amino acid analysis of two maize threonine-overproducing, lysine-insensitive aspartate kinase mutants. *Theor. Appl. Genet.* 89:767–774.

Muehlbauer GJ, Somers DA, Matthews BF, Gengenbach BG (1994b) Molecular genetics of the maize (Zea mays L.) aspartate kinase-homoserine dehydrogenase gene family. *Plant Physiol*. 106:1303–1312.

Nurit E, Tiessen A, Pixley K, Palacios-Rojas N (2009) A reliable and inexpensive colorimetric method for determining protein-bound tryptophan in maize kernels. *J. Agric. Food Chem.* 57:7233–7238.

Nuss, ET, Tanumihardjo SA (2010) Maize: A paramount staple crop in the context of global nutrition. *Comprehensive Rev. Food Sci. Food Safety* 9:417–436.

Ortiz-Monasterio JI, Palacios-Rojas N, Meng E, Pixley K, Trethowan R, Peña RJ (2007) Enhancing the mineral and vitamin content of wheat and maize through plant breeding. *J. Cereal Sci.* 46:293–307.

Osborne TB, Mendel LB (1914) Nutritive properties of protein of the maize kernel. *J. Biol. Chem.* 18:1–16.

Pfeiffer WH, McClafferty B (2007) HarvestPlus: Breeding crops for better nutrition. *Crop Sci.* 47:S88–S105.

Prasanna BM, Majumdar S, Chakraborti M, Hossain F, Manjaiah KM, Agarwal PK, Guleria SK, Gupta HS (2011) Genetic variability and genotype x environment interactions for kernel iron and zinc concentrations in maize (*Zea mays* L.) genotypes. *Indian J. Agric. Sci.* 81(8):704–711.

Prasanna BM, Pixley K, Warburton ML, Xie C (2010) Molecular marker-assisted breeding for maize improvement in Asia. *Molecular Breeding* 26:339–356.

Prasanna BM, Vasal SK, Kassahun B, Singh NN (2001) Quality protein maize. *Curr. Sci.* 81:1308–1319.

Raboy V (2002) Progress in breeding low phytate crops. *J Nutr.* 132:503S–505S.

Raboy V, Gerbasi PF, Young KA, Stoneberg SD, Pickett SG, Bauman AT, Murthy PP, Sheridan WF, Ertl DS (2000) Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. *Plant physiol.* 124:355–368

Raboy, V. 2009. Approaches and Challenges to Engineering Seed Phytate and Total Phosphorus. *Plant Science* 177:281–296.

Ribaut JM, Hoisington DA (1998) Marker assisted selection: new tools and strategies. *Trends Plant Sci.* 3:236–239.

Schmidt RJ, Burr FA, Aukerman MJ, Burr B (1990) Maize regulatory gene opaque-2 encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc. Natl. Acad. Sci. USA* 87:46–50.

Scott MP, Blanco M (2009) Evaluation of the grain methionine, lysine and tryptophan contents of maize (*Zea mays* L.) germplasm in the Germplasm Enhancement of Maize Project. *Plant Genetic Resources: Characterization and Utilization* 7:237–243.

Scott MP, Darrigues A, Stahly TS, Lamkey K (2008) Recurrent selection to alter grain methionine concentration and improve nutritional value of maize. *Crop Sci.* 48:1705–1713.

Shi J, Wang H, Hazebroek J, Ertl DS, Harp T (2005) The maize low-phytic acid 3 encodes a myo-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. *The Plant J.* 42:708–719.

Shi J, Wang H, Schellin K, Li B, Faller M, Stoop JM, Meeley RB, Ertl DS, Ranch JP, Glassman K (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nature Biotechnology* 25:930–937.

Shi J, Wang H, Wu Y, Hazebroek J, Meeley RB, Ertl DS (2003) The maize low-phytic acid mutant lpa2 is caused by mutation in an inositol phosphate kinase gene. *Plant Physiol.* 131:507–515.

Shiferaw B, Prasanna B, Hellin J, Banziger M (2011) Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security* 3:307–327.

Singleton WR (1939) Recent linkage studies in maize: V. opaque endosperm-2 (*o2*). *Genetics* 24:61–63.

Solomons NW (1993) Zinc deficiency. In: Benjamin Caballero (ed.) *Encyclopedia of Food Sciences and Nutrition.* 2nd ed. London: Elsevier, pp. 4988–4994.

Sommer A Davidson FR (2002) Assessment and control of vitamin A deficiency: The Annecy Accords. *J Nutr.* 132, S2845–S2850.

Sullivan JS, Knabe DA, Bockholt AJ, Gregg J (1989) Nutritional value of quality protein maize and food maize for starter and growth pigs. *J Anim. Sci.* 67:1285–1292.

Tanksley SD (1983) Molecular markers in plant breeding. *Plant Mol. Biol. Rep.* 1:1–3.

Ufaz S, Galili G (2008) Improving the content of essential amino acids in crop plants: Goals and opportunities. *Plant Physiol.* 147: 954–961.

UNSCN (2004) 5th Report on the world nutrition situation. Nutrition for improved development outcomes. United Nations System Standing Committee on Nutrition, Geneva, Switzerland.

Vallabhaneni R, Bradbury LMT, Wurtzel ET (2010) The carotenoid dioxygenase gene family in maize, sorghum and rice. *Arch. Biochem. Biophys.* 504:104–111.

Vallabhaneni R,Gallagher CE, Licciardello N, Cuttriss AJ, Quinlan RF, Wurtzel ET (2009) Metabolite sorting of a germplasm collection reveals the Hydroxylase3 locus as a new target for maize provitaminA bio- fortification. *Plant Physiol.* 151, 1635–1645.

Vasal SK (2001) High quality protein corn. In A Hallauer (ed.) *Specialty corn*. 2nd ed. CRC: Boca Raton, FL, pp. 85–129.

Vasal SK, Villegas E, Bjarnason M, Gelaw B, Goertz P (1980) Genetic modifiers and breeding strategies in developing hard endosperm *o2* materials. In: WG Pollmer, and RH Phipps (eds.) *Improvement of quality traits of maize grain and silage use*. The Hague, Boston, and London: Martinue Nijhoff, pp. 37–73.

Wang X, Larkins BA (2001a) Genetic analysis of amino acid accumulation in *o2* maize endosperm. *Plant Physiol.* 12:1766–1777.

Wang X, Lopez-Valenzuela JA, Gibbon BC, Gakiere B, Galili G, Larkins B (2007) Characterization of monofunctional aspartate kinase genes in maize and their relationship with free amino acid content in the endosperm. *J Exp. Bot.* 58:2653–2660.

WHO (2009) Global prevalence of vitamin A deficiency in populations at risk 1995-2005. In: *WHO Global Database on vitamin A deficiency*, Geneva: World Health Organization, pp. 1–55.

Wu Y, Messing J (2011) Novel genetic selection system for quantitative trait loci of Quality Protein Maize. *Genetics* 188:1019–1022.

Wurtzel ET, Cuttriss A, Vallabhaneni R (2012) Maize provitamin A carotenoids, current resources, and future metabolic engineering challenges. *Front. Plant Sci.* 3:29.

Xu Y, Crouch JH (2008) Marker assisted selection in plant breeding: From publications to practice. *Crop Sci.* 48:391–407.

Yan J, Bermudez-Kandianis CB, Harjes CE, Bai L, Kim E, Yang X, Skinner D, Fu Z, Mitchell S, Li Q, Salas-Fernandez M, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, DellaPenna D, Brutnell T, Buckler E, Warburton M, Rocherford T (2010) Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. *Nat Genet.* 42:322–327.

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around Tm-2 locus of tomato during backcross breeding. *Theor. Appl. Genet.* 77:353–359.

Chapter 11

Marker-Assisted Backcrossing Selection for High O/L Ratio in Cultivated Peanut

Padmalatha Koilkonda, Chikara Kuwata, Masanobu Fukami, Kenta Shirasawa, Koh Aoki, Satoshi Tabata, Makoto Hasegawa, Hiroyuki Kiyoshima, Shigeru Suzuki, Shigemi Sasamoto, Atsushi Kurabayashi, Hisano Tsuruoka, Tsuyuko Wada, and Sachiko Isobe

Abstract

Peanut (*Arachis hypogaea* L.) is an important food legume and oilseed crop cultivated worldwide. The quality of peanut oil depends on the properties of the oil, one of which is the oleic/linoleic acid ratio (O/L ratio). The two genes related to the O/L ratio, *ahFAD2A* and *ahFAD2B*, were previously identified in peanut. In this chapter, marker-assisted backcrossing selection (MABS) for high O/L ratio in peanut is demonstrated with a breeding line showing a high O/L ratio ('YI-0311') and an elite line with medium O/L ratio ('Nakateyutaka') as donor and backcross parents, respectively.

A total of 204 F_2 plants were obtained from 34 F_1 plants. Seven F_2 lines exhibited a high O/L ratio and had homozygous 'YI-0311' genotypes on the two *FAD2* loci. Four F_3 seeds were generated from each of the three F_2 plants, and a total of 108 BC_1F_1 seeds were developed from the 12 F_3 plants. Twelve of the 108 BC_1F_1 plants were selected according to the genotypes of 42 genome-wide (GW) DNA markers. A total of 178 BC_1F_2 plants were developed from the 12 BC_1F_1 plants, and 16 of the 178 showed a homozygous 'YI-0311' genotype on the *FAD2* loci. Of the 16 BC_1F_2 plants, eight were selected based on the genotypes of the 19 GW markers, and then backcrossed. A total of 26 BC_2F_1 plants were developed, and 10 BC_2F_1 plants were selected according to GW genotypes. Of the subsequently developed 205 BC_2F_2 plants, nine showed homozygous 'YI-0311' genotypes on the FAD2 loci. At present (January 2012), BC Cycle 3 is in process. In addition to low cost for selection of high O/L ratio plants, the higher homozygous backcrossing ratio suggested efficiency of MABS for high O/L ratio breeding in cultivated peanut.

Introduction

Peanut (*Arachis hypogaea* L.) is an important food legume and oilseed crop cultivated worldwide in temperate and tropical zones. It covers 23 million hectares, which produce about 50 million metric tons, with an average yield 1.6 tons/ha (FAOSTAT 2010). Cultivated peanuts

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

are used for the extraction of edible oil from seeds and kernels, in addition to several other uses, such as food and fodder. Peanut breeding is more frequently performed in the public sector rather than by private companies. One of the largest breeding stations is in ICRISAT (International Crop Research Institute for the Semi-Arid Tropics), where the world's largest peanut collection, of approx. 15,000 accessions, is housed. Several national breeding programs are also active, mainly in Asia and African countries. The most important objective in peanut breeding is improvement of yield, followed by acquisition of biotic and abiotic stress tolerance. Because peanut shows relatively higher drought tolerance than other legume crops, it is often cultivated in semidry areas. Thus, drought tolerance is one of the most important targets for the improvement of abiotic stress tolerance in peanut breeding.

Meanwhile, improvement of quality characteristics in seeds, such as the modification of chemical components and reduction of allergenicity and aflatoxins, are also important targets in breeding programs for industry use (Dwivedi et al. 2007). Peanut seeds contain mostly oil, which makes up 45–51% of the weight of dry seeds (López et al. 2000). Most of the oil extracted from peanut is composed of three major fatty acids, namely, a saturated fatty acid, a monounsaturated omega-9 fatty acid, and a polyunsaturated omega-6 fatty acid, known as palmitic (C16:0), oleic (C18:1), and linoleic (C18:2), respectively. However, including these three major fatty acids, up to 12 fatty acids have been identified in peanuts (Dean et al. 2009). Oleic and linoleic acids account for 80% of the fatty acids in peanut oil (oleic acid: 36.67%, linoleic acid: 15–43%) and determine the quality of the oil (Norden et al. 1987). Accordingly, the quality of peanut oil depends on the properties of the oil, one of which is the oleic/linoleic acid ratio (O/L ratio).

There is little evidence that a high O/L ratio affects other agro-morphological characteristics of groundnut, such as yield, oil content, protein content, seed size, and so forth (Moore et al. 1989). Oleic acid is a monounsaturated fatty acid that has the ability to reduce low-density lipoprotein (LDL) levels in humans without reducing the concentration of high-density lipoproteins (HDLs), while linoleic acid is a polyunsaturated fatty acid that can decrease the levels of LDL and HDL, which is unhealthy (Yin and Cui 2006). Oleic acid is less prone to oxidation and thereby possesses an extended shelf life. High-oleate oil makes hydrogenation unnecessary, avoiding additional costs and the generation of harmful trans-fatty acids (Chu et al. 2007, Pham et al. 2010). A high oleate diet has been shown to improve the blood lipoprotein profile, suppress tumorigenesis, reduce atherosclerosis, and ameliorate inflammatory and coronary heart diseases (Yu et al. 2008; Chu et al. 2009). In addition to their uses in food, high-oleic oils also have industrial applications. The industrial oleochemicals business is investigating the use of high-oleic vegetable oils as feedstock for the production of numerous products including cosmetics and machine lubricants (e.g., high-temperature engine, transmission, hydraulic, gear, and grease applications) (Butzen and Schnebly 2007). Thus, increased oleic acid levels in peanut will not only improve oil stability, but will also support the health of consumers in addition to providing benefits for industrial applications.

Peanut is an allotetraploid species (AABB genome; 2n=4X=40). It has been assumed that the species probably originated via a single hybridization event between the two diploid species, *A. duranensis* (AA genome) and A. *ipaënsis* (BB genome) (Krapovickas and Gregory 1994; Seijo et al. 2004 and 2007). Modern cultivars are generally classified into four botanical types, that is, Spanish, Valencia, Virginia, and Southeast runner, based on their morphological traits. Despite widespread morphological variations, extremely lower genetic diversity has been observed in peanut germplasms, a factor that has slowed advances of molecular genetics and MAS in peanuts (Halward et al. 1991; Kochert et al. 1996). To date, more than

6,000 DNA markers have been developed in the *Arachis* spp. (Pandey et al. 2011), for example, random amplified polymorphic DNA (RAPD) (Garcia et al. 1996); amplified fragment length polymorphisms (AFLP) (Giemens et al. 2002); inter simple sequence repeat polymorphisms (ISSR) (Raina et al. 2001); restriction fragment length polymorphisms (RFLP) (Halward et al. 1993; Burow et al. 2001); simple sequence repeats (SSRs) derived from genomic libraries (He et al. 2003; Moretzsohn et al. 2003; Ferguson et al. 2004; Moretzsohn et al. 2005; Naito et al. 2008); and sequences of expressed sequence tags (ESTs) (Proite et al. 2007; Koilkonda et al. 2011), and bacterial artificial chromosome (BAC)-end sequences (Wang et al. 2012). Recently, a total of 504 *AhMITE1* transposon insertion polymorphic markers (hereafter referred to as transposon markers) were also developed in peanut (Shirasawa et al. 2012).

Many linkage maps have been constructed from diploid wild and cultivated species, in addition to integrated maps based on cultivated peanut species (Halward et al. 1993; Moretzsohn et al. 2005; Leal-Bertioli et al. 2009; Moretzsohn et al. 2009; Varshney et al. 2009; Khedikar et al. 2010; Hong et al. 2010; Gautami et al. 2011; Qin et al. 2011; Sujay et al. 2011; Wang et al. 2012). Although a number of studies have attempted to develop DNA markers and genetic linkage maps for peanut, the linkage groups in the latest maps have not yet converged with the number of chromosome pairs (20), which suggests that the current DNA marker resources remain insufficient in the molecular genetics of peanut. Recently, the International Peanut Genome Initiative (IPGI) was created and began research on the peanut genome project. It is expected that the activity of the peanut genome project will accelerate advances in the molecular breeding and genetics of peanut.

In marker-assisted selection (MAS), markers showing linkage with targeted traits were used for the selection of favorable individuals from breeding populations. Such markers (hereafter referred to as selection markers) can be developed through two approaches: (1) quantitative trait loci (QTL) mapping, and (2) the candidate gene approach. Several studies of QTL mapping in peanut are published, and the most frequent target in this field has been drought tolerance (Varshney et al. 2009; Khedikar et al. 2010; Gautami et al. 2011; Ravi et al. 2011). According to the results of several studies, drought tolerance is controlled by multiple QTLs, and no major QTL exists. Therefore, further studies are required before MAS can be applied for the development of selection markers for drought tolerance. Unlike drought tolerance, major QTLs were identified for biotic stress tolerance, in particular, late leaf spot (Sujay et al. 2011) and tomato spotted wilt virus (Qin et al. 2011), which showed maximum phenotypic variance of 83.0% and 35.8%, respectively. The QTLs identified for these two disease resistance traits could be applied for MAS in peanut.

With respect to the use of a candidate gene approach for the development of selection markers, genes related to the O/L ratio have been well studied in peanut. In higher plants, oleic acid is synthesized from stearic acid and is converted to linoleic acid in a reaction catalyzed by two fatty acid desaturases encoded by the *SAD* and *FAD2* genes (Yin and Cui 2006; Clemente and Cahoon 2009). When the *FAD2* genes from the commonly cultivated peanut and a high-oleate mutant that contains >80% oleate and 2% linoleate in seed oil were compared, the high-oleate phenotype was found to be caused by a single nucleotide insertion (López et al. 2000; López et al. 2002; Yu et al. 2008). Moreover, the acyl carrier protein (ACP) is a central cofactor for de novo fatty acid synthesis, carrying the nascent acyl chains during the synthesis of acyl groups. Five different types of ACP genes were also cloned from peanut (Li et al. 2010).

As the O/L ratio is a key determinant of oil and nutritional quality, the selection of mutated alleles of *FAD2*, which is associated with oleic acid content in seeds (Martin and Rinne 1986; Garces and Mancha 1991; Lee and Guerra 1994), is a simple strategy to generate high-oleic acid crops

efficiently. In peanut, *ahFAD2A* and *ahFAD2B* have been identified on the A and B genomes, respectively, and the mutant alleles are reported to confer a high oleic/linoleic acid ratio (O/L ratio) (López et al. 2000; Bruner et al. 2001; López et al. 2002; Patel et al. 2004;Chu et al. 2009; Barkley et al. 2010). In previous studies, cleaved amplified polymorphic sequences (CAPS), allele specific polymerase chain reaction (AS-PCR), and real-time PCR genotyping assays were developed for the *FAD2A* and *FAD2B* genes (Chu et al. 2007; Chu et al. 2009; Barkley et al. 2010; Chen et al. 2010; Chu et al. 2011). Accordingly, MAS for high O/L ratio appears to be already in practical use. Chu and colleagues (2011) reported gene pyramiding by MAS for nematode resistance and high O/L content.

MAS is classified into two categories: marker-assisted backcrossing selection (MABS) and marker-assisted recurrent selection (MARS). Generally, the former is employed for the improvement of traits controlled by major genes, while the latter is for pyramiding multiple QTLs. Because previous studies suggest that the O/L ratio is controlled by the two candidate *FAD2* genes, MABS is favorable for the improvement of the O/L ratio in peanut. In this chapter, MABS for a high O/L ratio in peanut is demonstrated with a breeding line showing a high O/L ratio and an elite line with medium O/L ratio as donor and backcross parents, respectively. The MABS program was started in 2007, and the number of DNA markers developed was considerably lower at that time compared to the present. Polymorphic DNA markers and a linkage map were developed in parallel with the MABS program.

Materials and Methods

Breeding Scheme

The breeding scheme of the MABS program is shown in Figure 11.1. The two parental lines, 'Nakateyutaka' and 'YI-0311,' were crossed in 2007 at Chiba Prefectural Agriculture and Forestry Research Center. 'Nakateyutaka,' which was used as a seed and backcross parent, is one of the leading varieties in Japan, whose agronomic type is Virginia and shows a normal O/L ratio (O/L=1.16). It exhibits large seed size (approx. 89g/100 seeds), light brown seed coat color, short length, and *erect* branches. 'YI-0311' is a breeding line of the Chiba Prefectural Agriculture and Forestry Research Center with a high O/L ratio (O/L=48.2), and used as a pollen and donor parent. It is Southeast-runner type, and shows small seed size (approx. 50g/100seeds), light brown seed coat color, and long and prostate branches. The F_1 seeds were sown in the field in May 2008, and F_2 seeds were harvested in October. The O/L ratios of the F_2 seeds were analyzed, and F_2 seeds showing a high O/L ratio were sown in the greenhouse during the off-season for the production of F_3 seeds. In parallel, the availability of the *FAD2* genes as selection markers for a high O/L ratio was confirmed (details are described in the next sections). The F_3 seeds were sown in the field in 2009 and backcrossed with 'Nakateyutaka' to start backcross (BC) Cycle 1. The harvested BC_1F_1 seeds were grown in the greenhouse during the off-season, and genotypes were analyzed using genome-wide (GW) DNA markers. The BC_1F_2 seeds harvested from the selected BC_1F_1 plants based on their GW genotypes were sown in pots, and the *FAD2* loci genotypes were investigated. The BC_1F_2 plants showing 'YI-0311' genotypes on the *FAD2* loci were selected and transplanted to the field for the next backcrossing cycle (BC Cycle2). At present, BC Cycle 2 is complete and BC Cycle 3 is in process.

Polymorphic Analysis of the *FAD2* Genes

The previously reported single nucleotide polymorphism (SNP) in *ahFAD2A* and the transposon insertional polymorphism in *ahFAD2B* were investigated, based on the existence of polymorphisms in the parental

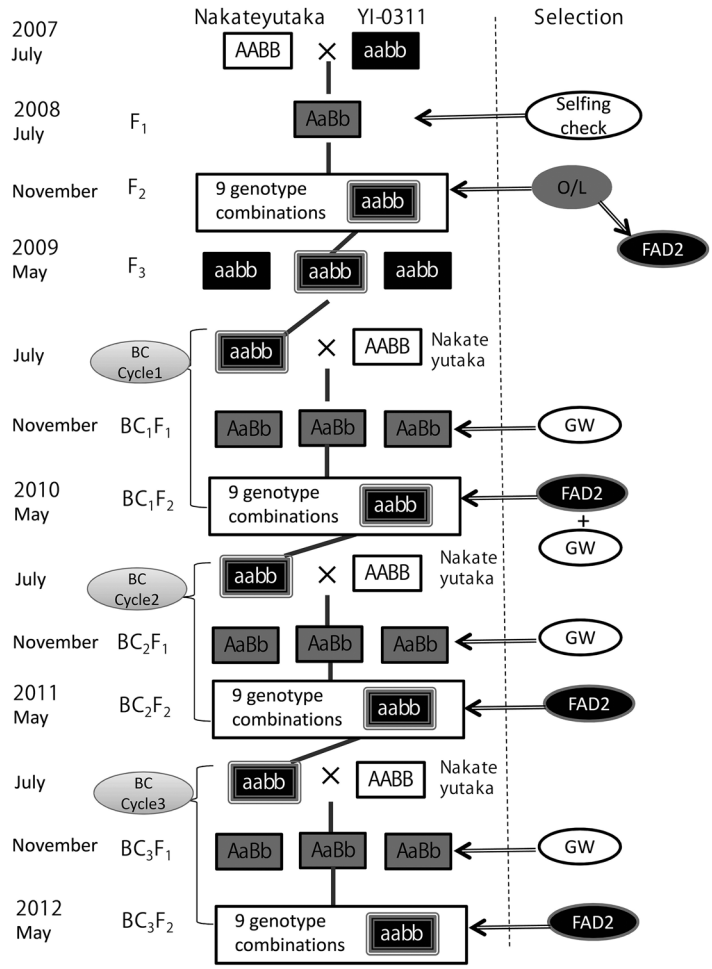


Fig. 11.1. MABS breeding scheme for high O/L ratio in peanut. '*AABB,*' '*AaBb,*' and '*aabb*' indicate genotypes of individuals on *ahFAD2A* and *ahFAD2B*. *A* and *a* show genotypes of 'Nakateyutaka' and 'YI-0311' on *ahFAD2A*, respectively, whereas *B* and *b* represent genotypes of 'Nakateyutaka' and 'YI-0311' on *ahFAD2B*, respectively. In F_2, BC_1F_2, BC_2F_2 and BC_3F_2 generations, nine genotype combinations were observed on the two *FAD2* loci: *AABB*, *AaBB*, *aaBB*, *AABb*, *AaBb*, *aaBb*, *AAbb*, *Aabb*, and *aabb*. The expected frequency of '*aabb*' genotype is 1/16, and plants showing '*aabb*' genotypes were selected for the next BC cycle. 'Self checking,' 'FAD2,' and 'GW' represent the investigation of genotypes of a DNA marker to check for existence of seeds developing from self-pollination, *ahFAD2A* and *ahFAD2B* genes, and genome-wide DNA markers.

lines and F_2 progenies (Jung et al. 2000; Lopéz et al. 2000; Patel et al. 2004). The SNP in *ahFAD2A* was identified by using the TaqMan assay with the primer pairs (5′-CCCTTCACTCTTGTCTATTAGTTCCTTAT-3′ and 5′-TGATACCTTTGATTTTGGTTTTGG-3′) and the TaqMan probes (FAM-labeled 5′-CCTCGACCGC**A**ACG-3′ for mutant allele and VIC-labeled 5′-CCTCGACCGC**G**ACG-3′ for wild-type alleles; Applied Biosystems, USA) on the 7900HT Fast Real-Time PCR System (Applied Biosystems, USA). The TaqMan assay

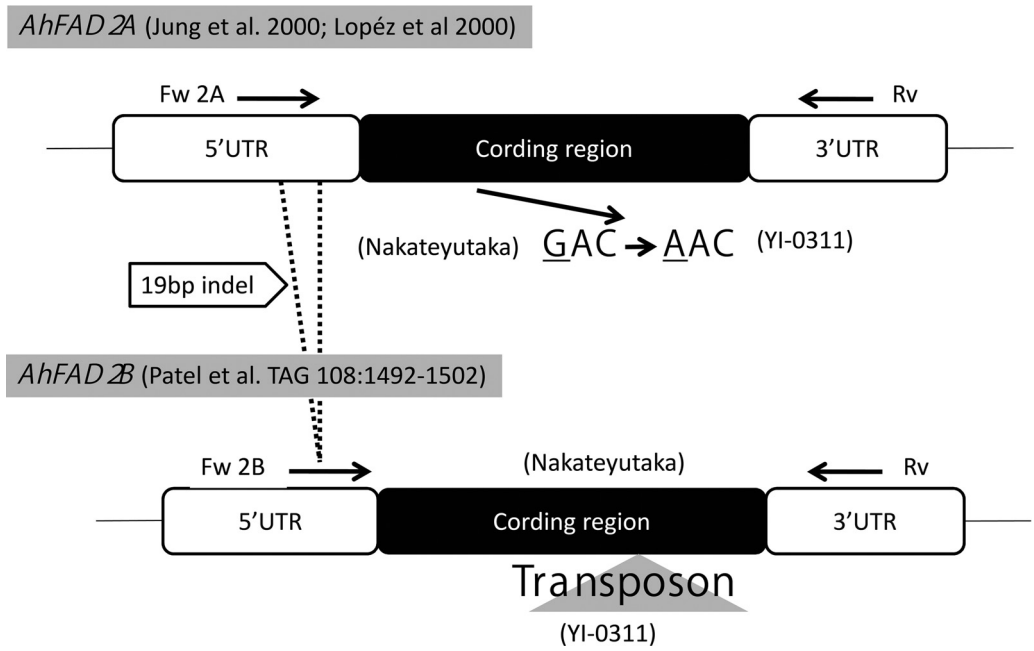


Fig. 11.2. Structures of *ahFAD2A* and *ahFAD2B* genes and mutational polymorphic sites between 'Nakateyutaka' and 'YI0311.'

was performed according to the protocol of the TaqMan Genotyping Master Mix (Applied Biosystems, USA).

The transposon insertional polymorphism on *ahFAD2B* was identified on a 2% agarose gel as a mobility difference of the DNAs amplified with primers, bF19: 5′-CAGAACCATTAGCTTTG-3′ and R1: 5′-CTCTGACTATGCATCAG-3′ (Patel et al. 2004). PCR reactions were performed in a 5 μl reaction volume using 0.5 ng of genomic DNA in 1X PCR buffer (Bioline, UK), 3 mM $MgCl_2$, 0.04 U of BIOTAQ DNA polymerase (Bioline, UK), 0.2 mM dNTPs, and 0.8 μM of each primer. The thermal cycling conditions were as follows: 1 min denaturation at 94°C; 35 cycles of 30 s denaturation at 94° C, 30 s annealing at 60° C, and 1 min extension at 72°C; and a final 3 min extension at 72° C.

The genotypes of 'YI-0311' identified on *ahFAD2A* and *ahFAD2b* were the same as the previously reported mutational alleles found on high O/L peanut lines (Jung et al. 2000; Lopéz et al. 2000; Patel et al. 2004, Figure 11.2), and the loci identified on the two candidate genes were used as selection markers for a high O/L ratio. However, because the TaqMan assay required expensive chemicals and facilities, the TaqMan SNP marker was converted to a PCR-based SNP marker that was able to identify the SNP on a 3% agarose gel for practical use (Hayashi et al. 2004). A mixture of the following four primers was used for PCR; FAD2AF: ATCCAAGGCTGCATTCTCCCT, FAD2AR4: CCTAGTGTGAGTGTGATGCAG, HachiAF: ACACCGGTTCCCTCGACCTCA and NakaR: GGTTTTGGGACAAACACTTCTTC. PCR reactions were performed in a 15 μl reaction volume using 1.5 ng of genomic DNA in 1X PCR buffer (Bioline, UK), 1.5 mM $MgSO_4$, 0.375 U of Takara Taq HS (Takara Bio Inc.), 0.2 mM dNTPs, and 0.25 μM of each primer. The thermal cycling conditions were as follows: 5 min denaturation at 94° C; 35 cycles of 30 s denaturation at 94° C, 30 s annealing at 55° C, and 1 min extension at 72° C.

Table 11.1. MABS based selection of high O/L ratio individuals in peanut

Generation	Number of all individuals	Number of selected individuals[a]	Number of individuals of FAD2/YI genotypes[b]	Number of used GW markers	Homozygous backcrossing ratio[c] of the all individuals		Homozygous backcrossing rati[c] of the selected individuals	
					Average	Range	Average	Range
F_1	58	34	–	1	–	–	–	–
F_2	204	3	7	–	–	–	–	–
F_3	12	–	–	–	–	–	–	–
BC_1F_1	108	12	–	42	55.4	38.5–69.2	62.9	52.6–68.4
BC_1F_2	178	8	16	19	62.9	33.3–93.3	70.2	53.3–80.0
BC_2F_1	26	10	–	34	83.5	73.5–91.2	87.9	85.3–91.2
BC_2F_2	205	9	9	–	–	–	–	–

[a]Individuals were selected based on genotypes of *FAD2* loci, GW markers and phenotypes.
[b]Number of individuals showing homozygous 'YI-0311' genotypes on *ahFAD2A* and *ahFAD2B* genes.
[c]Ratio of markers showing homozygous 'Nakateyutaka' genotypes to all the tested GW markers.

GW DNA Markers

The presence of selfing seeds in the F_1 generation was investigated using a simple sequence repeat (SSR) marker, PM204 (He et al. 2003). When the breeding program was started in 2007, only a few DNA markers showing polymorphisms between 'Nakateyutaka' and 'YI-0311' had been developed. Accordingly, we developed GW polymorphic markers and a linkage map in parallel with the breeding program (Koilkonda et al. 2011; Shirasawa et al. 2012). A total of 19-42 polymorphic SSR and transposon markers were used to assess GW genotypes in BC_1F_1, BC_1F_2 and BC_2F_1 progenies (Table 11.1). The GW DNA markers had not been mapped on a peanut linkage map until 2011 and the positions of the markers on the genome were therefore unknown when the genotypes were investigated.

Investigation of Fatty Acid Content in Seeds

The selected 32 F_1 parents and the two parental lines were sown in a peanut plant breeding field of the Chiba Prefectural Agriculture and Forestry Research Center in May 2008. The F_2 and the parental seeds were harvested in October and dried with shells for one month under open air conditions. One quarter of each dried seed was cut off, and 25 mg of seed material was homogenized using TissueLyzer (Qiagen, Hilden, Germany). The other parts of the seeds were sown in pots to obtain an F_3 generation. The homogenate was used for determination of fatty acid content. The quantification was performed using a GC-TOF-MS system consisting of a 6890N Network GC System (Agilent Technologies, U.S.A.) equipped with a DB-17MS column (30 m, I.D 0.25 mm, film 0.25 μm) (J&W Scientific, U.S.A.), coupled to Pegasus3 (Leco®, St. Joseph, MI, U.S.A.). Acquisition and analysis of mass spectrometry data were performed using ChromaTOFTM version 2.32 optimized for Pegasus (Leco®). The concentrations of oleic acid, linoleic acid, palmitic acid, and stearic acid were estimated from calibration curves obtained using the respective pure compound. The association between O/L ratio and genotypes on the two *FAD2* loci was detected with Genotype Matrix Mapping software and the following parameters: Max Length of Locus Combination=2; Min Number of Corresponding Samples=1; Search Range=auto (Isobe et al. 2007).

Results and Discussion

Association between FAD2 Genes and O/L Ratio

Four fatty acid compounds were identified in seeds including palmitic, stearic, oleic, and

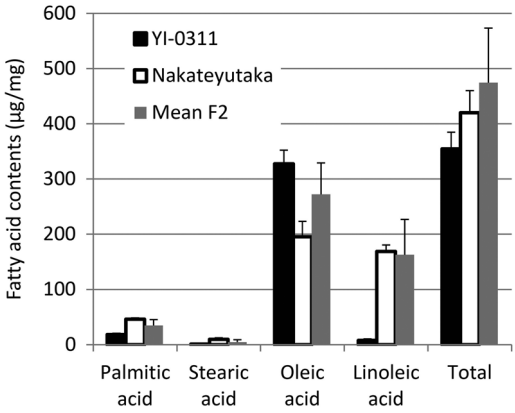


Fig. 11.3. Fatty acid content in the two parental lines and the 204 F_2 seeds. Black, white, and gray diagonal line bars indicate fatty acid content of 'YI-0311,' 'Nakateyutaka,' mean value of the total 204 F_2 seeds, and mean value of the selected three F2 seeds for high O/L ration, respectively. Black lines indicate standard deviation.

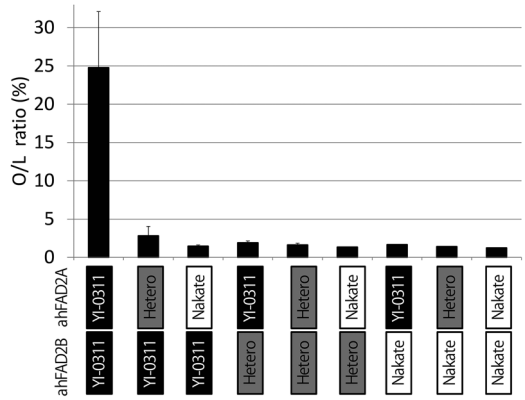


Fig. 11.4. Mean O/L ratio of the F_2 seeds of nine genotype combinations of *ahFAD2A* and *ahFAD2B* loci. 'YI-0311,' 'Hetero,' and 'Nakate' indicate homozygous 'YI-0311,' heterozygous, and homozygous 'Nakateyutaka' genotypes on the loci, respectively. Black line shows standard deviation.

linoleic acid (Figure 11.3). The content of palmitic and stearic acid was low, and the major fatty acids present in the seed were oleic and linoleic acids. The O/L ratios of the parental lines 'Nakateyutaka' and 'YI-0311' were 0.98 ± 0.11 and 49.4 ± 4.9, respectively. The O/L ratio of the F2 lines ranged from 0.84 to 35.2. The oleic acid content was positively correlated with the sum of the oleic acid and the linoleic acid content ($r=0.69$), and not significantly correlated with linoleic acid content ($r=0.13$) with a significant level of $P < 0.05$. A significant association between the O/L ratio and genotypes on *ahFAD2A* and *ahFAD2B* was identified by the GMM analysis with an F value=1619.7 (Figure 11.4). This result suggested that the genetic effects of the two loci in relation to the high O/L ratio were recessive, and a high O/L was associated with the combination recessive alleles of both *ahFAD2A* and *ahFAD2B*.

Brunner and colleagues (2001) reported that the linoleate content in cells expressing *ahFAD2A* was about 40% of that in cells expressing *ahFAD2B* when they were separately expressed in yeast (*Saccharomyces cerevisiae*). In our results, no significant differences were observed in either O/L ratio or linoleic acid content, except when both loci were derived from 'YI-0311.' This result suggests that there are no significant differences between the functions of the two *FAD2* genes in the peanut plant. Chu and colleagues (2009) investigated the genetic mutation of *FAD2* genes in peanut cultivars bred in the U.S.A., and discovered the existence of two types of mutations in *ahFAD2B* – transposon insertion and SNP – while a single mutation was observed in *ahFAD2A*. The identified mutations on *FAD2* genes in 'YI-0311' were same as those reported by Chu and colleagues (2009). This suggested that the diversity in the mutation alleles on the two *FAD2* genes is limited, and the markers developed in the present and previous studies is effective for multiple donor lines showing high O/L ratios (Chu et al. 2009; Barkley et al. 2010).

Practice of MABS for High O/L in Peanut

As indicated in Figure 11.1, the MABS program is currently in progress (2012 January), and BC cycle 3 has not been completed. In this section, we trace the progress of the MABS program from the beginning until BC Cycle2 (Table 11.1). A total of 58 seeds were obtained

from crosses between 'Nakateyutaka' and 'YI-0311.' An SSR marker, PM204, which showed polymorphisms between the parents, was used for genotyping of the seeds, and 34 of the 58 seeds were confirmed as hybrid. Six F_2 seeds were obtained from each of the 34 F_1 plants, thus, 204 F_2 plants were obtained in total. Of the 204, seven F_2 lines exhibited a high O/L (24.8 $\pm$ 6.8 on average) ratio and had homozygous 'YI-0311' genotypes on the two *FAD2* loci (Figure 11.4). Three of the seven F_2 plants were selected based on plant vigor growing in the greenhouse during off-season. The F_2 plants came into bloom in January 2009 and F_3 seeds were obtained in May 2009. Theoretically, F_2 plants could be used as parents of BC_1F_1, however, we used F_3 generation for backcrossing, largely because of insufficient temperature in the greenhouse for obtaining a high success ratio of backcrossing. Therefore, four F_3 seeds were generated from each of the three F_2 plants in the greenhouse during off-season and were sown in the field, then crossed with 'Nakateyutaka.' A total of 108 BC_1F_1 seeds were developed from the 12 F_3 plants, and genotypes were evaluated using 15 SSR and 27 transposon markers. Of a total of 42 markers, 33 were codominant polymorphic markers, while four and five markers were dominant markers of 'Nakateyutaka' and 'YI-0311,' respectively. The six markers identified the homozygous genotype of 'Nakateyutaka,' and the other 36 markers showed polymorphisms within the BC_1F_1 plants. Twelve of the 108 BC_1F_1 plants were selected according to the genotypes. The average ratio of the markers showing homozygous 'Nakateyutaka' genotypes (hereafter referred to as homozygous backcrossing ratio) was 55.4% in the all BC_1F_1 plants, while that in the selected 12 BC_1F_1 plants was 62.9%.

A total of 178 BC_1F_2 plants were developed from the 12 BC_1F_1 plants, and 16 of the 178 showed a homozygous 'YI-0311' genotype on the *FAD2* loci. In order to increase the homozygous backcrossing ratio, eight of the 16 BC_1F_2 plants were selected based on the genotypes of the 19 GW markers. The 19 GW markers were screened from the 42 markers used in the selection of BC_1F_1 plants. The average homozygous backcrossing ratio was 62.9% in all BC_1F_2 plants, while it was 70.2% in the selected eight BC_1F_2 plants. 'Nakateyutaka' was then backcrossed to the eight BC_1F_2 plants, and 26 BC_2F_1 plants were developed. Thirty-four markers selected from the 42 markers tested in the BC_1F_1 were used for GW genotype analysis, and 10 BC_2F_1 plants were selected. The average homozygous backcrossing ratio was 83.5% in all BC_2F_1 plants, while it was 87.9% in the selected 10 BC_2F_1 plants. A total of 205 BC_2F_2 plants were developed from the 10 BC_2F_1 plants, and nine of the 205 showed homozygous 'YI-0311' genotypes on the FAD2 loci. All nine BC_2F_2 plants were designated for use in the next BC Cycle3; the GW genotypes of BC_2F_2 plants were therefore not analyzed.

Transition of GW Genotypes in the MABS

A linkage map, which had been developed based on 186 F_2 plants during the MABS process, was constructed in 2011. A total of 326 segregated loci were mapped onto 19 linkage groups (LGs) on 1332.9 cM. Homeologous group (HG) numbers of the LGs were identified based on previously published maps (Moretzsohn et al. 2005; Fonceka et al. 2009; Leal-Bertioli et al. 2009; Moretzsohn et al. 2009; Varshney et al. 2009). Three corresponding LGs were generated in HG 2 (HG2.1, HG2.2, and HG2.3), while single LGs were developed in HG6 (LG6.2) and HG7 (LG7.1). No corresponding LG was identified as HG10, and one LG that showed no correspondence to previously reported maps was named LGX. The two *FAD2* loci were mapped onto LG9.1 and LG9.2.

The average ratios of alleles derived from the donor parent 'YI-0311' (hereafter AR*y*) on mapped loci are shown in Figure 11.5 for 17 LGs of the F_2 and the BC_2F_1 populations. The numbers of genotyped plants in the populations

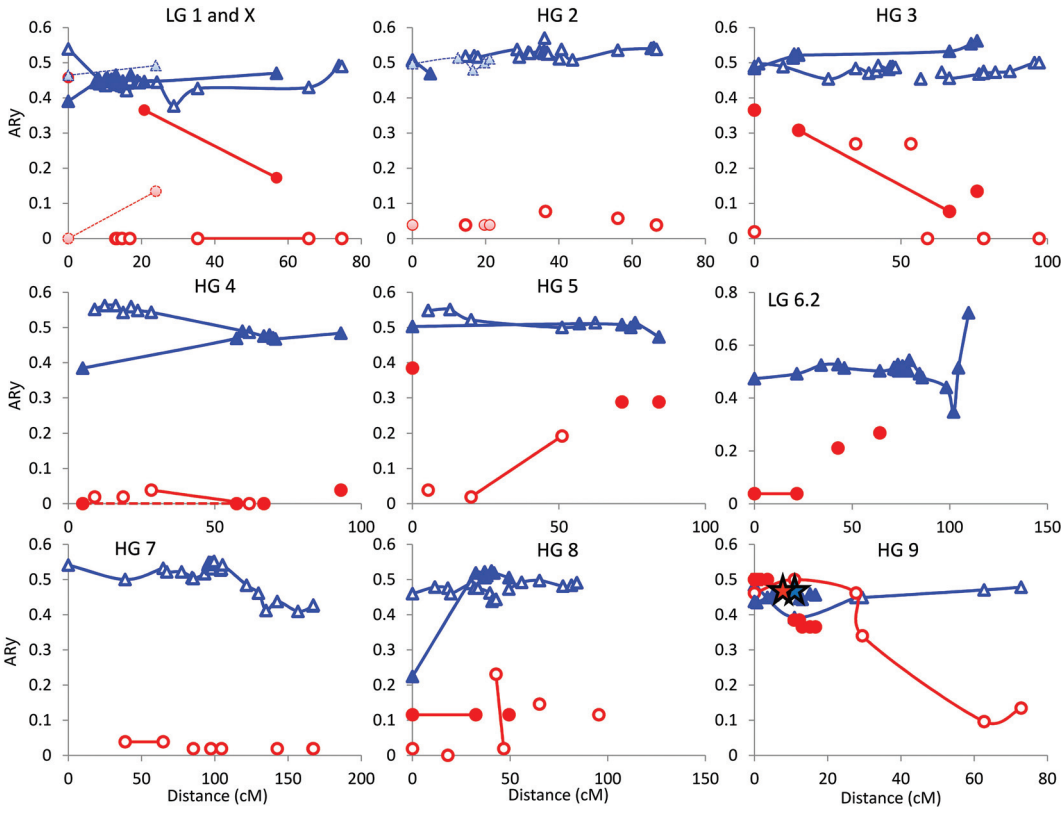


Fig. 11.5. Average ratios of the 'YI-0311' allele (AR*y*) on loci in the 186 F_2 and the 26 BC_2F_1 plants. AR*y* at a given locus was calculated by using the equation: AR*y* = (*h*+*y*×2) / (*x*+*h*+*y*) ×2, where *h*, *x*, and *y* are the numbers of homozygous alleles of 'Nakateyutaka,' 'YI-0311,' and heterozygous allele in the F_2 or BC_2F_1 populations, respectively. The graphs are classified by homeologous linkage groups (HGs) 1 to 9. LGX is shown with HG1. The X and Y axes in the graphs correspond to genetic distance on each linkage group (cM) and AR*y*, respectively. Triangles show the F2 population, while circles show the BC_2F_1 population. Open and solid symbols indicate homeologous LG with appendix 1 (ex. LG1.1), and appendix 2 (ex. LG1.2), respectively. Light-colored symbols represent LG with appendix 3 (LG2.3) or LGX. LG2.2, LG6.1, and LG7.2 were not included in the figures because of incomplete LG development or lower number of mapped markers. Red and blue stars indicate the *ahFAD2A* and *ahFAD2B* loci, respectively. For a color version of this figure, please refer to the color plate.

were 186 and 26 in the F_2 and the BC_2F_1 populations, respectively. AR*y* was calulated by using the following equation:

$$\mathrm{AR}y = (h + y \times 2) / (x + hy) \times 2,$$

where *h*, *x*, and *y* are the numbers of homozygous alleles of 'Nakateyutaka,' 'YI-0311,' and the heterozygous allele in the F_2 or BC_2F_1 populations, respectively. AR*y* values in the F2 population were approximately 0.5 on most genomic regions in the F_2 population. If random plants were selected during backcrossing, the expected AR*y* in a BC_2F_1 population was estimated as $1/2^3 = 0.125$. Of the 17 LGs, seven LGs showed AR*y* values less than 0.125 on whole regions of the LGs. Six of the 17 LGs indicated higher AR*y* values than 0.125 on partial regions of LGs. The AR*y* values of the other LGs (LG1.2, LG5.2, LG9.1, and LG9.2) were higher than 0.125 on most of regions of the LGs. The higher AR*y* values on LG9.1 9.2 seemed to be caused by the two *FAD2* loci. Because progenies with donor parental genotypes on the two *FAD2* loci

were selected, genome regions derived from the donor parent were not excluded. The reason for the higher ARy values on LG1.2 and 5.2 is not clear.

Comparison with the Other Agarose Gel Base Genotyping Technique for *FAD2* Loci

In this study, we first used the TaqMan assay for SNP identification on the *FAD2A* locus for genotyping of F_2 individuals, then the agarose gel base SNP identification assay (Hayashi et al. 2004) was employed for MAS, since our breeding program preferred SNP genotyping with simple and inexpensive equipment. As with our study, Chen and colleagues (2010) also reported an agarose gel base allele-specific PCR assay for identification of SNPs on *FAD2* loci, for simplifying the techniques and decreasing the cost of SNP genotyping. In the study by Chen and colleagues (2010), five primer sequences (one forward primer and four reverse primers) were developed. One of the four reverse primers was used for amplification of control DNA fragments, while the other three reverse primers were designed to amplify allele-specific amplicons. Wild and mutation-specific amplicons were obtained from individual PCR, then the both wild and the mutation-specific amplicons were mixed and loaded in agarose gel for electrophoresis.

While Chen and colleagues (2011) proposed a two-step PCR for identifying mutation alleles on the *FAD2* loci, in our study a one-step PCR was employed for identification of the *FAD2A* mutation. PCR was performed with a mixture of four primers, that is, a control forward primer (FAD2AF), a control reverse primer (FAD2AR), a wild type-specific reverse primer (NakaR), and a mutant-specific forward primer (HachiAF). A primer pair made up of FAD2AF and FAD2AR was used for obtaining control amplicons (293bp), whereas primer pairs FAD2AF and NakaR, and HachiAF and FAD2AR were used for obtaining wild (127bp) and mutant-specific (209bp) amplicons, respectively. In order to disturb the amplification of the *FAD2B* locus, terminal sequences of the control primer pairs were used as *FAD2A* specific bases. In addition, the third bases from the terminals of each of the primers were designed as mismatch bases for obtaining a high specificity of PCR (Hayashi et al. 2004). It is well known that the result of a PCR often changes depending on the enzymes used and the thermal cycler. Because the two-step PCR proposed by Chen and colleagues (2010) used more simple primer combinations than are used in the one-step PCR, it is expected that the two-step PCR will deliver a more stable result under diverse lab conditions. Meanwhile, the advantage of the one-step PCR over the two-step PCR is the former's lower cost and less intensive labor.

Comparison with the Other Breeding Program for High O/L with MAS

Chu and colleagues (2011) reported MAS for high O/L and nematode resistance in peanut. There are many parameters in our study that are similar to and others that contrast with the work carried out by Chu and colleagues (2011). In the study of Chu and colleagues (2011), two lines, 'Georgia-02C' and 'Florida-07,' were used as donors of the high O/L alleles and crossed with 'Tifguard.' 'Tifguard' was a donor of nematode resistance traits and used as maternal and recurrent parent. The mutation allele of 'Georgia-02C' and 'Florida-07' in *FAD2A* locus was same as 'YI-0811,' while that in the *FAD2B* locus was different from 'YI-0811,' that is, 441_442insA mutation. In the study of Chu and colleagues (2011), CAPS markers were first used for SNP identification, then the HybProbe assay (Bernard et al. 1998) was employed for polymorphic analysis of the 441_442insA mutation, in order to increase through-put and accuracy of genotyping. Both studies changed the genotyping assay to fit the breeding program, but the directions of modification were opposite. The two contrasting results indicate that there is no single correct answer for a genotyping system that uses MAS,

and modification of the system depending on the situation of the breeding program is important.

Another remarkable difference between the two studies appeared in the breeding scheme. In this study, backcrossing and selfing were alternately performed, while in the study performed by Chu and colleagues (2011), three backcrosses were continuously done first, then two selfings were performed. From the point of view of shortening the breeding period, the scheme used by Chu and colleagues (2011) was better than that used in our study. The main reason for alternating backcrossing and selfing in this study was the difficulty of performing backcrossing with the lower temperatures prevalent during off-season in the greenhouse. In order to compensate for the disadvantage, GW genotyping was performed to decrease ARv values across the genome. The ideal scheme of MABS for high O/L would be integration of the schemes used in the two studies: two backcrossings per year with genotyping of GW and FAD2 loci markers in each generation, then selfing with the genotyping of GW markers.

Merits of MABS and Future Prospects

In this breeding program, we paid approximately US$5 per sample for the chemicals used in the determination of fatty acid content, and two days of labor were required for the determination. On the other hand, the cost of the chemicals and the labor required for the investigation of genotypes on the two *FAD2* loci were approximately USD$1 per sample and one day, respectively. Although it was not used in the breeding program, a direct PCR approach for genotyping would drive the cost of materials and the required labor to less than US20¢ and half a day, respectively. These figures indicate the efficiency of MAS for high O/L ratio breeding in peanut, from the standpoint of breeding costs.

With regard to backcrossing with MAS, the results in Table 11.1 suggest a higher efficiency of marker-assisted selection compared to random selection. However, we recognize that the benefits of MABS were not fully tested in this breeding program for the following reasons. One was the lack of a sufficient number of GW polymorphic markers and information on their position on a linkage map. Large differences in the genomic backgrounds were observed across LGs in the BC_2F_1 generation. Although seven LGs were almost completely replaced by homozygous 'Nakateyutaka' genomes, four LGs still kept the 'YI-0311' genomes. The utilization of graphical genotypes would enhance selection pressure and improve the strategic selection in each generation. The MABS was very effective in many cereal crops such as barley for disease resistance, maize for earliness and yield, rice for bacterial blight, wheat for powdery mildew, and so forth (Jefferies et al. 2003; Bouchez et al. 2002; Chen et al. 2001; Zhou et al. 2005). Moreover, the yield of soybean was increased by using MABs to introgress a yield QTL from a wild accession into commercial genetic backgrounds (Concibido et al. 2003). Those species have now developed thousands of GW polymorphic markers and are able to break precisely a linkage between target genes and unfavorable genomic regions. As with those crops, recent advances in molecular genetics in peanut, especially the increasing number of developed markers and constructed linkage maps, should enable a more strategic selection than was attainable in the present study.

The other factor that limited the efficiency of MABS was the amount of breeding material used. Because selection can be done in the early growth period or in seeds, larger numbers of individuals can be used in MAS than in conventional breeding. The higher ratio of genomic regions derived from the donor genome on the four LGs in the BC_2F_1 generation indicates the need for an increased number of progenies in order to obtain more recurrent parent specific homozygous genomic regions in the generation. Increasing the amount of breeding material used is a promising solution for that problem. However, limitations in facilities and the labor required for backcrossing and plant management often

restrict the up-scaling of breeding materials used in MABS. In addition, backcrossing was not performed in off-season in this study because of lower temperatures in the greenhouse. If backcrossing were to be performed in off-season as well, the breeding term would be shortened as described by Chu and colleagues (2011). Until now, most of the studies on the molecular breeding of peanut have focused on the development of selection markers and demonstration of their utility. However, it would be important to determine an ideal balance between the required amounts of plant material and the markers used under limited facility and labor conditions. The success of MABS is not achieved only by introducing molecular work in a breeding program. The entire design of the breeding procedure including field and greenhouse management plays an important role for obtaining maximum results from MAS.

References

Barkley, N.A., Chenault, K.D., Chamberlin, C., Wang, M.L., Pittman, R.N. (2010) Development of a real-time PCR genotyping assay to identify high oleic acid peanuts (*Arachis hypogaea* L.). *Mol Breed* 25:541–548.

Bernard, P.S., Lay, M.J., Wittwer, C.T. (1998) Integrated amplification and detection of the C677T point mutation in the methylenetetrahydrofolate reductase gene by fluorescence resonance energy transfer and probe melting curves. *Anal. Biochem.* 255:101–110.

Bouchez, A., Hospital, F., Causse, M., Gallais, A., Charcosset, A. (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics* 162:1945–1959.

Bruner, A.C., Jung, S., Abbott, A.G., Powell, G.L. (2001) The naturally occurring high oleate oli character in some peanut varieties results from reduced oleoyl-PC desaturase activity from mutation of aspartate 150 to asparagine. *Crop Sci* 41:522–526.

Burow, M.D., Simpson, C.E., Starr, J.L., Paterson, A.H. (2001) Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): Broadening the gene pool of a monophyletic polyploid species. *Genetics* 159:823–837.

Butzen, S., Schnebly, S. (2007) High oleic soybean. *Crop Insights*17:1.

Chen, S., Xu, C.G., Lin, X.H., Zhang, Q. (2001) Improving bacterial blight resistance of '6078,' an elite restorer line of hybrid rice, by molecular marker-assisted selection. *Plant Breed* 120:133–137.

Chen, Z., Wang, M.L., Barkley, N.A., Pittman, R.N. (2010) A simple allele-specific PCR assay for detecting *FAD2* alleles in both A and B genomes of the cultivated peanut for high-oleate trait selection. *Plant Mol Biol Rep* 28:542–548.

Chu, Y., Ramos, L., Holbrook, C.C., Ozias-Akins, P. (2007) Frequency of a loss-of-function mutation in oleoyl-pc desaturase (*ahFAD2A*) in the mini-core of the U.S. Peanut Germplasm Collection. *Crop Sci* 47: 2372–2378.

Chu, Y., Holbrook, C., Ozias-Akins, P. (2009) Two alleles of *ahFAD2B* control the high oleic acid trait in cultivated peanut. *Crop Sci* 49:2029–2036.

Chu, Y., Wu, C.L., Holbrook, C.C., Tillman, B.L., Person. G., Ozias-Kins, P. (2011) Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *The Plant Genome* 4:110–117.

Clemente, T.E., Cahoon, E.B. (2009) Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiol* 151:1030–1040.

Concibido, V.C., La Vallee, B., Mclaird, P., Pineda, N., Meyer, J., Hummel, L., Yang, J., Wu, K., Delannay, X. (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theor Appl Genet* 106:575–582.

Dean, L.L., Hendrix, K.W., Holbrook, C.C., Sanders, T.H. (2009) Content of some nutrients in the core of the core of the Peanut Germplasm Collection. *Peanut Sci* 36: 104–120.

Dwivedi S.L., Bertioli, D.J., Crouch, J.H., Valls, J.F., Updhyaya, H.D., Fávero, A., Moretzsohn, M., Paterson, A.H. (2007) Peanut. In: Kore, C. (ed.) *Genome mapping and molecular breeding in plants*, Vol. 2, *Oilseeds* Berlin and Heidelberg: Springer-Verlag, pp. 115–151.

FAOSTAT (2006) FAO production yearbook 60. FAOSTAT, Rome, Italy.

Ferguson, M.E., Burow, M.D., Schulze, S.R., Bramel, P.J., Paterson, A.H., Kresovich, S., Mitchell, S. (2004) Microsatellite identification and characterization in peanut (*A. hypogaea* L.). *Theor Appl Gene*t 108:1064–1070.

Foncéka, D., Hodo-Abalo, T., Rivallan, R., Faye, I., Sall, M.N., Ndoye, O, Fávero, A.P. Bertioli, D., Glaszmann J.C., Courtois, B., Rami, J.F. (2009) Genetic mapping of wild introgressions into cultivated peanut: A way toward enlarging the genetic basis of a recent allotetraploid. *BMC Plant Biol* 9:103.

Garces, R., Mancha, M. (1991) *In vitro* oleate desaturase in developing sunflower seeds. *Phytochemistry* 30:2127–2130.

Garcia, G.M., Stalker, H.T., Shroeder, E., Kochert, G. (1996) Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* into *Arachis hypogaea*. *Genome* 39:836–845.

Gautami, B., Pandey, M.K., Vadez, V., Nigamm S.N., Ratnakumar, P., Krishnamurthy, L., Radhakrishnan, T., Gowda, M.V.C., Natasu, M.L., Hoisington, D.A.,

Knapp, S.J., Varshney, R. (2011) Quantitative trait locus analysis and construction of consensus genetic map for drought tolerance traits based on three recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol Breed* doi: 10.1007/s11032-011-9660-0.

Gimenes, M.A., Lopes, C.R., Valls, J.F.M. (2002) Genetic relationships among Arachis species based on AFLP. *Genet Mol Biol* 25:349–353.

Halward, T., Stalker, H.T., Kochert, G. (1993) Development of an RFLP linkage map in diploid peanut species. *Theor Appl Genet* 87:379–384.

Hayashi, K., Hashimono, N., Daigen, M., Ashikawa, I. (2004) Development of PCR-based SNP markers for rice blast resistance at the *Piz* locus. *Theor Appl Genet* 108:1212–1220.

He, G., Meng, R., Newman, M., Gao, G., Pittman, R.N., Prakash, C.S. (2003) Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol* 3: 3.

He, G., Meng, R.H., Newman, M., Gao, G.Q., Pittman, R.N., Prakash, C.S. (2003) Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol* 3:3.

Hong Y., Chen X., Liang X., Liu H., Zhou G., Li S., Wen S., Holbrook CC, Guo B. (2010) A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. *BMC Plant Biol* 10:17.

Isobe, S., Nakaya, A., Tabata, S. (2007) Genotype matrix mapping: searching for quantitative trait loci interactions in genetic variation in complex traits. *DNA Res* 14: 217–225.

Jefferies, S.P., King, B.J., Barr, A.R., Warner, P., Logue, S.J., Langridge, P. (2003) Marker-assisted backcross introgression of the Yd2 gene conferring resistance to barley yellow dwarf virus in barley. *Plant Breed*. 122: 52–56.

Jung, S., Powell, G., Moore, K., Abbott, A. (2000) The high oleate trait in the cultivated peanut (*Arachis hypogaea* L). II. Molecular basis and genetics of the trait. *Mol Gen Genet* 263:806–811.

Khedikar, Y.P., Gowda, M.V., Sarvamangala, C., Patgar, K.V., Upadhyaya, H.D., Varshney, R.K. (2010) A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). *Theor Appl Genet* 121:971–984.

Kochert, G., Halward, T., Branch, W.D., Simpson, C.E. (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theor Appl Genet* 81: 565–570.

Koilkonda, P., Sato, S., Tabata, S., Shirasawa, K., Hirakawa, H., Sakai, H., Sasamoto, S., Watanabe, A., Wada, T., Kishida, Y., Tsuruoka, H., Fujishiro, T., Yamada, M., Kohara, M., Suzuki, S., Hasegawa, M., Kiyoshima, H., Isobe, S. (2011) Large-scale development of expressed sequence tag-derived simple sequence repeat markers and diversity analysis in *Arachis* spp. *Mol Breed* doi: 10.1007/s11032-011-9604-8.

Krapovickas, A. Gregory, W.C. (1994) Taxonomía del género *Arachis* (Leguminosae). *Bonplandia* 8:1–186.

Leal-Bertioli, S.C., Jose, A.C., Alves-Freitas, D.M., Moretzsohn, M.C., Guimaraes, P.M., Nielen, S., Vidigal, B.S., Pereira, R.W., Pike, J., Favero, A.P., Parniske, M., Varshney, R.K., Bertioli, D.J. (2009) Identification of candidate genome regions controlling disease resistance in *Arachis*. *BMC Plant Biol* 9:112.

Lee, M.S., Guerra, D.J. (1994) Biochemical characterization of temperature-induced changes in lipid metabolism in a high oleic acid mutant of *Brassica rapa* Arch. *Biochem Biophys* 315:203–211.

Li, M.J., Wang, X.J., Su, L., Bi, Y.P., Wan, S.B. (2010) Characterization of five putative acyl carrier protein (ACP) isoforms from developing seeds of *Arachis hypogaea* L. *Plant Mol Biol Rep.* doi: 10.1007/s11105-009-0160-x.

López, Y., Nadaf, H.L., Smith, O.D., Connell, J.P., Reddy, A.S., Fritz, A.K. (2000) Isolation and characterization of the Δ12- fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in spanish markettype lines *Theor Appl Genet* 101:1131–1138.

López, Y., Nadaf, H.L., Smith, O.D., Simpson, C.E., Fritz, A.K. (2002) Expressed variants of Δ^{12}-fatty acid desaturase for the high oleate trait in Spanish market-type peanut lines. *Mol Breed* 9:183–190.

Martin, B.A., Rinne, R.W. (1986) A comparison of oleic acid metabolism in the soybean (*Glycine max* [L.] Merr.) genotypes Williams and A5, a mutant with decreased linoleic acid in the seed. *Plant Physiol* 81:41–44.

Moore, K.M., Knauft, D.A. (1989) The inheritance of high oleic acid in peanut. *J. of Heredity* 80:252–253.

Moretzsohn, M.C., Barbosa, A.V., Alves-Freitas, D.M., Teixeira, C., Leal-Bertioli, S.C., Guimaraes, P.M., Pereira, R.W., Lopes, C.R., Cavallari, M.M., Valls, J.F., Bertioli, D.J., Gimenes, M.A. (2009) A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. *BMC Plant Biol* 9:40.

Moretzsohn, M.C., Hopkins, M.S., Mitchell, S.E., Kresovich, S., Valls, J.F., Ferreira, M.E. (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biol* 4:11.

Moretzsohn, M.C., Leoi, L., Proite, K., Guimaraes, P.M., Leal-Bertioli, S.C., Gimenes, M.A., Martins, W.S., Valls, J.F., Grattapaglia, D., Bertioli, D.J. (2005) A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). *Theor Appl Genet* 111:1060–1071.

Naito, Y., Suzuki, S., Iwata, Y., Kuboyama, T. (2008) Genetic diversity and relationship analysis of peanut germplasm using SSR markers. *Breed Sci* 58: 293–300.

Norden, A.J., Gorbet, D.W., Knauft, D.A., Young, C.T. (1987) Variability in oil quality among peanut genotypes in the Florida breeding program. *Peanut Sci* 14:7–11.

Pandey, M.K., Monyo, E., Ozias-Akins, P., Liang, X., Guimarães, P., Nigam, S.N., Upadhyaya, H.D., Janila, P., Zhang, X., Guo, B., Cook, D.R., Bertioli, D.J., Michelmore, R., Varshney, R.K. (2011) Advances in *Arachis* genomics for peanut improvement. *Biotechnol Adv* doi: 10.1016/j.biotechadv.2011.11.001.

Patel, M., Jung, S., Moore, K., Powell, G., Ainsworth, C., Abbott, A. (2004) High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. *Theor Appl Genet* 108:1492–1502.

Patel, M., Jung, S., Moore, K., Powell, G., Ainsworth, C., Abbott, A. (2004) High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. *Theor Appl Genet* 108:1492–1502.

Pham, A., Lee, J., Shannon, J.G., Bilyeu, K.D. (2010) Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait. *BMC Plant Biol* 10:195.

Proite, K., Leal-Bertioli, S.C., Bertioli, D.J., Moretzsohn, M.C., da Silva, F.R., Martins, N.F., Guimarães, P.M. (2007) ESTs from a wild *Arachis* species for gene discovery and marker development. *BMC Plant Biol* 7:7.

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., He, G., Wang, M.L., Zhang, X., Holbrook, C.C., Ozias-Akins, P., Guo, B. (2011) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. *Theor Appl Genet* doi: 10.1007/s00122-011-1737-y.

Raina, S.N., Rani, V., Kojima, T., Ogihara, Y., Singh, K.P., Devarumath, R.M. (2001) RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Genome* 44:763–772.

Ravi, K., Vadez, V., Isobe, S., Mir, R.R., Guo, Y., Nigam, S.N., Gowda, M.V., Radhakrishnan, T., Bertioli, D.J., Knapp, S.J., Varshney, R.K. (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). *Theor Appl Genet* 122:1119–1132.

Seijo, G., Lavia, G.I., Fernandez, A., Krapovickas, A., Ducasse, D.A., Bertioli, D.J., Moscone, E.A. (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. *Am J Bot* 94:1963–1971.

Seijo, J.G., Lavia, G.I., Fernandez, A., Krapovickas, A., Ducassem D, Mosconem, E.A. (2004) Physical mapping of the 5S and 18S–25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaënsis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). *Am J Bot* 2004, 91:1294–1303.

Shirasawa, K., Hirakawa, H., Tabata, S., Hasegawa, M., Kiyoshima, H., Suzuki, S., Sasamoto, S., Watanabe, A., Fujishiro, T., Isobe, S. (2012) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. *Theor Appl Genet* doi: 10.1007/s00122-012-1798-6.

Sujay, V., Gowda, M.V.C., Pandey, M.K., Bhat, R.S., Khedikar, Y.P., Nadaf, H.L., Gautami, B., Sarvamangala, C., Lingaraju, S., Radhakrishan, T., Knapp, S.J., Varshney, R.K. (2011) Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol Breed* doi: 10.1007/s11032-011-9661-z.

Varshney, R.K., Bertioli, D.J., Moretzsohn, M.C., Vadez, V., Krishnamurthy, L., Aruna, R., Nigam, S.N., Moss, B.J., Seetha, K., Ravi, K., He, G., Knapp, S.J., Hoisington, D.A. (2009) The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). *Theor Appl Genet* 118:729–739.

Wang, H., Penmetsa, R.V., Yuan, M., Gong, L., Zhao, Y., Guo, B., Farmer, A.D., Rosen, B.D., Gao, J., Isobe, S., Bertioli, D.J., Varshney, R.K., Cook, D.R., He, G. (2012) Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol* 12:10.

Yin, D., Cui, D. (2006) Identification and molecular phylogenetic relationships of Δ12-fatty acid desaturase in *Arachis*. *Euphy* 150:347–354.

Yu, S., Pan, L., Yang, Q., Min, P., Ren, X., Zhang, H. (2008) Comparison of the Δ12 fatty acid desaturase gene between high-oleic and normal-oleic peanut genotypes. *J Genet Genomics* 35:679–685.

Zhou, R.H., Zhu, Z.D., Kong, X.Y., Huo, N.X., Tian, Q.Z., Li, P., Jin, C.Y., Dong, Y.C., Jia, J.Z. (2005) Development of wheat near-isogenic lines for powdery mildew resistance. *Theor Appl Genet* 110, 640–648.

Chapter 12

Genomics-Assisted Breeding for Tomato Fruit Quality in the Next-Generation Omics Age

Matthew P. Kinkade and Majid R. Foolad

Abstract

The cultivated tomato, *Solanum lycopersicum* L., is the most consumed and most popular vegetable crop in the world. It is the primary source of the carotenoid lycopene, a highly beneficial dietary antioxidant whose consumption may reduce the incidence of certain cancer and heart diseases in humans. Tomato is also an important model system for genetics and genomics studies that have led to many discoveries, including identification and development of some of the earliest molecular markers and genetic maps, fine mapping and cloning of the first plant disease resistance gene, and fine mapping and cloning of the first QTL. Marker-assisted selection (MAS) has been employed extensively in tomato for improving many simple traits. However, MAS has not been frequently used to advance complex traits in tomato, although many QTLs have been identified for various quantitative traits. In this chapter we look beyond the heavily reviewed QTL analysis approaches and focus mainly on the use of new "omics" technology and its potential use for tomato breeding, in particular for improving fruit quality. With the dawn of the genomics and next-generation sequencing ages, the role of genomic tools in applied tomato breeding is changing. The tomato genome has been sequenced and the information is freely available. Genomic and transcriptomic resources and bioinformatic methods have become available, metabolomic methods have been established, segregating populations have been analyzed for alterations in key metabolic traits on an "omic" scale, and large, multi-faceted omics databases have been constructed. Reverse genetics approaches, such as TILLING, have recently been employed to produce novel disease resistance and fruit quality traits in tomato. Mutagenized populations have been developed for use in TILLING approaches, and the bioinformatic workflows to handle high-throughput identification of mutations in candidate genes have been published. Thus, the pillars now exist upon which a genomics-assisted breeding scheme could be devised for tomato. The challenge is how to seamlessly incorporate these types of analyses, selection methods, and tools into a practical breeding program, and determine whether or not the time and expense required for such studies can be justified for use in contemporary tomato variety development programs.

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

Introduction

Tomato (*Solanum lycopersicum* L.) is an economically and nutritionally important crop grown throughout the world, and has been used as a model organism for genetics, genomics and physiological studies for the past 70 years. Many factors have contributed to its suitability for both basic and applied research; this includes ease of culture, short life cycle, photoperiod insensitivity, high self fertility and homozygosity, great reproductive potential, ease of controlled pollination and hybridization, amenability to asexual propagation and whole plant regeneration, and availability of a wide array of mutant and genetic stocks (http://tgrc.ucdavis.edu/; http://www.sgn.cornell.edu/). In addition, *S. lycopersicum* (hereafter, "tomato") is a diploid species with a rather small genome (~0.95 pg/1C, 950 Mbp) that lacks extensive gene duplication. Throughout the past 70 years, modulating tomato fruit quality to suit grower and end-user specifications has been a focal point in many tomato breeding programs (Foolad 2007a). As with nearly all crops in the United States, practical tomato breeding was originally led primarily by researchers at public institutions, and breeding efforts and various germplasm collection expeditions by C.M. Rick and others have formed the basis of the present-day tomato cultigen. Because of the justified focus on tomato by early genetics and molecular biology researchers, there have also been many widely-heralded discoveries using the tomato system, including but not limited to identification and development of some of the earliest plant molecular markers and genetic maps (Tanksley et al. 1982; Tanksley and Orton 1983), and fine mapping and cloning of the first plant disease resistance gene, *Pto* (Martin et al. 1993a; Martin et al. 1993b), and the first QTL, *fw2.2* (Frary et al. 2000). Tomato has also been an excellent system for breeding purposes, evidenced by the continual introgression of desirable genes and phenotypes from wild species and improvement of the crop disease resistance, fruit quality, and yield (Foolad 2007a).

Brief History of Tomato

The *Solanum* genus is thought to have evolved approximately 12 million years ago (Mya), with the tomato clade (section Lycopersicon) radiating from the potato clade (section Solanum) at about 7 Mya (Wikström et al. 2001). Since then, tomato species have evolved to inhabit a vast array of elevations, soil types, and climates (Taylor 1986). In fact, although a tropical plant, tomatoes are now grown in some form in every region of the world, from the tropics to within a few degrees of the Arctic Circle (Foolad 2007a). *Solanum* sect. Lycopersicon is monophyletic, consists of 13 different species, and has been thoroughly reviewed (Labate et al. 2007). The present-day cultivated tomato is the product of more than six centuries of domestication and selective breeding, presumably initiated by Mayan agriculturalists (Kalloo 1991; Peralta and Spooner 2005). The cultivated tomato is thought to have originated either in Mexico or Peru, and an ensuing controversy over the exact geographic origin has persisted since the late 19th century (Peralta and Spooner 2005). However, despite the wide distribution of the genus in the Andean region, Mexico has been considered the most likely center of domestication (Rick 1976b). Regardless of origin and place of domestication, natives were already cultivating tomato when European explorers arrived in the Americas in the late 15th century, and subsequently the Spaniards introduced the crop to the Old World early in the 16th century (Rick 1976b; 1978; Peralta and Spooner 2005). Europeans were initially suspicious of tomato due to its morphological resemblance to poisonous nightshade, thus preventing widespread use of the tomato as a food until the late 19th century (Rick 1978). Commercial production of tomato in the U.S. began in 1847 at Lafayette College, in Easton, Pennsylvania, leading to a major vegetable production industry in the mid 20th century (Foolad

2007a). Since then, the tomato has become an integral part of the human diet as well as an incredibly useful organism for the study of plant genetics and breeding.

Economic and Nutritional Value of Tomato

The tomato is a major vegetable crop produced and consumed around the world, and has enjoyed a vast expansion in production over the last 50 years. In 1961, total world production of tomato was about 26.2 million tons; in 2009, total world production was about 131 million tons (FAOSTAT 2012), as a result of increased acreage and advances in mechanized agriculture, industrial processing, and plant breeding. China is the world leader in tons of tomatoes produced (~41.9 million tons), followed by the United States [14.2 million tons; (FAOSTAT 2012)]. However, it is estimated that the agricultural value of tomatoes produced in the U.S. is more than twice that of tomatoes produced in China [gross production value per ton in constant 2004–2006 US$; (FAOSTAT 2012)]. In addition to economic value, fresh tomatoes and tomato-based products constitute an important supplier of nutrients to humans by virtue of total volume consumed. According to the United States Department of Agriculture, a 100 g serving of tomato paste contains 36 mg calcium, 1.014 g potassium, 21.9 mg vitamin C, 38.5 mg choline, 900 μg β-carotene, 28.8 mg lycopene, and smaller amounts of other vitamins, minerals, and bioactive components (USDA 2012). Tomato is the premier dietary source of the carotenoid lycopene; approximately 90% of the average human's yearly intake of lycopene is from tomato or tomato-based products (Rao and Rao 2007). Recently, this compound has been identified as a highly beneficial dietary antioxidant via studies that have shown that the presence of this compound is inversely related to the occurrence of certain types of cancer, as well as chronic heart diseases (Rao and Rao 2007). The mode of action of lycopene is thought to be via its system of conjugated double bonds, enabling the compound to quench reactive oxygen species (ROS), triplet chlorophylls (in plants and phototrophic bacteria), and UV radiation (Demmig-Adams and Adams 2002). It has been shown in human epidemiological studies that intake of lycopene results in a heightened total antioxidant capacity, and lower levels of oxidized protein and serum lipids (Rao 2004). This finding is significant because increased levels of oxidation are correlated with increased incidence of various diseases, including heart disease, cancer, inflammation, and macular degeneration (Demmig-Adams and Adams 2002). A nested case-control study concluded that there was a significant inverse relationship between plasma lycopene concentration, dietary intake, and prostate cancer risk in humans older than 65 without a family history of incidence (Wu et al. 2004). These findings imply that dietary antioxidants may have a prohibitive effect on spontaneous prostate cancer occurrence, but this effect may be secondary to the influence of a strong hereditary component. Despite the conclusions of several epidemiological studies, which indicate that increased plasma lycopene may have a prohibitive effect against the incidence of prostate cancer, a recent critical review of all studies concluded that a defensible link between lycopene and reduced cancer incidence could not be made due to various confounding factors (Von Löw et al. 2007). The U.S. Food and Drug Administration recently issued a report examining two requests by several commercial entities wishing to present qualified health claims on food labels regarding the ability of lycopene to prevent certain human cancers, and the agency concluded that there was little evidence supporting the association of tomato consumption with reduced cancer risk (Kavanaugh et al. 2007). It is expected that more definitive clinical or in vitro trials will be conducted in the near future in order to support or refute this link. Although much has been accomplished in terms of identifying the effect of lycopene on disease prevalence, its actual mode of action

at a molecular level has yet to be definitively elucidated.

Practical Breeding Considerations

One of the goals of pre-breeding research in tomato is the discovery of novel genetic elements that could be used by tomato breeders to modulate important traits in a practical manner. Since transgenic approaches for tomato genetic improvement have remained generally unacceptable to consumers, new genetic variation has either been introgressed into the cultigen from closely related wild species or generated via mutagenesis. Due to the fact that most tomato fruit quality traits, such as soluble solids content (SSC), viscosity, and carotenoid accumulation, exhibit complex inheritance and interact with the environment, the quantitative trait locus (QTL) approach to identifying useful genetic variation has been utilized extensively for pre-breeding purposes. This technique also has allowed researchers to develop mapping populations and identify QTLs within the context of germplasm improvement. With the development of publicly available molecular maps and core sets of markers, numerous quantitative genetics studies have been conducted using various types of intra- and interspecific populations of tomato, and hundreds of putative QTLs affecting fruit quality traits such as SSC, carotenoid content, fruit size, acidity, firmness, and taste have been reported and extensively reviewed elsewhere (Foolad 2007a, b; Labate et al. 2007). Concurrently, much of the commercial-scale tomato breeding capacity shifted from public to private institutions. Thus, published results from subsequent verification of identified QTLs in relevant genetic backgrounds and environments has been limited, with a few notable exceptions (e.g., Lecomte et al. 2004; Yates et al. 2004; Chaïb et al. 2006; Kinkade 2010). Further, most detected QTLs have not been used for practical breeding purposes, with a few exceptions, such as QTLs LIN5 and LS1, which purportedly increase fruit SSC and are found in some of today's processing tomato cultivars (Fridman et al. 2000; Schaffer et al. 2000; Causse et al. 2004; Foolad 2007a). The use of reverse genetics approaches, such as "targeting induced local lesions in genomes" (TILLING), for applied breeding purposes has, until recently, been hindered by the considerable time, labor, and cost required to produce and screen mutant populations. Recently this approach has been used in tomato to produce novel disease resistance (Piron et al. 2010) and fruit quality traits (Gady et al. 2012). With the advent of high-throughput point mutation detection schemes utilizing cheap next-generation sequencing technologies and novel bioinformatic pipelines, mutation breeding has re-emerged as an attractive approach to generating new genetic variation in tomato breeding material (Gady et al. 2009; Rigola et al. 2009).

In this chapter we look beyond the heavily reviewed QTL analyses that have dominated tomato literature until recently. Although many QTLs have been identified for various fruit quality traits in tomato, extremely few are actively used in today's commercially grown tomato cultivars (Foolad 2007a; Kinkade 2010). There are several legitimate reasons for this discrepancy. Many detected QTLs and their associated molecular markers are based on interspecific mapping populations of tomato, and thus may be specific to those populations. Subsequent verification of such QTLs, their individual effects, and associated markers in breeding populations is a time-consuming, expensive, and risky task. This is especially true when the QTLs are originally described in genetic populations distantly related from or irrelevant to a given breeder's germplasm and/or in different agricultural environments from a given breeder's target climatic region. Also, the lack of a universal set of high-throughput marker assays that can be utilized to discriminate between genotypes within the cultivated tomato germplasm has hindered the widespread use of genomics information to assist the breeding process. In tomato, these concerns and roadblocks are

slowly being dealt with, as discussed later in this chapter.

The most popular inbred development technique currently used by practical tomato breeding programs is pedigree-based selection, with the goal of producing desirable inbred lines and subsequently F_1 hybrid cultivars utilizing the inbred lines as parents. By nature, this process favors traits that are simply inherited and can be visually evaluated. Further, traits such as yield, vine cover and size, disease resistance, and maturity are generally considered higher breeding priorities than modulating complex fruit quality traits such as viscosity and carotenoid content. With a suitably large number of diverse materials entering the inbred development pipeline, one can obtain desired fruit quality parameters using phenotypic selection; but obtaining inbred lines with the combination of these parameters alongside improved disease resistance and/or acceptable yield remains difficult, since selection in early generations focuses primarily on horticultural traits, not fruit quality. For example, in a processing tomato breeding program, a plant with an unacceptably small vine but desirable viscosity is not useful and often would not be selected. Thus, in a hypothetical breeding program where marker-assisted selection (MAS) is not applied for fruit quality traits, the number of inbred lines with desirable horticultural characteristics, disease resistance, and thick viscosity at the end of an inbred development pipeline is a function of the number of F_2 populations used, the frequency and number of alleles conferring the desirable fruit quality characteristics, the presence of genetic linkages, the breeder's skill, and random chance. In most cases, this situation results in a low probability of success, as well as a highly inefficient germplasm development program. In tomato, MAS is widely used for single-gene traits, such as disease resistance or determinacy, but traits with more complicated inheritance, such as sugar profile, viscosity, and carotenoid profile, are still largely subject to traditional phenotypic selection (Foolad 2007a). However, if one desires "enhanced β-carotene content" as a breeding objective, while also taking into account the grower requirements for successful tomato varieties, utilizing solely traditional breeding techniques is exceedingly inefficient (Barone et al. 2009). This area of difficulty is precisely where "omics" tools applied to large, relevant, diverse populations, using large numbers of informative molecular markers and high-throughput phenotyping protocols, can have the largest impact on practical tomato breeding.

With the dawn of the age of genomics and next-generation sequencing, the role of genomics in applied tomato breeding has come into focus. The overall goal of applying genomics strategies to aid breeding, that is, association of phenotypic variation with sequence variation on a genome-wide scale, will not change. Successful application of next-generation "omics" techniques in combination with conventional breeding techniques may result in the identification and commercial application of novel traits, as seen with other crop species (Tuberosa and Salvi 2006; Tuberosa et al. 2007; Dahmani-Mardas et al. 2010). Such traits may be difficult to deal with using conventional breeding techniques alone. For example, various tomato fruit quality traits are expensive to measure, and modulating these traits at the phenotypic level often results in negative consequences to fruit yield (see Schauer et al. 2006). The tomato genome has been sequenced and the information is freely available (Tomato Genome Consortium 2012), genomic and transcriptomic resources and bioinformatic methods are publicly available, metabolomic methods have been established and reported in the literature, and research groups are beginning to analyze segregating populations for alterations in key metabolic traits on an "omic" scale, while concurrently collecting data from relevant agronomic traits. Further, mutagenized tomato populations have been developed for use in TILLING approaches (Menda et al. 2004; Minoia et al. 2010; Okabe et al. 2011) and the bioinformatic workflows to handle

high-throughput identification of mutations in candidate genes have been published (Gady et al. 2009; Rigola et al. 2009). All of this is leading to the capability of using contemporary combinatorial "omic" methods to achieve breeding outcomes in tomato; whether or not practical tomato breeding outcomes will be obtained and/or explicitly reported in the literature using these methods remains to be seen.

Key Advances Enabling Genomics-Assisted Breeding in Tomato

The tools required for genomics-assisted breeding in tomato are available: the tomato genomic sequence has become available for public use, early proof-of-concept "omics" studies have been conducted on genetic populations of tomato, large-scale marker identification projects have produced multitudes of markers for use within the tomato cultigen, new TILLING populations have been developed and characterized, and large, multi-faceted omics databases have been constructed, which allow users to compare multi-omics datasets in order to construct correlational networks of genes. Yet, as in many other crop species, single-gene MAS (mainly for disease resistance) remains the most widely-used molecular breeding technique in practice for tomato, and the status of "genomics-assisted breeding" of tomato in the literature remains nascent at best. In fact, a literature search of "genomics assisted breeding of tomato" currently returns a litany of review articles but scant primary research reports. Nevertheless, the pillars upon which a genomics-assisted breeding scheme could be devised now exist for tomato. In this section, we describe these resources and some early examples from the literature that have begun to bridge the gap between basic genomics and practical breeding outcomes related to tomato fruit quality. The challenge is how to seamlessly incorporate these types of analyses and selection methods into a practical breeding program, and determine whether or not the time and expense required for such studies can be justified to the contemporary tomato breeder.

Early Breeding Research and the *S. pennellii* LA716 Introgression Lines

The cultivated tomato, *S. lycopersicum* L., has become a model system for quantitative genetic studies for a variety of reasons. Tomato is a diploid organism and a self-pollinator, seed-to-seed generation time is as little as three months, controlled hybridizations are easy to conduct, wide phenotypic variation exists within the available germplasm, and diverse genetic populations are relatively simple to construct (Foolad 2007a, b; Labate et al. 2007). The body of work led by C. M. Rick and colleagues paved the way for tremendous genetic improvement of tomato varieties using accessions from the related wild tomato species (i.e., *Solanum* sect. Lycopersicon species other than *S. lycopersicum*). Based on collections and botanical studies of wild tomato species as well as isozyme surveys of genetic variation within *Solanum* sect. Lycopersicon, which indicated extremely low genetic variability within the cultivated species, it was determined that wild germplasm held the key to the improvement of many important agricultural traits in tomato (*see* Rick and Fobes 1974, 1975; Rick 1976a, 1978, 1979). As a result, some breeders and geneticists delved back into the wild tomato germplasm (accessions of botanically related, but undomesticated tomato *Solanum* species) in order to identify and transfer beneficial traits into breeding lines. As early as 1982, molecular markers (specifically isozymes) were being used to map tomato genes involved with phenotypic characters (Tanksley et al. 1982). These developments resulted in the creation of segregating interspecific populations of tomato, and, combined with collaboration between public and private researchers, enabled the genetic analysis of various phenotypic traits (reviewed in Foolad 2007a; b). Since then, a wealth of QTL

mapping studies has been conducted using the tomato system (reviewed in Foolad 2007a and Labate 2007), which has subsequently sparked in-depth transcriptomic and metabolomic studies, in addition to sequencing the tomato genome (Mueller et al. 2009). Combined with the tomato genome sequence, such information can now be leveraged in the search for candidate genes underlying detected QTLs, or in the search for new QTLs. As a result, researchers have used segregating interspecific populations to identify new alleles affecting various traits and to subsequently clone and characterize them (reviewed in Foolad 2007a). A majority of studies has been conducted using most widely known tomato population, the *S. pennellii*-based introgression lines [ILs; (Eshed and Zamir 1995)]. Many other tomato "immortalized" mapping populations with useful traits have been constructed (see Foolad 2007b; Ashrafi et al. 2009) and are expected to be heavily utilized in future for both strategic and applied breeding research.

Although more time-consuming to develop than many other mapping populations, IL populations can be more useful for QTL mapping and phenotyping, since any significant difference(s) between an IL and the recurrent parent is due to a single introgressed segment of the donor *S. pennellii* genome. However, once QTLs are detected, the physical size of the genomic bin in which each QTL resides may be quite large; therefore time saved by removing background wild genomic intervals prior to QTL analysis may be moot, as further backcrossing, marker mapping, recombinant selection, and validation experiments are still required prior to practical application. Thus we posit that the overall "work" required to identify and transfer QTLs to elite germplasm using ILs is similar to other population types (see Barone et al. 2009). Nevertheless, in tomato, the *S. pennellii*-based IL population has been the foundation of many genetic discoveries and continues to be a useful tool for contemporary omics experimentation in the public sector (Liu et al. 2003; Rousseaux et al. 2005; Schauer et al. 2005; Schauer et al. 2006; Fraser et al. 2007a; Lippman et al. 2007; Bermudez et al. 2008; Schauer et al. 2008). QTLs for altered soluble solids content, yield, fruit size, metabolic traits, and many others, have been identified using the ILs, and some of these QTLs are actively used by commercial breeding programs (Rousseaux et al. 2005; Lippman et al. 2007). This population has also been used for metabolomic studies (Schauer et al. 2006; Fraser et al. 2007a; Lippman et al. 2007; Bermudez et al. 2008; Schauer et al. 2008), which will be discussed later in this chapter. In addition, markers developed using the interspecific populations constructed for quantitative genetics studies helped form the basis of the tomato genome sequencing project.

Sequencing of the Tomato Genome

The tomato genome sequencing project officially began in 2004 as an international collaboration among ten countries (Mueller et al. 2009). A BAC-by-BAC approach, beginning with physical mapping and BAC (bacterial artificial chromosome) tiling followed by "golden path" sequencing, was initially proposed, and as sequencing technology advanced rapidly during the 2000s, this approach was utilized to buttress next-generation sequencing datasets (Mueller et al. 2009). A formal report on the genome sequence and insights into the evolution of tomato has recently been published (Tomato Genome Consortium 2012). An FTP site within the Solanaceae Genomics Network (SGN; http://solgenomics.net) for mass data download has also been developed for researchers wishing to conduct their own annotation and/or sequence investigation (such as repeat masking, transposon detection, gene identification, etc.; Mueller et al. 2009). Gene identification using the raw sequence and the MAKER annotation tool (Cantarel et al. 2007), followed by track visualization using Apollo (Lewis et al. 2002), is now possible and has been

conducted for QTL candidate gene identification and marker development in targeted genomic regions (Kinkade 2010; Kinkade and Foolad, unpubl. results).

New Tomato Genomics Resources

A publicly available genome sequence has supported the direct application of genomics to tomato breeding on a number of levels. For the first time, tomato researchers are now able to identify potential markers directly within regions of interest, design primers using the genome sequence, and screen for polymorphism within their population of interest (Kinkade 2010; Kinkade and Foolad, unpubl. data). In addition, as genotyping-by-sequencing (GBS) becomes more cost-effective, a reference genome will assist researchers in mapping the tens of thousands of putative single nucleotide polymorphisms (SNPs) identified by this process within contemporary breeding populations. An extension of this approach has been taken up by the Solanaceae Coordinated Agricultural Project (SolCAP; http://solcap.msu.edu). Focused on improving genomics resources and training for potato, tomato, and pepper, and translating these resources into applied outcomes, SolCAP has introduced a highly successful series of training articles and videos on its eXtension website (http://www.extension.org/plant_breeding_genomics). SolCAP has identified thousands of SNPs within the cultivated tomato (Sim et al. 2012) and potato (Hamilton et al. 2011) germplasm and made them available on an Infinium array for the breeding community to utilize. At the very least, this project has provided successful training tools and a substantial, universal set of informative SNP markers to the broader tomato community. Applied breeding outcomes in tomato by the SolCAP consortium have yet to be formally published; future research as part of SolCAP is anticipated to elucidate alleles controlling variation in carbohydrate and vitamin metabolism.

In parallel to the SolCAP and tomato sequencing projects, the Tomato Functional Genomics Database [TFGD; (http://ted.bti.cornell.edu/; Fei et al. 2011)] has made a variety of expressed sequence tag (EST) datasets, metabolite analyses, microarray experiments, and small RNA experiments available for public query and analysis. Rather than simply act as a repository for experimental results, the TFGD has expanded its capabilities for analysis of microarray data, sRNA (small RNA) experiments, and metabolomic datasets (Fei et al. 2011). This has enabled researchers to analyze data in a consistent, reliable manner – one of the key developments that must occur if we are to build upon previous knowledge and utilize it in a practical way. Each of these resources has been integrated (or at least linked) with SGN, which has positioned SGN as the current nexus for tomato omics information (maps, markers, analysis tools, phenotypes, and genotypes). A "breeder's toolbox" has been made available for researchers searching for markers, phenotypes, and/or QTLs, although the practical utility of information contained within this portal is currently limited. Nonetheless, massive amounts of tomato genomics information and resources are freely available through SGN, and these resources continue to support applied researchers.

Lastly, reverse genetics approaches to tomato germplasm improvement have emerged as a cost-effective way to generate novel genetic variation. With the development of efficient bioinformatics pipelines to detect point mutations in large populations, coupled with massively parallel sequencing strategies utilizing cheap next-generation sequencing technologies, TILLING can now compete with forward genetics strategies for pre-breeders' attention. Combined with the fact that a large body of basic plant biology research has been conducted using the tomato system, which can be leveraged to target specific genes or gene families involved in important phenotypes, this approach is well-suited for use in tomato genetic improvement. In addition, researchers can obtain and evaluate many

lines with mutations in candidate genes much more quickly than developing materials using the QTL method. Another advantage to TILLING compared to the QTL approach is that the mutations are already incorporated into a completely isogenic background; once useful traits are identified, they can be easily transferred to breeding populations. TILLING populations have been developed using the processing tomato lines M82 (Menda et al. 2004) and Red Setter (Minoia et al. 2010), as well as the laboratory-friendly Micro-Tom (Okabe et al. 2011). At the very least, these populations are important proof-of-concept examples for other researchers, and will most likely serve as integral resources for pre-breeding efforts in the future. Gady et al. (2009) and Rigola et al. (2009) report three different high-throughput methods of detecting point mutations in tomato EMS (ethyl methanesulfonate) mutant populations, and each method results in an acceptable mutation detection rate in the populations analyzed. Both reports describe the use of next-generation sequencing technologies coupled with elegantly tailored pooling strategies.

Fruit Quality Traits Targeted for Genomics-Assisted Breeding in Tomato

As discussed earlier, many agronomically important traits in tomato can be easily phenotyped and as a result, conventional phenotypic selection and field-based breeding practices remain most effective for developing improved tomato breeding lines and commercial cultivars. However, other traits, such as sugar metabolite profile, carotenoid content and profile, and flavor profile are more difficult to phenotype in breeding-scale systems and therefore tend to be neglected by practical breeding programs. These complex traits have been targeted by tomato researchers in an attempt to provide tools and analyses that utilize contemporary protocols and equipment, complement conventional breeding practices, and produce real outcomes. Such studies are described in the following sections.

Primary Metabolites

The observed phenotypes associated with fruit quality traits are inherently the result of metabolite production and composition within the fruit, and the genetic components of the phenotypic variances for such traits have been shown to be substantial. Early studies reporting metabolic profiling of segregating populations and attempting to construct correlational networks of genes involved with metabolic QTLs have made significant progress (Schauer et al. 2005; Schauer et al. 2006; Fraser et al. 2007a; Bermudez et al. 2008; Schauer et al. 2008). Recently tomato genomics-assisted breeding research has benefited from several studies aiming to identify and describe variation in primary metabolites. These metabolites play important roles in the determination of key tomato fruit quality attributes, such as soluble solids and acidity. What is widely accepted is that primary and secondary metabolite levels can be substantially different between tomato genotypes, even within breeding germplasm, and also differ as a result of environmental conditions. Deeper understanding of the inheritance of primary metabolite QTLs and thorough investigation of the ultimate effect (or lack thereof) of modulating the accumulation of these metabolites on fruit quality is still required in order to establish legitimacy of these QTLs. Recent studies have sought to explore these questions using the tomato system (Schauer et al. 2006; Bovy et al. 2007; Bermudez et al. 2008; Schauer et al. 2008).

Current high-throughput metabolic profiling protocols are amenable to quantifying aqueous compounds; therefore, such species of compounds were the focus of the first metabolite profiling studies in tomato. Overy and colleagues (2005) examined the metabolic profiles of one *S. lycopersicum* accession, one *S. pennellii* accession, and 5 *S. pennellii* IL accessions, and demonstrated that metabolic profiling combined

with principal component analysis could indeed be used to distinguish tomato genotypes. A similar, yet more robust experiment was carried out by Schauer and colleagues (2005), who reported GC/MS (gas chromatography/mass spectrometry) profiling of accessions of 5 *Solanum* species and described substantial differences in sugars, organic acids, amino acids, and other metabolites between the different species. Neither of these studies examined large populations in an agricultural setting, as opposed to studies conducted by Schauer and colleagues (2006 and 2008), who reported extensive evaluations of the *S. pennellii* IL population for metabolic profile as well as agronomic traits. Schauer et al. (2006) were the first to evaluate an entire genetic population of tomato for metabolite profile, in an agriculturally relevant setting, along with agronomic traits such as soluble solids, yield, fruit size, and so forth. This study elucidated specific, genotype-dependent differences in specific metabolites, such as higher fructose in introgression line 6-3, and also established positive and negative correlational networks between specific metabolites as well as between various classes of metabolites and agronomic traits. While many metabolic QTLs were identified, many of these were also coupled with harvest index-associated QTLs, casting some doubt on the practical utility of the detected QTLs (Schauer et al. 2006). It remains unclear to what extent negative linkage drag influenced the findings from this study. However, it is clear from this study that metabolite QTLs do exist in wild germplasm, metabolic networks can be constructed in tomato using wild germplasm sources, metabolic data can be collected in parallel to agronomic and horticultural data, key nodes of regulation can be identified on an omics scale, and these nodes may be exploited to alter metabolic traits in tomato.

A later study examined the inheritance of these primary metabolic traits in the same population (Schauer et al. 2008). Building on the results of the previous study, three years' worth of metabolite data were combined to identify metabolite QTLs conserved across years and to examine the heritability of the traits. In addition to the homozygous *S. pennellii* IL (introgression lines) population, the researchers also analyzed heterozygous ILs [ILHs; (Semel et al. 2006)]. From the 43 QTLs detected over the three years of experiment, a striking number of them were fairly heritable. Although some heritability estimates fluctuated widely over the three years of this study, this is to be expected when dealing with traits that are subject to environmental conditions (Schauer et al. 2008). In addition, it was reported that most metabolite QTLs were dominant, which is a promising result from a breeding standpoint, since most tomato breeders are attempting to develop hybrid cultivars. The researchers also compared metabolic networks between the IL and ILH populations, and found that a significant amount of the previously detected associations between metabolite QTLs and harvest index were abolished in the heterozygotes (Schauer et al. 2008). This finding is an important discovery for breeders attempting to modulate these traits. Bermudez and colleagues (2008) took a candidate gene approach in order to further study QTLs identified by Schauer and colleagues (2006) and to associate sequence variation within candidate gene regions with the previously detected QTLs. The approach utilized available genomic resources to (1) identify potential candidate genes within genomic regions of interest, (2) compare allelic variation between the two original parental lines, and (3) correlate transcription of some candidate genes with metabolite abundance. Although Bermudez and colleagues (2008) were limited by the lack of a complete genome sequence, this did not prevent the identification of many promising candidate alleles underlying previously detected QTLs; in addition, the approach simultaneously produced gene-specific markers that could be used in practical application.

There is still a long way to go in the effort to combine metabolite profiling, genomics, and practical breeding for tomato fruit quality. However, the use of immortalized populations

(ILs), common phenotyping platforms, transparent bioinformatic methods, and relevant environmental conditions strengthened and legitimized the findings of Schauer and colleagues (2006 and 2008). It is hoped that future studies adhere to these examples. The resource limitations of the Bermudez and colleagues (2008) study are no longer relevant, now that a genome sequence and a large set of genome-wide markers amenable to high-throughput assays are available. These studies set the stage for the next generation of genomics-assisted breeding strategies in tomato.

Carotenoids

Carotenoids are C_{40} terpenoid compounds found abundantly in tomato fruit. Ripe tomato fruits contain significant amounts of lycopene and β-carotene (the precursor of vitamin A), and small amounts of phytoene, phytofluene, lutein, and zeaxanthin. The concentration of each of these compounds depends on the genotype, the maturity stage of the fruit, and the environmental conditions in which the plants are grown. The insolubility of carotenoids in aqueous solutions creates challenges for high-throughput quantification of these compounds, and therefore carotenoid data was not reported in the Schauer and colleagues (2006 and 2008) studies. This class of secondary metabolites is ubiquitous throughout nature; the presence of conjugated double bonds in a long polyene chain renders carotenoids indispensable for several reasons, including quenching ROS generated by electron transport mechanisms and photosynthetic reactions, harvesting photons, and, due to their distinct spectral qualities, attracting pollinators and seed dispersal agents (Hirschberg 2001; Bramley 2002). Throughout evolutionary time, carotenoids have been recruited for a variety of extremely important cellular purposes, including light harvesting, dissipation of excess solar energy, scavenging of ROS, strengthening sight (in animals), and the biosynthesis of hormones (Vershinin 1999).

Since animals cannot synthesize these compounds de novo, carotenoid levels in animals are directly related to the amount ingested through the diet (Demmig-Adams and Adams 2002). One such compound, lycopene, has been co-opted by some plants to attract seed dispersal agents through its bright red appearance. Lycopene is responsible for the deep red hue of ripe tomato fruit and is the most abundant carotenoid in this organ, and since deep red fruit color is a conventional tomato breeding objective, elevating lycopene content is a parallel objective to fruit color. The biosynthesis, and to some extent regulation, of carotenoid accumulation is fairly well understood, although some uncertainty remains regarding the exact mechanism that funnels metabolites into the carotenoid pathway. These topics have been extensively reviewed and discussed (Hirschberg 2001; Fraser and Bramley 2004; Fraser et al. 2008).

As follows from current knowledge (and speculation) surrounding carotenoids and human health, increasing levels or altering types of carotenoids in tomato is of importance from a nutritional standpoint. β-carotene is the only carotenoid defined as a "nutrient" by the USDA, yet a large proportion of the relevant literature has focused on modulating lycopene content in tomato, as lycopene is the most predominant carotenoid in the red tomato fruit (Foolad 2007a). Also, elevated fruit carotenoid content in tomato can be considered a value-added trait, potentially resulting in increased consumer desire for such varieties or food products, although the effect of increased carotenoids on tomato product sales has yet to be definitively quantified. A recent sensory study concluded that consumers associate increased tomato fruit quality and better overall taste with red color; however, under conditions in which the fruit color was masked, the association was abolished (Stommel et al. 2005). Regardless of whether or not lycopene actually provides a measurable health benefit, when faced with the choice between a deeply red tomato and a lightly red or yellow tomato, consumers seem to prefer the former. Stommel and colleagues (2005) claim this preference results from a visual association between intense red fruit color and better

overall quality and taste. In addition, for processing tomatoes, the raw tomatoes used for red paste must meet a certain grade of redness in order to be accepted by processing plants. Following this line of thought, a better understanding of the genetic factors affecting lycopene accumulation in tomato will allow for more successful traditional breeding and metabolic engineering of this important crop, with the eventual goals of increased economic benefit for breeders and growers (by elevating demand, not necessarily price) as well as potentially increased nutritional benefit to consumers.

Due to the nascent status of omics-based research to identify novel carotenoid regulatory targets in tomato and the relative ease of selecting tomatoes of the desired color, commercial breeders continue to employ a series of mutations identified in the mid-20th century in order to modulate carotenoid type and accumulation level. Several naturally occurring tomato mutations in structural enzymes of the carotenoid biosynthetic pathway have been observed, reported in the literature, and are actively employed for breeding purposes. Of these, *r* (null mutation in *PSY*), β (*LCY-B* not down-regulated in response to ripening), *Delta* (lycopene-*δ*-cyclase not down-regulated in response to ripening), *tangerine* (*t*; mutation in *CRTISO*) and *old-gold* (*og*) or *crimson* (og^c; mutation in *LCY-B*), are most significant, owing to their drastic fruit phenotypes (Ronen et al. 1999; Ronen et al. 2000; Liu et al. 2003). Among these, *og*/og^c is the only mutant that results in increased lycopene accumulation, albeit at the expense of β-carotene. The og^c mutation is widely used in commercial varieties to increase fruit color intensity (Faria et al. 2003). Commercial varieties harboring some of the other mutations have been developed and used as ingredients in an expanding array of processed food products or sold for fresh market consumption. Reports of QTLs that modulate carotenoid content have been published and reviewed by Foolad (2007a).

While much is known about the synthesis of carotenoids in plants, the "next frontier" is the elucidation of how carotenoid accumulation is regulated in tomato fruit and the harnessing of such knowledge for breeding purposes (Fraser and Bramley 2004). Unlike most plants, tomato hyper-accumulates lycopene during fruit ripening by converting chloroplasts into chromoplasts, and has evolved a specialized system to specifically produce this compound in comparatively large amounts. Intuitively, this would require a constant supply of metabolic precursors, a consistently high amount of active biosynthetic machinery, and a tightly controlled metabolic flux specifically favoring carotenoid biosynthesis over other competing pathways. Thus, there are two main questions that have yet to be answered in this regard: What is the specific regulatory mechanism utilized by tomato to convert fruit from a photosynthetic tissue into a highly specialized carotenoid factory? How does this mechanism differ between low-lycopene *S. lycopersicum* varieties and high-lycopene wild tomato accessions, such as some *S. pimpinellifolium* accessions, assuming carotenoid biosynthetic genes are functioning normally in both cases?

It is generally hypothesized that lycopene accumulation in ripe tomato fruit is influenced by the action of many regulatory genes (the specific natures of most are unknown), and in most segregating populations, variation in lycopene content is continuous, which indicates quantitative inheritance of the trait. Also of note, it has been observed that environmental conditions, as well as fruit size (which has been negatively correlated with lycopene content), highly influence the final concentration of lycopene in the fruit. So, intuitively, one would conclude that lycopene accumulation in tomato fruit must depend not only on the action of biosynthetic genes, which have been the major focus of basic researchers until recently, but also on as-of-yet unidentified regulatory interactions (which may be involved in metabolic flux, transcriptional and post-transcriptional regulation, response to light, or a combination of these), and indirect mechanisms present in the plant and the environment

that influence ripening, nutrient acquisition, and photosynthetic productivity.

Studies focusing on the application of contemporary genomics information to tomato breeding for the purpose of modulating carotenoid levels have been conducted. A TILLING approach was employed by Gady and colleagues (2012) in order to identify and describe two different mutations in the *PSY1* gene that exhibited significant phenotypic effects. Using a mutagenized population described in Gady and colleagues (2009), the researchers identified M2 individuals with point mutations in the *PSY1* sequence and produced tomato lines with null and weak alleles of *PSY1* that resulted in altered carotenoid accumulation (Gady et al. 2012). This study is an example of the utility of TILLING for applied breeding objectives. Now that rapid and cost-effective identification of mutants is possible, it is expected that TILLING approaches will be employed by pre-breeders to generate novel genetic variation in more genes that are involved in fruit color, carotenoid accumulation, sugar accumulation and content, ripening, and more.

Concurrently, forward genetics approaches are still routinely employed to mine existing genetic variation in closely related wild species of tomato. One recent multi-year QTL study identified two QTLs, stably inherited from a *S. pimpinellifolium* accession, that significantly increased fruit lycopene content in a recombinant inbred line (RIL) population (Ashrafi et al. 2009, 2012). This research was followed up with a marker-assisted backcross (MABC) experiment to simultaneously verify the QTL effects, fine map the QTL regions, and develop near-isogenic lines useful for breeding purposes (Kinkade 2010; Kinkade and Foolad, unpubl. results). In this case, a large number of molecular markers were required to execute the MABC breeding and fine mapping experiments effectively; therefore, new simple sequence repeat (SSR) markers were developed utilizing available tomato genomic BAC end sequences physically located near the QTL regions. Field experiments were conducted using BC_2 and BC_2F_2 populations segregating for different QTL interval sizes, and fruit lycopene content was measured by high performance liquid chromatography (HPLC). As a result, the phenotypic effect of one QTL (*lyc12.1*) was verified as stable in the genetic background analyzed (Kinkade 2010; Kinkade and Foolad, unpubl. data). In addition, *lyc12.1* was delimited to a ~0.3 cM region of chromosome 12, near-isogenic lines (NILs) with *lyc12.1* in a relevant cultivated tomato genetic background were developed, and tightly-linked co-dominant markers for tracking the QTL were produced (Kinkade 2010; Kinkade and Foolad, unpubl. data). The results of this research were made possible in part by the availability of tomato genomics resources, which were successfully employed to translate findings from a QTL study into a validated breeding tool for increasing lycopene content. Further, this series of studies indicates that current tomato genomics tools are useful enough to develop tiny introgressions containing useful alleles from wild accessions and transfer them to cultivated genetic backgrounds with minimal linkage drag. However, more high-throughput techniques are required to develop correlational networks and identify MAS targets for carotenoid regulation from large populations in an efficient manner.

Omic-scale identification of carotenoid regulatory mechanisms in segregating populations has just begun to be conducted, and some of the necessary tools have been developed for the tomato system. Fraser and colleagues (2007b) demonstrated that high-throughput metabolic profiling techniques based on MALDI/TOF-MS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) could be employed to evaluate tomato populations segregating for carotenoid accumulation level and carotenoid profile using crude extracts. Because quantifying the various species of carotenoids in large populations is laborious and expensive, this has tremendous utility for genomics-assisted breeding for increased carotenoid content. Although it has not been definitively

demonstrated that MALDI/TOF-MS can completely replace conventional HPLC (largely due to the fact that lycopene and β-carotene have identical m/z values), this technique offers rapid identification of high- or low-accumulating genotypes and also has the ability to detect other carotenoids in tandem.

Future Directions

It is hoped that subsequent studies adopt the systems biology approach of Schauer and colleagues (Schauer et al. 2005, 2006, 2007) in conjunction with the carotenoid profiling techniques pioneered by Fraser and colleagues (Fraser and Bramley 2004; Fraser et al. 2007a,b, 2008), and utilize relevant genetic backgrounds if there is truly a desire to translate forward genetics findings into breeding practices. The systems approach takes advantage of genetic variation to produce correlational networks and infers key regulators of important traits from these networks. Research teams can then employ contemporary high-throughput techniques to identify (or generate) genetic variation within these regulatory genes, in order to produce novel metabolic traits that can be tested under relevant agricultural conditions. It is not out of the realm of consideration that continued tomato omics studies can, by exploiting the well-developed genomics resources for tomato (Mueller et al. 2005) and the continuously-maturing techniques for integrating various types of omics data, fuel the development of tomato omics resources similar to those developed for yeast (Myers et al. 2005) or other biological systems.

The exploration, characterization and incorporation of abundantly available natural genetic variation, or generation of variation using mutagenesis in agriculturally relevant genetic backgrounds, still represents an opportunity for practical breeders to identify and incorporate novel alleles for fruit quality traits into the cultigen. As described above, genomic and metabolomic experimental tools have been developed using tomato as a model system to mine this variation. With the advent of genotyping by sequencing and SNP arrays, the inability to obtain extremely large sets of informative polymorphic markers within the tomato cultigen is no longer an issue. There is no reason why a similar overall approach to that of Davuluri and colleagues (2005) cannot be executed for carotenoid regulatory targets identified by correlational networks while using publicly acceptable sources of genetic variation, rather than genetic engineering. Early applied TILLING studies have produced tomato materials with altered carotenoid content (Gaby et al. 2012) and potyvirus resistance (Piron et al. 2010). Other examples include the development of melon materials with improved shelf life characteristics (Dahmani-Mardas et al. 2010). This effort will require a well-coordinated collaborative effort among bioinformaticians, chemists, molecular biologists, and breeders, and alignment on a practical, concrete breeding objective to pursue with such an approach.

In general, tomato fruit quality is considered by practical tomato breeders to be a lower priority than yield, disease resistance, and other production-related traits. There are extremely valid and unavoidable reasons for this. However, consumers are increasingly dissatisfied with the quality of fresh tomatoes available to them (Kolata 2012). With the technologies available to applied research and development entities (public and private), it is now feasible to generate high-yielding, durable tomato varieties with superior fruit quality if the desire to do so exists. In the case of altering tomato carotenoid accumulation, we speculate that the main reason this has not been widely pursued is that consumers are unwilling to pay more for increased antioxidant content, and thus this trait has not been a priority for breeders, growers, or processors. As a result, currently, private research and development firms are not willing to invest heavily in this area, nor are seed companies willing to license such traits for hefty sums from the public sector, though this may not be the case in the future. On the other hand, primary metabolic traits can be related to yield and

compose the final fruit quality attributes that are prioritized by the customers of seed companies. Association of individual metabolite levels or metabolite profiles with critical fruit quality metrics, such as soluble solids and viscosity, concurrent with field testing in relevant genetic backgrounds, will hasten the incorporation of genomics-assisted findings into tomato breeding practices.

References

Ashrafi, H., Kinkade, M.P., and Foolad, M.R. 2009. A new genetic linkage map of tomato based on a *Solanum lycopersicum* × *S. pimpinellifolium* RIL population displaying locations of candidate resistance ESTs. *Genome* 52:935–956.

Ashrafi, H., Kinkade, M., Merk, H., and Foolad, M.R. 2012. Identification of novel QTLs for lycopene content and other fruit quality characteristics in a Solanum lycopersicum x S. pimpinellifolium RIL population. *Mol Breeding* published online; DOI 10.1007/s11032-011-9643-1.

Barone, A., Di Matteo, A., Carputo, D., and Frusciante, L. 2009. High-throughput genomics enhances tomato breeding efficiency. *Current Genomics* 10:1–9.

Bermudez, L., Urias, U., Milstein, D., Kamenetzky, L., Asis, R., Fernie, A.R., Van Sluys, M.A., Carrari, F., and Rossi, M. 2008. A candidate gene survey of quantitative trait loci affecting chemical composition in tomato fruit. *Journal of Experimental Botany* 59:2875–2890.

Bovy, A., Schijlen, E., and Hall, R. 2007. Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): The potential for metabolomics. *Metabolomics* 3: 399–412.

Bramley, P.M. 2002. Regulation of carotenoid formation during tomato fruit ripening and development. *Journal of Experimental Botany* 53:2107–2113.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., and Yandell, M. 2007. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18:188–196.

Causse, M., Duffe, P., Gomez, M.C., Buret, M., Damidaux, R., Zamir, D., Gur, A., Chevalier, C., Lemaire-Chamley, M., and Rothan, C. 2004. A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. *Journal of Experimental Botany* 55:1671–1685.

Chaïb, J., Lecomte, L., Buret, M., and Causse, M. 2006. Stability over genetic backgrounds, generations, and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. *Theoretical and Applied Genetics* 112:934–944.

Dahmani-Mardas, F., Troadec, C., Boualem, A., Leveque, S., Alsadon, A.A., Aldoss, A.A., Dogimont, C., and Bendahmane, A. 2010. Engineering melon plants with improved shelf life using the TILLING approach. *PLoS ONE* 5:e15776.

Davuluri, G.R., van Tuinen, A., Fraser, P.D., Manfredonia, A., Newman, R., Burgess, D., Brummell, D.A., King, S.R., Palys, J., Uhlig, J., et al. 2005. Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nat Biotech* 23:890–895.

Demmig-Adams, B., and Adams, W.W. 2002. Antioxidants in Photosynthesis and Human Nutrition. *Science* 298:2149–2153.

Eshed, Y., and Zamir, D. 1995. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141:1147–1162.

FAOSTAT. 2012. Food and Agriculture Organization of the United Nations. http://faostat.fao.org.

Faria, M.V., Maluf, W.R., Azevedo, S.M.d., Andrade Junior, V.C.d., Gomes, L.A.A., Moretto, P., and Licursi, V. 2003. Yield and post-harvest quality of tomato hybrids heterozygous at the locialcobaça, old gold-crimson or high pigment. *Genetics and Molecular Research* 2: 317–327.

Fei, Z., Joung, J.-G., Tang, X., Zheng, Y., Huang, M., Lee, J.M., McQuinn, R., Tieman, D.M., Alba, R., Klee, H.J., et al. 2011. Tomato Functional Genomics Database: A comprehensive resource and analysis package for tomato functional genomics. *Nucleic acids research* 39:D1156–D1163.

Foolad, M.R. 2007a. Genome mapping and molecular breeding of tomato. *Int'l J Plant Genomics* Vol. 2007, Article ID 64358, 52 pages. doi:10.1155/2007/64358:1-52.

Foolad, M.R. 2007b. Molecular mapping, marker-assisted selection and map-based cloning in tomato. In: R.K. Varshney and R. Tuberosa (eds.) *Genome Assisted Crop Improvement. Vol. 2, Genome Applications in Crops*. Dordrecht, The Netherlands: Springer, pp. 307–356.

Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J.-P., Meller, J., Elber, R., Alpert, K.B., et al. 2000. *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science* 289:85–88.

Fraser, P.D., and Bramley, P.M. 2004. The biosynthesis and nutritional uses of carotenoids. *Progress in Lipid Research* 43:228–265.

Fraser, P.D., Enfissi, E., and Bramley, P.M. 2008. Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. *Arch Biochem and Biophys* 483: 196-204.

Fraser, P.D., Enfissi, E.M., Goodfellow, M., Eguchi, T., and Bramley, P.M. 2007a. Metabolite profiling of plant carotenoids using the matrixassisted laser desorption ionization time-of-flight mass spectrometry. *Plant J.* 49:552–564.

Fraser, P.D., Enfissi, E.M.A., Halket, J.M., Truesdale, M.R., Yu, D., Gerrish, C., and Bramley, P.M. 2007b. Manipulation of phytoene levels in tomato fruit: Effects on isoprenoids, plastids, and intermediary metabolism. *Plant Cell* 19:3194–3211.

Fridman, E., Pleban, T., and Zamir, D. 2000. A recombination hotspot delimits a wild-species quantitative trati locus for tomato sugar content to 484 bp within an invertase gene. *Proc. Natl. Acad. Sci. U.S.A.* 97:4718–4723.

Gady, A.L.F., Hermans, F.W.K., Van de Wal, M.H.B.J, van Loo, E.N., Visser, R.G.F., and Bachem, C.W.B. 2009. Implementation of two high through-put techniques in a movel application: Detecting point mutations in large EMS mutated plant populations. *Plant Methods* 5:13.

Gady, A.L.F., Vriezen, W.H., Van de Wal, M.H.B.J., Huang, P., Bovy, A.G., Visser, R.G.F., and Bachem, C.W.B. 2012. Induced point mutations in the phytoene synthase 1 gene cause difference in carotenoid contant during tomato fruit ripening. *Mol Breeding* 29:801–812.

Hamilton, J.P., Hansey, C.N., Whitty, B.R., Stoffel, K., Massa, A.N., Van Deynze, A., De Jong, W.S., Douches, D.S., and Buell, C.R. 2011. Single Nucleotide Polymorphism discovery in elite North American potato germplasm. *BMC Genomics* 12:302.

Hirschberg, J. 2001. Carotenoid biosynthesis in flowering plants. *Current Opinion in Plant Biology* 4:210–218.

Kalloo, G. 1991. *Genetic improvement of tomato*. Berlin and Heidelberg: Springer-Verlag.

Kavanaugh, C.J., Trumbo, P.R., and Ellwood, K.C. 2007. The U.S. Food and Drug Administration's evidence-based review for qualified health claims: Tomatoes, lycopene, and cancer. *J. National Cancer Institute* 99:1074–1085.

Kinkade, M.P. 2010. *Identification and isolation of genetic elements conferring increased tomato fruit lycopene content.* PhD diss., Department of Horticulture, The Pennsylvania State University, University Park, 136 pages.

Kolata, G. 2012. Flavor is price of scarlet hue of tomatoes, study finds. *The New York Times*, published June 28, 2012.

Labate, J.A., Grandillo, S., Fulton, T., Muños, S., Caicedo, A.L., Iris Peralta6, Ji, Y., Chetelat, R.T., Scott, J.W., Gonzalo, M.J., et al. 2007. Tomato. In: C. Kole (ed.) *Genome mapping and molecular breeding in plants*. Berlin: Springer-Verlag, pp. 1–125.

Lecomte, L., Duffé, P., Buret, M., Servin, B., Hospital, F., and Causse, M. 2004. Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. *Theor Appl Genet* 109:658–668.

Lewis, S., Searle, S., Harris, N., Gibson, M., Iyer, V., Richter, J., Wiel, C., Bayraktaroglu, L., Birney, E., Crosby, M., et al. 2002. Apollo: A sequence annotation editor. *Genome Biol.* 33 (12):research0082.0081-0082.0014.

Lippman, Z., Semel, Y., and Zamir, D. 2007. An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Current Opinion in Genetics and Development* 17:545–552.

Liu, Y.-S., Gur, A., Ronen, G., Causse, M., Damidaux, R., Duffe, P., Buret, M., Hirschberg, J., and Zamir, D. 2003. There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotech. J.* 1:195–207.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D., and Tanksley, S.D. 1993a. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262:1432–1436.

Martin, G.B., Vicente, M.C.d., and Tanksley, S.D. 1993b. High-resolution linkage analysis and physical chraracterization of the *Pto* bacterial resistance locus in tomato. *Mol. Plant-Micr. Interact.* 6:26–34.

Menda, N., Semel, Y., Peled, D., Eshed, Y., and Zamir, D. 2004. In silico screening of a saturated mtuation library of tomato. *Plant J.* 38:861–872.

Minoia, S., Petrozza, A., D'Onofrio, O., Piron, F., Mosca, G., Sozio, G., Cellini, F., Bendahmane, A., and Carriero F. 2010. A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC Research Notes* 3:69.

Mueller, L.A., Lankhorst, R.K., Tanksley, S.D., Giovannoni, J.J., White, R., Vrebalov, J., Fei, Z., van Eck, J., Buels, R., Mills, A.A., et al. 2009. A snapshot of the emerging tomato genome sequence. *The Plant Genome* 2:78–92.

Mueller, L.A., Solow, T.H., Taylor, N., Skwarecki, B., Buels, R., Binns, J., Lin, C.-W., Wright, M.H., Ahrens, R., Wang, Y., et al. 2005. The SOL genomics network: A comparative resource for Solanaceae biology and beyond. *Plant Physiol.* 138:1310–1317.

Myers, C., Robson, D., Wible, A., Hibbs, M., Chiriac, C., Theesfeld, C., Dolinski, K., and Troyanskaya, O. 2005. Discovery of biological networks from diverse functional genomic data. *Genome Biology* 6:R114.

Okabe, Y., Asamizu, E., Saito, T., Matsukura, C. Ariizumi, T., Bres, C., Rothan, C., Mizoguchi, T., and Ezura, H. 2011. Tomato TILLING technology: Development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries. *Plant Cell Physiol.* 52:1994–2005.

Overy, S.A., Walker, H.J., Malone, S., Howard, T.P., Baxter, C.J., Sweetlove, L.J., Hill, S.A., and Quick, W.P. 2005. Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. *Journal of Experimental Botany* 56:287–296.

Peralta, I.E., and Spooner, D.M. 2005. History, origin and early cultivation of tomato (Solanaceae). In: M.K. Razdan and A.K. Mattoo (eds.) *Genetic improvement of solanaceous crops: Tomato*. Enfield, NH: Science Publishers, pp. 1–24.

Piron, F., Nicolai, M., Minoia, S., Piednoir, E., Moretti, A., Salgues, A., Zamir, D, Caranta, C. Bendahmane, A. 2010. An induced mutation in tomato eIF4E leads to immunity to two potyviruses. *PLoS ONE* 5:e11313.

Rao, A.V. 2004. Processed tomato products as a source of dietary lycopene: Bioavailability and antioxidant properties. *Canadian Journal of Dietetic Practice and Research* 65:161–165.

Rao, A.V., and Rao, L.G. 2007. Carotenoids and human health. *Pharmacological Research* 55:207–216.

Rick, C.M. 1976a. Natural variability in wild species of *Lycopersicon* and its bearing on tomato breeding. *Genet. Agraria* 30:249–259.

Rick, C.M. 1976b. Tomato, *Lycopersicon esculentum* (Solanaceae). In: N.W. Simmonds (ed.), *Evolution of crop plants*. London: Longman, pp. 268–273.

Rick, C.M. 1978. The tomato. *Sci. Amer.* 23:76–87.

Rick, C.M. 1979. Potential improvement of tomatoes by controlled introgression of genes from wild speies. In: *Proc. Conf. Broadening Genetic Base of Crops*. Pudoc: Wageningen, pp. 167–173.

Rick, C.M., and Fobes, J.F. 1974. Association of an allozyme with nematode resistance. *Rpt. Tomato Genetic Coop.* 24:25.

Rick, C.M., and Fobes, J.F. 1975. Allozyme variation in the cultivated tomato and closely related species. *Bul. Torrey Bot. Club* 102:376–384.

Rigola, D., van Oeveren, J., Janssen, A., Bonne, A., Schneiders, H., van der Poel, H.J.A., van Oursouw, N.J., Hogers, R.C.J., de Both, M.T.J., and van Eijk, M.J.T. 2009. High-throughput detection of induced mutations and natural variation using KeyPoint technology. *PLoS ONE* 4:e4761.

Ronen, G., Carmel-Goren, L., Zamir, D., and Hirschberg, J. 2000. An alternative pathway to beta-carotene formation in plant chloroplast discovered by map-based cloning of Beta and old-gold color muation in tomato. *Proc Natl Acad Sci USA* 97:11102–11107.

Ronen, G., Cohen, M., Zamir, D., and Hirschberg, J. 1999. Regulation of carotenoid biosyntheis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down regulated during ripening and is elevated in the mutant *Delta*. *Plant J.* 17:341–351.

Rousseaux, M.C., Jones, C.M., Adams, D., Chetelat, R., Bennett, A., and Powell, A. 2005. QTL analysis of fruit antioxidants in tomato using Lycopersicon pennellii introgression lines. *Theoretical and Applied Genetics* 111:1396–1408.

Schaffer, A.A., Levin, I., Oguz, I., Petreikov, M., Cincarevsky, F., Yeselson, Y., Shen, S., Gilboa, N., and Bar, M. 2000. ADPglucose pyrophosphorylase activity and starch accumulation in immature tomato fruit: The effect of a Lycopersicon hirsutum-derived introgression encoding for the large subunit. *Plant SciencE* 152: 135–144.

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Melis, A., Bruedigam, C., Kopka, J., et al. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotech* 24:447-454.

Schauer, N., Semelb, Y., Balboa, I., Steinfatha, M., Repsilbera, D., Selbiga, J., Plebanb, T., Zamir, D., and Fernie, A.R. 2008. Mode of inheritance of primary metabolic traits in tomato. *The Plant Cell* 20:509–523

Schauer, N., Zamir, D., and Fernie, A.R. 2005. Metabolic profiling of leaves and fruit of wild species tomato: A survey of the *Solanum lycopersicum* complex. *J. Exp. Bot.* 56:297–307.

Semel, Y., Nissenbaum, J., Menda, N., Zinder, M., Krieger, U., Issman, N., Pleban, T., Lippman, Z., Gur, A., and Zamir, D. 2006. Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. of Sci. U.S.A.* 103:12981–12986.

Sim, S., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M., Van Deynze, A., Hamilton, J., Buell, C.R., Causse, M., Wijeratne, S., and Francis, D.M. 2012. Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE* 7: e40563.

Stommel, J.R., Abbott, J., Saftner, R.A., and Camp, M. 2005. Sensory and objective quality attributes of beta-carotene- and lycopene-rich tomato fruit. *J. Amer. Soc. Hort. Sci.* 130:244–251.

Tanksley, S.D., Medina-Filho, H., and Rick, C.M. 1982. Use of naturally-occuring enzyme variation to detect and map genes controlling quantitative traits in an interspecific cross of tomato. *Heredity* 49:11–25.

Tanksley, S.D., and Orton, T.J. 1983. *Isozymes in plant genetics and breeding*. Amsterdam: Elsevier.

Taylor, J.B. 1986. Biosystematics of the tomato. In: J.G. Atherton, and J. Rudich (eds.) *The tomato crop: A scientific basis for crop improvement*. New York: Chapman & Hall, pp. 1–34.

Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400):635–641.

Tuberosa, R., and Salvi, S. 2006. Genomics approaches to improve drought tolerance in crops. *Trends in Plant Science* 11:415–412.

Tuberosa, R., Salvi, S., Giuliani, S., Sanguineti, M.C., Bellotti, M., Conti, S., and Landi, P. 2007. Genome-wide approaches to investigate and improve maize response to drought. *Crop Science* 47:1–23.

USDA. 2012. National Nutrient Database for Standard Reference. http://ndb.nal.usda.gov.

Vershinin, A. 1999. Biological functions of carotenoids: Diversity and evolution. *Biofactors* 10:99–104.

Von Löw, E.C., Perabo, F.G., Siener, R., and Muller, S.C. 2007. Facts and fiction of phytotherapy for prostate cancer: A critical assessment of preclinical and clinical data. *In Vivo* 21:189–204.

Wikström, N., Savolainen, V., and Chase, M.W. 2001. Evolution of the angiosperms: Calibrating the family tree. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* 268:2211-2220.

Wu, K., Erdman, J.W., Schwartz, S.J., Platz, E.A., Leitzmann, M., Clinton, S.K., DeGroff, V., Willett, W.C., and Giovannucci, E. 2004. Plasma and dietary carotenoids, and the risk of prostate cancer. *Cancer Epidemiology Biomarkers & Prevention* 13:260–269.

Yates, H.E., Frary, A., Doganlar, S., Frampton, A.J., Eannetta, N.T., Uhlig, J., and Tanksley, S.D. 2004. Comparative fine mapping of fruit quality QTLs on chromosome 4 introgression derived from two wild tomato species. *Euphytica* 135:283–296.

Chapter 13

Improvement of Yield per se in Sugarcane

M. Gouy, S. Nibouche, J.Y. Hoarau, and L. Costet

Abstract

Sugarcane is the main source of sucrose in the world and has also become one of the main sources of renewable energy, thanks to ethanol and electricity production. The main objectives of breeding programs are improving cane biomass and qualities. The genetics of current sugarcane cultivars (*Saccharum* spp.) are extremely complex, owing to a high polyploid genome of 10 to 12 homeologous chromosome sets, resulting from the interspecific origin of two polyploid ancestral genomes. A century of breeding efforts based on elite domesticated progenitors improved by wild introgressions has helped build one of the most productive biomass crops. However, in recent decades the percentage increase in annual sugarcane yield has been lower than in other crops. Until now, cultivar improvement has relied on traditional breeding programs that require 12 to 15 years of expensive field evaluations. This article discusses the most recent advances in genomics applications to support traditional breeding. In the past two decades, efforts have been made to construct genetic maps, study quantitative trait loci (QTL), and start association mapping studies. Today, genomic selection (GS) approaches to select individuals for advancement in the breeding process is believed to be appropriate for complex traits such as yield, thanks to improved estimates of marker effects combined with a better grasp of small-effect QTLs. GS is particularly attractive in the highly polyploid context of sugarcane, where the nature of yield genetic determinism is presumably highly quantitative. Frequent genotype x environment interactions add to the challenges associated with QTL detection. The focus in this article is on research areas that are likely to result in concrete improvements through advancing the incorporation of association genetics-based approaches. A strategy to design a breeder-friendly marker system is also presented. Plant growth modeling could provide sounder ecophysiological parameters for describing the complex biological process underlying yield and sucrose elaboration. Such models could overcome traditional problems caused by G x E interactions and consequently improve both QTL detection and GS approaches.

Introduction

Sugarcane is a giant perennial grass two to four meters in height that is propagated clonally using stem cuttings. Sugarcane is cultivated with one plant crop and several ratoon crops, each about 12 months in length after an annual

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

harvest. Sugarcane combines a high polyploidy, a heterozygous genome, and C4 photosynthetic metabolism, making it one of the most efficient converters of light into chemical energy. Sugarcane is an important cash crop in tropical and subtropical areas and is grown in about 100 countries (FAOSTAT 2012). It is mainly grown for the production of sucrose, which accumulates at high concentrations in the stem, and the very high productivity of this crop makes it one of the most economically attractive crops for the production of energy from biomass (Waclawovsky et al. 2010). Apart from sucrose, many by- or co-products (bagasse, molasses, bioethanol, mud) used for a wide range of industrial applications (livestock feed, energy or paper production, bio-fertilization, bio-refinery, particleboard, etc.) can also be obtained from sugarcane. In 2010, more than 23 million hectares of sugarcane were harvested worldwide, ranking only 12th in crop harvested area, far behind wheat (217 million hectares), maize (161 million hectares), and rice (153 million hectares). However sugarcane was by far the most harvested and processed product in the world, with 1.7 billion tons of cane harvested and more than 133.6 million tons of sucrose and 54 million tons of molasses produced (FAOSTAT 2012). Sugarcane is the main source of sucrose worldwide, representing more than 75% of world production. After maize, it is the second largest source of sugar-based bioethanol (first generation bioethanol), with more than 26 billion liters produced. Bioethanol is mainly produced in Brazil where more than half the area under cane is dedicated to it (Dal-Bianco et al. 2011). In Brazil, sugar-based ethanol and bagasse-based electricity produced from sugarcane together represent 17% of national energy consumption, which makes sugarcane one of the main sources of renewable energy and one of the main alternatives to fossil fuel (Brazilian Ministério de Minas e Energia 2011; Matsuoka et al. 2009). Leaving aside the controversy about the social and environmental costs of converting land to grow biofuel crops (De Araujo and Moura 2011; Duailibe 2010; Walker 2009), sugarcane is arguably one of the best crops for biofuel production in terms of net energy value (NEV), when considering all resource inputs (Garoma et al. 2011). Moreover, NEV is expected to increase with the development of cellulosic ethanol (second-generation bioethanol) (Dias et al. 2011a; Dias et al. 2011b), since the lignocellulosic fraction of sugarcane is estimated to represent 50% of its energy potential (Botha 2009; Manners 2011). In favorable agro-climatic zones, sugarcane fresh biomass frequently exceeds 100 tons of cane/ha/year using ordinary fertilization input levels. This outstanding biomass production, combined with the development of cellulosic ethanol and genetic transformation technologies, means sugarcane will play a major role among industrial plants in meeting the demand for sugar, renewable energy, and bio-products (Figure 13.1) (Arruda 2011).

Currently increased yield is the main objective of sugarcane genetic improvement. Until now, sucrose yield per unit area was the main goal of sugarcane breeders since this parameter is the most closely correlated with both farmers' and industrial incomes (Jackson 2005). However breeding either for both sucrose and biomass or for only biomass, as already suggested by Alexander (1985), is clearly one of the new reasons for cultivating sugarcane, given the urgent need to reduce the consumption of fossil fuels. Depending on industrial applications, breeding objectives can be summarized in three sugarcane ideotypes: the traditional "sugar" ideotype bred for sugar (∼13% sugar, ∼12% fiber); the "energy cane type I," bred for both sugar (∼13%) and fiber (∼17%); and the "energy cane type II," bred only for fiber (∼5% sugar, ∼30% fiber) production (Tew and Cobill 2008). Sugarcane yield depends on two key components: the quantity of biomass and the quality of the dry matter (DM). The quantity of cane biomass depends on agro-morphological traits related to plant architecture (tillering, stalk height, and diameter), which are usually significantly genetically correlated

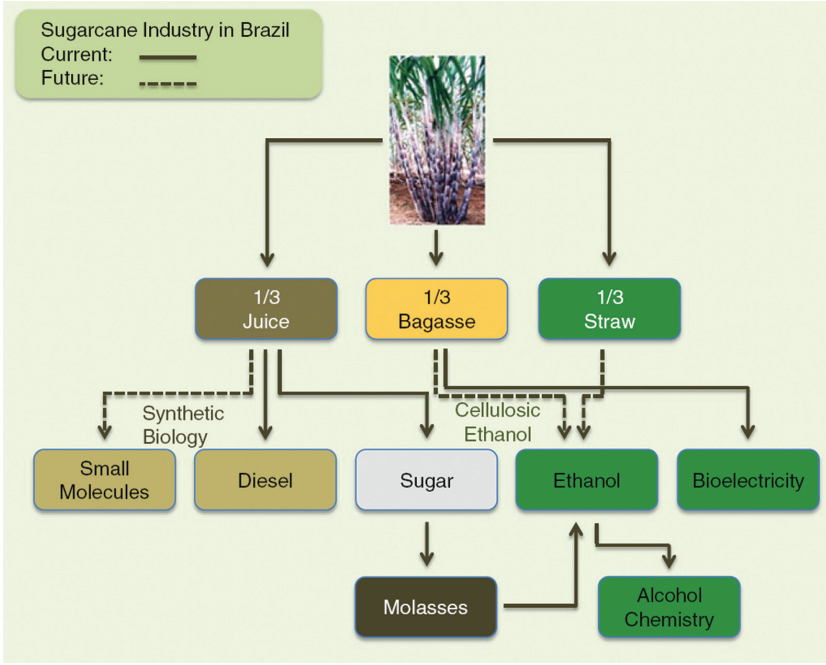


Fig. 13.1. Current and future prospects for the sugarcane industry in Brazil (according to Arruda 2011; with kind permission from Springer Science and Business Media). For a color version of this figure, please refer to the color plate.

(Gravois and Milligan 1992; Jackson 1994; Kang et al. 1989). Cane DM quality can be assessed based on its soluble and insoluble fractions. In the soluble fraction (cane juice), DM is estimated by measuring pol (sucrose content) and brix (total soluble DM) content. The brix:pol ratio determines the purity of industrial juice. Combined with fiber content (insoluble DM), this information allows for estimation of the percentage of sucrose content per fresh cane weight and finally the sugar yield per unit area.

Significant increases in yield have been obtained by classical breeding. The average worldwide yield of fresh cane increased by 41% over the last 50 years, from 50 tons per hectare in 1961 to 71 tons per hectare in 2010 (Figure 13.2) (FAOSTAT 2012). This increase is the combined result of genetic progress and improved agronomic practices (irrigation, fertilization, etc.). In terms of relative value (+41%), the gain in sugarcane yield over the same period was half that of sugar beet and silage maize and about a quarter of that of grain crops such as maize, rice, and wheat. However the gain in terms of absolute value (+ 0.43 tons/ha/year) reveals a parallel evolution between sugarcane and sugar beet or silage maize. In the sugar industry of Barbados (Simmonds 1979) and Louisiana (Edme et al. 2005), in recent decades, an average of 1% annual increase in yield can be attributed to genetic progress. Genetic gain for sugarcane yield is generally explained by an increase in biomass yield rather than an increase in sugar content (Jackson 2005; Kang et al. 1983), although some improvements in sucrose content may have been made in some areas such as Louisiana (Lingle et al. 2010; Lingle et al. 2009). Although significant regular increases in yield have been obtained through conventional breeding, sugarcane still appears to be far from achieving its theoretical agronomic potential (Waclawovsky et al. 2010).

The demand for sugarcane for the production of sucrose and ethanol along with all its valuable co-products will increase with the

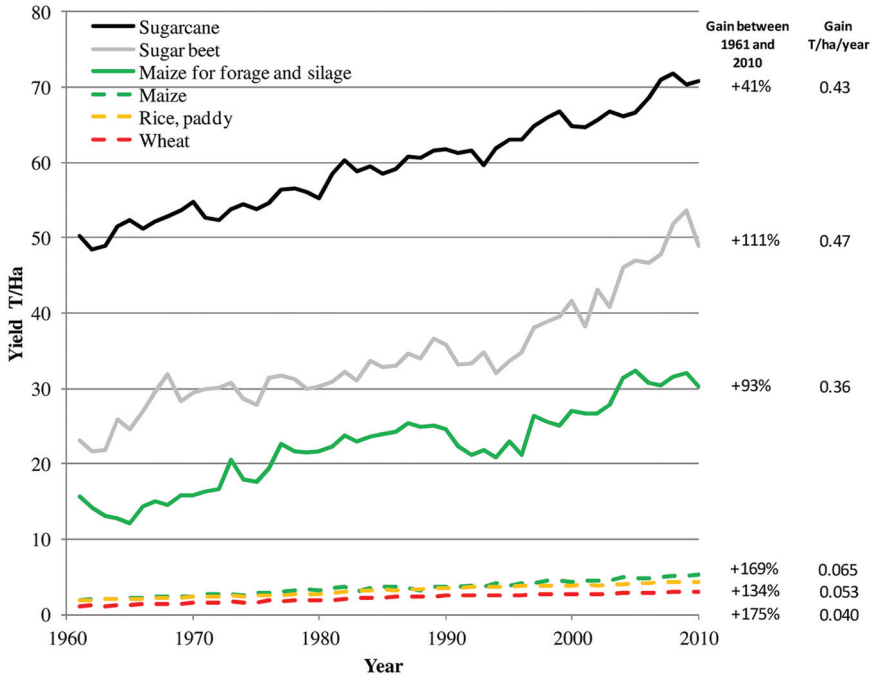


Fig. 13.2. Changes in sugarcane yield between 1961 and 2010. Comparison with other grass crops (FAOSTAT 2012). For a color version of this figure, please refer to the color plate.

growing world population (Anonymous 2012). These increasing demands will stimulate new scientific research investments to improve the efficiency of varietal development programs. Using conventional experimental approaches, breeding new sugarcane cultivars requires at least 10 to 15 years of selection. In the past two decades, much research has been dedicated to developing genomic tools that have improved our understanding of the genetic and genomic organization of sugarcane with the aim of facilitating varietal development. However, today the direct use of these tools in breeding remains challenging. Few applications derived from these genomic data have been implemented in sugarcane breeding programs and improved yields. The main reasons for these few applications could be the polygenic nature of yield control associated with the high complexity of the sugarcane genome.

In this chapter, after a review of the sugarcane evolution and breeding history that led to the present-day cultivar genomes (*Saccharum* spp.) and a summary of yield improvement, we discuss the genomic tools developed for sugarcane genome analyses, their potential applications, and the challenges involved in their use for yield improvements in "real life" breeding programs.

History of Sugarcane Yield Improvement

Evolution and Domestication

Sugarcane (*Saccharum* spp.) belongs to the Poaceae family and the Andropogoneae tribe, along with maize and sorghum. The former '*Saccharum* complex' concept encompasses the five closely related genera *Saccharum*, *Erianthus*, *Sclerostachya*, *Narenga*, and *Miscanthus,* which are characterized by cross-fertility (Daniels and Roach 1987). These genera are a potential source of diversity for sugarcane breeding. Species in the '*Saccharum* complex' are all highly

polyploid and frequently aneuploid. Domestication of sugarcane began several thousand years ago in Southeast Asia and the Pacific Islands, where inhabitants chewed its soft stem. Grivet and colleagues (2004; 2006) recently proposed a convincing scenario for sugarcane evolution and domestication in a review compiling historical, botanical, and molecular data, based on nuclear and mitochondrial probes as well as molecular cytogenetic evidence. Data generally do not support any evolutionary path through distant crosses involving representatives of different genera. These authors showed that the sugarcane lineage arose directly from the genus *Saccharum* independent of the closely related genera *Miscanthus* and *Erianthus*. The *Saccharum* genus is divided into six interfertile so-called 'species' (*S. spontaneum* L., *S. robustum* Brandes ex Jeswiet ex Grassl, *S. officinarum* L., *S. barberi* Jeswiet, *S. sinense* Roxburgh and *S. edule* Hasskarl.). Basically sugarcane emerged from an initial divergence between the two ancestral wild species *S. robustum* (2*n*=6*x*-8*x*=60-80) and *S. spontaneum* (2*n*=5*x*-16*x*=40-128) that have high fiber but very low sugar content. *S. spontaneum* comprises highly polymorphic phenotypes with many aneuploid forms and is considered to be autopolyploid. *S. officinarum* (2*n*=8*x*=80), with its juicy, sugar-rich, thick stalk called 'noble cane,' was domesticated by ancestral Melanesian populations in New Guinea from a particular cytotype of *S. robustum*. The two other sugar-producing domesticated 'species,' *S. barberi* (2n=104 to 128) and *S. sinense* (2n=81 to 124), formerly cultivated in Asia, have been shown to be natural hybrids between *S. officinarum* and *S. spontaneum* (D'Hont et al. 2002). Sugarcane domestication was the first stage of yield increase, which provided cultivated forms with high productive potential owing to their thick stalk that accumulated large amounts of sucrose. Two of the three sweet domesticated species, *S. barberi* and *S. officinarum*, spread throughout intertropical areas, thanks to human migration and European colonization. *S. officinarum* was the main cultivated species until the end of the 19th century.

Yield Increase Brought by Wild Introgressions

The second step of yield improvement relied on the 100 years of breeding that began at the end of the 19th century with the exploitation of sugarcane fertility (Figure 13.3). The first breeding stations were established in Barbados and Java at the end of the 1880s (Daniels and Roach 1987). After making efforts aimed at intraspecific breeding within *S. officinarum*, breeders rapidly observed that interspecific crosses between *S. officinarum* and *S. spontaneum* were better able to overcome the diseases that threatened sugar industries. Adding *S. spontaneum* in the crosses provided resistance to mosaic and Sereh, both serious diseases that occurred in the 1920s. Agronomic performance was also improved because of *S. spontaneum*'s good ratooning

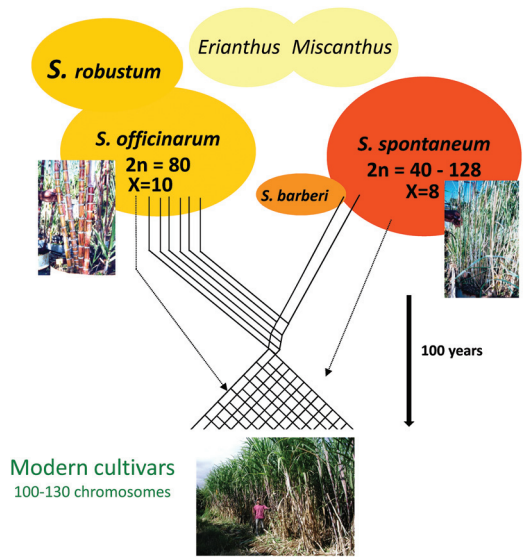


Fig. 13.3. Schematic breeding history of sugarcane, showing a second significant step in yield improvement in the 20th century. Modern cultivars (*Saccharum* spp.) rely on a few crosses undertaken 100 years ago between *S. officinarum* (the high-sucrose domesticated species) and *S. spontaneum* (a wild species without sucrose), involving a limited number of founders. These first interspecific hybrids were backcrossed with the high-sucrose species, and subsequently complex recurrent intercrossing of the best products gave rise to modern cultivars. For a color version of this figure, please refer to the color plate.

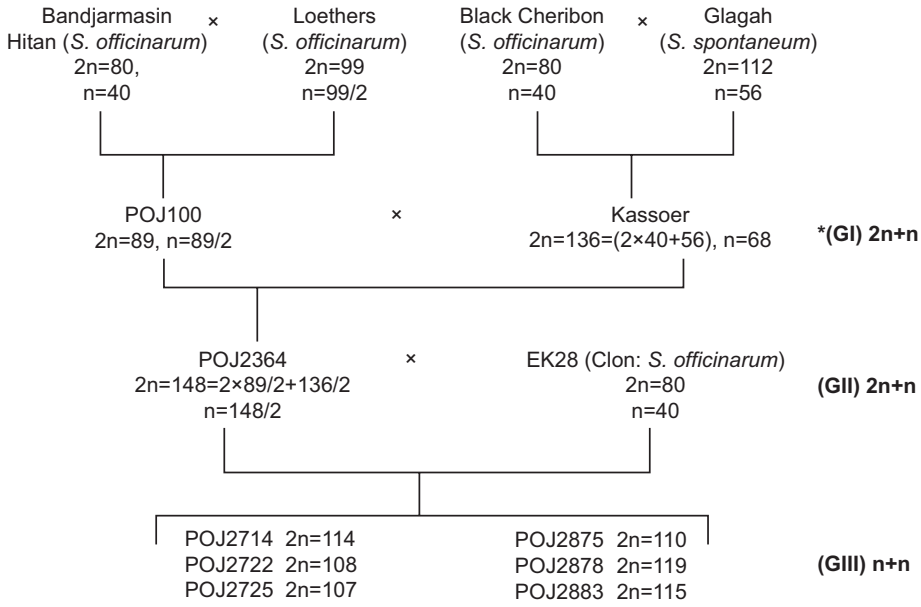


Fig. 13.4. 'Nobilization' breeding scheme that led to the few typical Javanese POJ interspecific founder cultivars obtained after three generations following ancestral species and used in most breeding programs worldwide in the 20th century. After a first interspecific (F1) hybridization (Kassoer) and a backross (BC) with a 'noble' cane (POJ 100) in which the somatic number (2*n*) of the two female genitors is transmitted, a second BC with a 'noble cane' (EK28) gave rise to outstanding 'nobilized' products (from Campo Zabala, 2010, with kind permission from the International Society of Sugar Cane Technologists).

ability and high tillering. Hybrids were produced from these two complementary species through an introgression process called "nobilization" (Figure 13.4). 'Nobilization' consists of an initial interspecific (F1) cross (♀ *S. officinarum* x ♂ *S. spontaneum*) followed by a few backcrosses (BC) with different 'noble' clones (*S. officinarum*). In F1 and BC1 crosses, *S. officinarum* used as female genitor has the particularity of transmitting its somatic (2*n*) chromosome number (Bremer 1922; Piperidis et al. 2010). The resulting products of introgression (BC2, BC3, etc.) thus rapidly recovered the high sugar-content phenotype characteristic of the 'noble' species (*S. officinarum*) along with sustainable yield due to newly acquired disease resistance. 'Nobilization' was first exploited at the Proefstation Ost Java (Java). In 1921, this breeding station created the famous 'nobilized' cane, POJ 2878. Its yield was 35% higher than the other varieties used at that time (Jeswiet 1930). The Sugarcane Breeding Institute at Coimbatore (India) used *S. sinense* and *S. barberi* as the male parent in their nobilization schemes and created famous cultivars such as Co 213, Co 281 and Co 290. Most of the superior 'nobilized' canes that were created between 1920 and 1930 spread throughout the world during the course of the 20th century. Since then, in all countries, improvement of sugarcane has been based on the recurrent intercrossing of elite cultivars derived from these few initial interspecific 'nobilized' founders, followed by mass selection among progenies. One of the main consequences of these breeding schemes is a relatively narrow genetic basis of modern

sugarcane germplasm (*Saccharum* spp.). The number of ancestral accessions that gave rise to the interspecific founders used as progenitors in breeding programs worldwide does not exceed 20 accessions originating from *S. officinarum* species and even fewer from *S. spontaneum* species (Arceneaux 1967; Roach 1989).

Modern Sugarcanes and Yield Progress

Modern cultivars have a large, complex aneuploid and polyploid genome consisting of 100-130 chromosomes of about 10 Gbp (D'Hont 2005). Seventy to eighty percent of the chromosomes are inherited from *S. officinarum*, 10 to 20% are inherited from *S. spontaneum,* and 10 to 20% are derived from recombination of the two species (Figure 13.5) (Cuadrado et al. 2004; D'Hont et al. 1996; Piperidis and D'Hont 2001; Piperidis et al. 2010). Modern sugarcane cultivars are highly heterozygous genotypes resulting from the well-known relative intolerance of the plant to inbreeding. As shown in Figure 13.5b, the genome of modern interspecific and aneuploid sugarcanes comprises 10 homeology groups (HG), each containing between 11 and 14 chromosomes. Chromosomes mainly form bivalents at meiosis, but, depending on the genotype and the HG, pairing may vary from preferential (0 to 40%) to complete (100%) affinities, leading to complex chromosome segregation patterns associating disomic with more or less polysomic behavior (Jannoo et al. 2004). This complex chromosome segregation pattern may invalidate several assumptions underlying the theory of quantitative genetics developed for conventional diploid (Hogarth 1977) or autopolyploid species. Using these models, estimates of genetic variance components derived from experimental mating designs are believed to be frequently biased. The development of specific population genetics or quantitative genetic models adapted to sugarcane is highly unlikely. This makes classical sugarcane breeding somewhat more

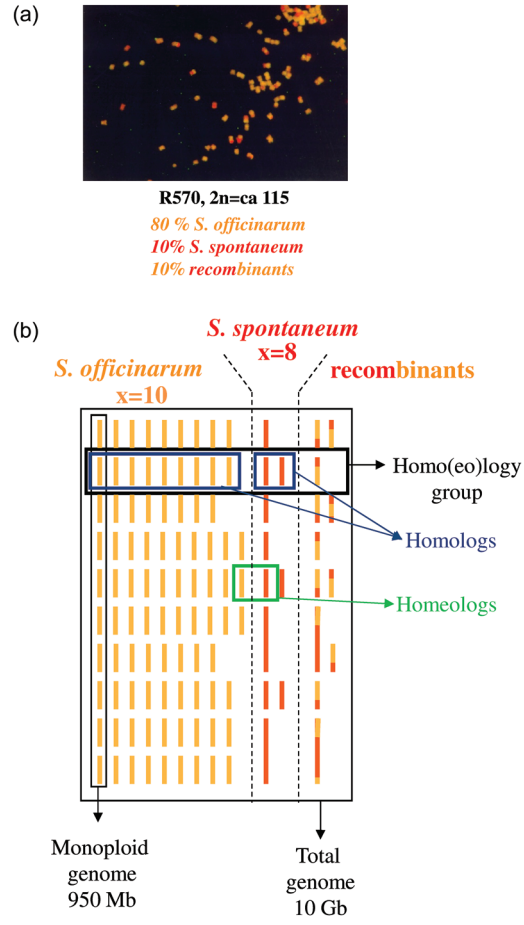


Fig. 13.5. Double structure of the genome of modern sugarcane cultivars. About 80% of the chromosome is inherited from *S. officinarum,* 10% from *S. spontaneum* and 10% are recombinant.
(a) *Genomic in situ hybridization* (GISH) using labeled total DNA of *S. officinarum* (yellow fluorescence) and of *S. spontaneum* (red fluorescence) on chromosome preparations of the cultivar R570 (from D'Hont et al. 1996, with kind permission from Springer Science and Business Media).
(b) Schematic representation of the genome of modern interspecific cultivars deduced from molecular cytogenetic data and mapping works. Modern cultivars are highly polyploid and aneuploid with about 120 chromosomes. Colored bars correspond to chromosomes. Chromosome colors are functions of their origins: yellow for *S. officinarum* and red for *S. spontaneum*. Chromosomes from the same row are homologous (or homeologous). Chromosomes of the *S. officinarum* and *S. spontaneum* part of the genome of the modern cultivars are distributed in 10 (X=10) and 8 (X=8) homology groups respectively. (Modified from Grivet and Arruda 2002). For a color version of this figure, please refer to the color plate.

empirical than for other crops. However, despite this tricky context, many breeders succeeded in obtaining new local cultivars (Machado 2001) that contributed to the steady increase in yield as evidenced by world statistics (Figure 13.2). On one hand, in the high polyploid and heterozygote sugarcane, a lot of crosses between elite lines frequently lead to large agronomic segregation of many traits, providing opportunities for genetic progress. On the other hand, conventional breeding programs have to be rather large (and expensive) to be efficient because of (1) the marked disjunction of the target traits of the breeders (yield components, disease resistance, etc.) in the currently 'unfixed' breeding germplasm, and (2) the absence of any information on the genetic basis of agronomic traits likely to be useful to rationalize experimental costs. Another reason for the relatively high cost of conventional breeding programs is the frequent genotype x environment interaction observed in sugarcane (Jackson and Hogarth 1992; Kang and Miller 1984), which requires investments in large experimental networks. Moreover the ratooning ability of selection candidates needs to be monitored over several crop cycles before the best sustainable elites suitable for semi-perennial cultivation can be identified. All these particularities explain why sugarcane breeding programs still rely on massive screening of millions of progenies. Programs call for tremendous experimental resources. Between 7 and 10 years are needed to select breeding parents and between 12 and 15 years to identify a commercial cultivar after initial crossing (Cheavegatti-Gianotto et al. 2011).

Further improvements in sugarcane yield are expected to come from a better understanding of the genetic bases of yield components that could facilitate the development of marker-assisted breeding approaches. However, the polygenic nature of the factors controlling the expression of yield traits, enhanced by the high ploïdy level of sugarcane, might hinder the use of molecular-assisted breeding approaches in sugarcane breeding programs.

Marker-Assisted Selection Related to Yield Component Traits

Marker-assisted selection (MAS) is based on the exploitation of linkage disequilibrium (LD) between markers and quantitative trait loci (QTLs). LD is the non-random association of alleles at distinct loci. If markers are tightly linked to many and/or prominent QTLs underlying the variation of any trait of interest, direct selection is possible using markers instead of phenotypic data, which can be time consuming to acquire. In this case, the advantages of MAS over phenotypic selection should enable a gain both in time and in cost, which could be even greater because MAS could be efficiently applied at a relatively early stage in the breeding program (depending on the unit cost of genotyping). The advantages and efficiency of MAS are based on the fact that QTL effects need to be accurately estimated and stable across genetic backgrounds and environments and over time. The prohibitive cost of sugarcane breeding programs provided motivation for testing the use of MAS breeding approaches for yield traits in sugarcane in order to identify elite genotypes as early as possible.

In sugarcane, detection of QTLs for MAS experiments relies on two strategies involving the study of highly heterozygous clones: QTL mapping of bi-parental crosses or association mapping using a panel of clones.

QTL Studies

QTLs genetic studies have been carried out on sugar yield, on cane yield, and on their agronomic components (tillering, stalk length and diameter, and brix). A total of 14 QTL studies based on ten different bi-parental progenies are reported in the literature (Sills et al. 1995; Ming et al. 2001; Hoarau et al. 2002; Ming et al. 2002a; Ming et al. 2002b; Jordan et al. 2004; Da Silva and Bressiani 2005; Reffay et al. 2005; Aitken et al. 2006; Aitken et al. 2008; Piperidis et al. 2008; Alwala et al. 2009; Pinto et al. 2010; Pastina et al. 2012). Table 13.1

Table 13.1. Review of allele tagging experiments related to sugarcane yield component traits, based on QTL studies or association mapping studies. Experimental context, QTL detection methods and results.

Population	Population size	Marker type	No. of markers	No. of locations	No. of years (crop cycle)[a]	Statistical analysis and threshold[b]	Traits[c]	QTLs detected across experiments	R^2 individual		References
									Min	Max	
Interspecific crosses											
La Purple (*S. off*) x Mol 5829 (*S. rob*)	44	RAPD	83	1	1(cp)	SM (marker wise p value <0.1)	Stalk number	2	–	–	Sills et al. (1995)
							Stalk diameter	3	–	–	
							Pol	4	–	–	
							Fiber	4	–	–	
							Plot biomass weight	2	–	–	
Green German (*S. off*) x IND81-146 (*S. spont*)	264	RFLP	475	1	1(cp)	SM (suggestive level# marker wise p value <0.003) IM (LOD score >2.5)	Sugar content	14	0.04	0.13	Ming et al. 2001
							Fiber	19	0.06	0.13	Ming et al. 2002b
							Stalk number	2	0.05	0.15	
							Sugar yield	3	0.05	0.11	
							Pol	2	0.05	0.09	
							Stalk weight	10	0.05	0.13	
							Stalk height	3	0.05	0.08	Ming et al. 2002a
PIN84-1 (*S. spont*) x Muntok Java (*S. off*)	239	RFLP	260	1	1(cp)	SM (suggestive level#, marker wise p value <0.003) IM (LOD score >2.5)	Sugar content	22	0.05	0.21	Ming et al. 2001
							Fiber	1	0.07	0.07	Ming et al. 2002b
							Stalk number	1	0.06	0.06	
							Sugar yield	7	0.05	0.15	
							Pol	12	0.04	0.15	
							Stalk weight	25	0.05	0.16	
							Stalk height	53	0.05	0.23	Ming et al. 2002a

(continued)

Table 13.1. *(Continued)*

Population	Population size	Marker type	No. of markers	No. of locations	No. of years (crop cycle)[a]	Statistical analysis and threshold[b]	Traits[c]	QTLs detected across experiments	R^2 individual		References
									Min	Max	
'Louisiana Striped' (*S. off)* x 'SES 147B' (*S. spont*)	100	AFLP, SRAP and TRAP	650	1	2(cp,r1)	IM and CIM (LOD >3.01); nonparametric discriminant analysis	Early brix	8	0.03	0.34	Alwala et al. (2009)
					1(cp)		Late brix	6	0.04	0.27	
					1(r1)		Early pol	8	0.08	0.15	
					1(cp)		Late pol	3	0.09	0.10	
Crosses involving modern cultivars											
selfing of R570	295	AFLP	1180	1	2(cp, r1)	SM (marker wise p value < 0.005)	Brix	2	0.03	0.04	Hoarau et al. (2002)
							Stalk height	1	0.03	0.03	
							Stalk diameter	2	0.03	0.05	
							Stalk number	1	0.07	0.07	
Q117 x 74C42	108	RFLP and RAFs (radio-labelled amplified fragments)	258	2	2(cp; r1)	SM (marker wise p value <0.01)	Stalk number	16	–	–	Jordan et al. (2004)
							Sucker number	14	–	–	
Q117 x MQ77–340 two moderns cultivars	232	AFLP and SSR	400 (only for MQ77-340)	1	2(cp, r1)	SM (marker wise p value <0.01)	Early season CCS	2	0.03	0.04	Reffay et al. (2005)
							Early season brix	1	0.03	0.03	
							Early season pol	1	0.03	0.03	
					1(cp)	SM (marker wise p value <0.01)	Late season CCS	5	0.03	0.07	
							Late season brix	5	0.03	0.06	
							Late season pol	5	0.03	0.07	
							Fiber	2	0.02	0.03	
							Stalk weight	8	0.02	0.07	
							Cane yield	3	0.02	0.03	
							Sugar yield	3	0.02	0.03	

			>1000	1	2(cp, r1)	SM (marker wise p value <0.01)	Early season CCS	8	0.03	0.06	Piperidis et al. (2008)
							Early season brix	6	0.03	0.07	
							Early season pol	5	0.03	0.07	
					1(cp)		Late season CCS	14	0.03	0.07	
							Late season brix	15	0.03	0.07	
							Late season pol	14	0.03	0.06	
IJ76-514 (*S. off*) x Q165 (modern cultivars)	230	AFLP; SSR	2238	2	2(cp; cp)	SM (suggestive level# marker wise p value <0.003) (analysis with combined data)	Early season brix (%)	9	0.04	0.05	Aitken et al. (2006)
							Mid season brix (%)	11	0.04	0.07	
							Early season Pol (%)	5	0.04	0.05	
							Mid season Pol (%)	10	0.04	0.06	
				1	2(cp; r2)	SM (suggestive level# marker wise p value <0.003)	Cane yield (t/ha)	2	0.04	0.06	Aitken et al. (2008)
							Stalk height (cm)	1	0.05	0.05	
							Stalk diameter (mm)	4	0.03	0.08	
							Stalk number	2	0.04	0.06	
				2	3 (cp; cp; r2)		Stalk weight (cm)	1	0.05	0.10	
SP80-180 x SP80-4966 two pre-commercial cultivars	108	EST derived RFLP	16	1	1(cp)	SM (marker wise p value <0.01)	Pol	1	0.24	Da Silva et Bressiani (2005)	

(continued)

Table 13.1. *(Continued)*

Population	Population size	Marker type	No. of markers	No. of locations	No. of years (crop cycle)[a]	Statistical analysis and threshold[b]	Traits[c]	QTLs detected across experiments	R^2 individual		References
									Min	Max	
	100	RFLP among them 35 are EST-derived	222	1	2(cp,r1)	SM (marker wise p value <0.05)	Pol	7	0.04	0.13	Pinto et al. (2010)
							Fiber	7	0.04	0.11	
							Cane yield	7	0.04	0.18	
							Sugar yield	5	0.04	0.20	
		RFLP, EST-RFLP, EST-SSR	741	2	3(cp, r1,r2)	Interval mapping and SM marker wise p value <0.01), mixed models	Cane yield	13	–	–	Pastina et al. (2012)
							Sugar yield	14	–	–	
							Fiber	11	–	–	
							Sucrose content	8	–	–	
Panel of modern cultivars											
Panel of modern clones. Half Diversity panel and half progenies from 31 biparental crosses	480	DArTs	1531 (discrete) 15360 (continous)	3	1(cp)	SM (marker wise p value <0.01) Several mixed models	CCS	42 (discrete markers) 377(continuous)	–	–	Wei et al. (2010)
							Cane yield	47 (discrete markers) 352 (continuous)	–	–	

[a] cp: plant crop; r1: first ratoon; r2: second ratoon.
[b] SM: single marker analysis; IM: interval mapping; CIM: composite interval mapping.
[c] CCS: commercial cane sugar.
#: suggestive level according to Lander and Kruglyak (1995).

summarizes QTL detection methods and results. All these QTL studies relied on pseudo 'F2' populations derived from selfed or bi-parental crosses of highly heterozygous sugarcane clones. Depending on the study concerned, populations were genotyped with different types of molecular markers (RAPD, RFLP, AFLP, SSR, TRAP, SRAP, and DArTs). Linkage maps (based on associations of markers in the coupling phase), were constructed mainly using single-dose (SD) markers but occasionally double-dose (DD) markers (Alwala and Kimbeng 2010; Wu et al. 1992). Detection of associations of marker/traits was most often based on marker-by-marker analysis of variance, or more rarely on interval mapping. Among genetic maps, the two most advanced ones for cultivar R570 (Hoarau et al. 2001) and cultivar Q165 (Aitken et al. 2005), which contain around 1,000 markers, still cover less than 50% of the sugarcane genome (Piperidis et al. 2008). Two reasons related to the high polyploidy context may explain this incomplete coverage of current interspecific cultivars: (1) the size of their genome, which is estimated to be around 17,000 cM (Hoarau et al. 2001), and (2) the high chromosome redundancy within the fraction of the genome inherited from *S. officinarum* (about 80%), which implies a lower frequency of single-dose (SD) markers compared to the frequency of SD markers within the genome fraction inherited from *S. spontaneum* (Grivet et al. 1996; Hoarau et al. 2001; Rossi et al. 2003).

Interspecific Crosses

The first category of QTL studies for yield components was carried out on interspecific crosses, with the aim of maximizing segregation of the phenotypic variation in yield. These studies involved *S. officinarum*, the domesticated species, with either the wild species *S. robustum* in one cross (Sills et al. 1995) or the wild species *S. spontaneum* in three crosses (Alwala et al. 2009; Ming et al. 2002a; Ming et al. 2001; Ming et al. 2002b). The first research was conducted by Sills and coworkers (1995) on Purple (*S. officinarum*) x Mol 5829 (*S. robustum*). They identified 12 markers linked to five traits associated with yield, but the size of the progeny was small (44) and the *P* value threshold was high (0.1). Ming and co-workers (2001) studied sugar content by measuring sucrose-related traits (brix and pol) in two interspecific populations – Green German (*S. officinarum*) x IND81-146 (*S. spontaneum*), and PIN 84-1 (*S. spontaneum*) x Muntok Java (*S. officinarum*) – of similar size (264 and 239 individuals respectively). Despite differences in genotyping efforts between the two studies, a similarly high percentage of the total phenotypic variation of sucrose content was explained in both studies by the markers detected, which were R^2=65% (14 SD markers) and 68% (22 SD markers), respectively. In these two populations, Ming and colleagues (2002a, 2002b) also detected a total of 82 QTLs associated with traits related to cane yield components (stalk weight and height, tillering, and fiber content). Considering all traits together (sugar content and cane yield components), individual R^2 markers frequently ranged from 4% to about 16% and up to 23% for one particular trait (Ming et al. 2001; Ming et al. 2002a; Ming et al. 2002b). These relatively high upper values in individual R^2 ranges can obviously be ascribed to the interspecific nature of the mapping populations that offered large segregations for most of the traits, possibly due to a few rare alleles of major effects. Congruently, a second team, also studying an interspecific cross (*S. officinarum* x *S. spontaneum*) (Alwala et al. 2009) reported markers of high individual effect size (up to 34% for brix content) in a smaller population (only 100 individuals). In addition, research by the teams of Alwala and Ming (Ming et al. 2001; Ming et al. 2002a; Ming et al. 2002b; Alwala et al. 2009) revealed the existence of unfavorable alleles for sucrose content in the favorable ancestor (*S. officinarum*) and the reverse (favorable alleles in *S. spontaneum*), even if the direction of the majority of the QTLs contributing to trait variation remains congruent with agronomic predictions. These results suggest the existence

of potential additional gains for sucrose content within 'noble' species and illustrate the potential advantage of using MAS approaches to purge the domesticated species of alleles that reduce sucrose content.

Modern Cultivar Crosses

The second category of QTL studies was based on five crosses involving modern cultivars (Table 13.1): selfing of R570 (Hoarau et al. 2002), SP80-180 × SP80-4966 (Da Silva and Bressiani 2005; Pinto et al. 2010; Pastina et al. 2012), Q117 × 74C42 (Jordan et al. 2004), Q117 × MQ77-340 (Reffay et al. 2005; Piperidis et al. 2008), and a cross between the *S. officinarum* clone IJ76-514 and the modern cultivar Q165 (Aitken et al. 2006; Aitken et al. 2008). All these studies used classical measurements of yield components such as height, diameter, weight or number of stalks, and measurements of the quality of sugarcane juice (brix and pol). Three of these studies based on the largest population size (230 to 295 individuals) revealed numerous QTLs for each trait with small individual R^2 values usually ranging between 3% and 8% (Hoarau et al. 2002; Reffay et al. 2005; Aitken et al. 2006; Aitken et al. 2008; Piperidis et al. 2008). These R^2 values are lower than values found in previous works by Ming and colleagues (Ming et al. 2001; Ming et al. 2002a; Ming et al. 2002b) for F1 interspecific populations of similar size (239 and 264 individuals, respectively) with R^2 that were up to 23%. These findings show that the frequency of alleles for major effects for yield-related traits is lower in modern varieties. This could be due to the fact that: (1) the most unfavorable alleles in the *S. spontaneum* part of the modern sugarcane genome are likely to have been eliminated by 'nobilization,' and (2) some of the most favorable alleles have already been 'fixed' in several copies by recurrent selection, which makes their detection less likely (absence of visible segregation). As a consequence of their relatively modest effect size, the number of QTLs detected in modern cultivars mainly depended on the thresholds used to qualify an association as statistically significant. Like QTL studies of interspecific crosses (Sills et al. 1995; Ming et al. 2002b; Alwala et al. 2009), some modern QTL studies revealed 'significant' digenic interactions (Hoarau et al. 2002; Aitken et al. 2006; Aitken et al. 2008; Pinto et al. 2010). These digenic interactions are responsible for a substantial portion of phenotypic variance, illustrating the potential importance of epistasis in the genetic control of yield components. In several QTL studies, populations were phenotyped in successive crop cycles (Hoarau et al. 2002; Reffay et al. 2005; Piperidis et al. 2008; Pinto et al. 2010) and sometimes in several locations (Jordan et al. 2004; Aitken et al. 2006; Aitken et al. 2008; Pastina et al. 2012). Although Aitken and colleagues (2006) combined data across locations for statistical analysis, in most of these studies, detection of marker-trait associations was carried out separately for each crop cycle and environment. They usually revealed a small number of overlapping sets of QTLs, however the direction of the marker effect on the trait value (either positive or negative) may always be conserved (Hoarau et al. 2002). These findings illustrate the fact that detecting QTLs in modern cultivars is highly sensitive to the effects of statistical thresholds. Only one work, recently published by Pastina and colleagues (2012), describes the use of mixed models for detection of yield QTLs, revealing significant interactions for all traits for QTL x crop cycle, QTL x environment, and QTL x crop cycle x environment. Piperidis and colleagues (2008) compared the location of QTLs for brix across four modern cultivar maps (R570, MQ77-340, Q117, and Q165), based on the use of a few neutral SSR primers scattered throughout the eight homo(eo)logy groups (HG). Two of the eight HGs were seen to contain marker-trait associations for brix, in two or three out of the four maps, suggesting common loci of interest in these HGs among cultivars. These results illustrate the value of conducting meta-QTL analyses to reveal key alleles that could be targeted using MAS approaches in genetic improvement program.

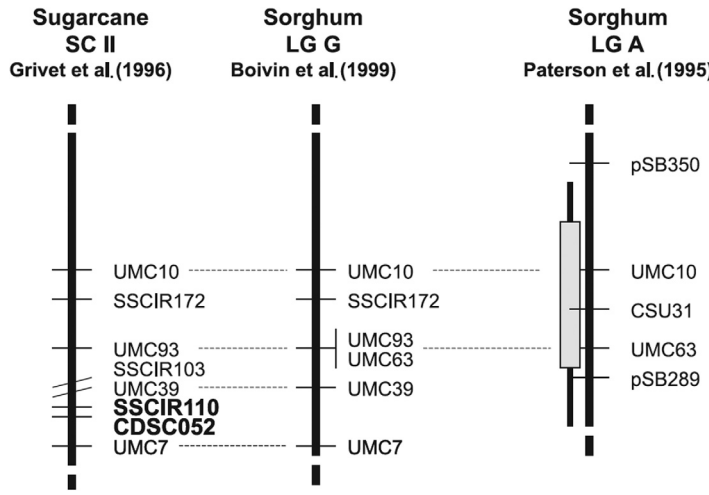


Fig. 13.6. Alignment of a chromosome region of sorghum bearing QTLs related to the regrowth (rectangle) in the linkage group A (LG A) (Paterson et al. 1995) with the orthologous region in sugarcane chromosome II (SC II) of sugarcane containing markers (SSCIR 110 and CDSC052 in bold) associated with QTLs for suckering (from Jordan et al. 2004, © 2008 Canadian Science Publishing or its licensors. Reproduced with permission).

Another strategy for tracking down alleles related to yield components in the very large sugarcane genome would be to take advantage of the relatively good synteny relationships between sugarcane and diploid grass models (Glaszmann et al. 1997; Ming et al. 1998; Jannoo et al. 2007; Le Cunff et al. 2008; Wang et al. 2010). Ming and colleagues (2002a; 2001) found syntenic regions between sugarcane and maize or sorghum that contain QTLs controlling sugar content, plant height, number of stalks, and flowering in sugarcane. Like what is shown in Figure 13.6 using heterologous Restriction Fragment Length Polymorphism (RFLP) probes, Jordan and colleagues (2004) found seven QTL colocalizations between sugarcane and sorghum related to tillering and rhizomatousness traits. These results clearly demonstrate the potential of allele-tagging strategies based on the exploitation of synteny.

Several studies have used the large Expressed Sequence Tag (EST) database available for sugarcane (Vettore et al. 2003) to develop molecular markers (RFLP, Simple Sequence Repeats [SSR], and Target Region Amplification Polymorphism [TRAP]) showing homology with candidate genes involved in yield elaboration (Da Silva and Bressiani 2005; Alwala et al. 2009; Pinto et al. 2010). This strategy allowed direct mapping of the genes of interest, as demonstrated by Da Silva and Bressiani (2005) and Pinto and colleagues (2010), who described the development of EST-RFLP markers and found a sucrose synthase EST-RFLP marker associated with sugar content in an SP80-180 x SP 80-4966 cross. The sugarcane EST resources were also the source of the discovery of single nucleotide polymorphisms (SNPs) (Grivet et al. 2003; Cordeiro et al. 2006; McIntyre et al. 2006). The ecotilling strategy that enabled the discovery of SNP in a target EST was tested by McIntyre and colleagues (2006) and Aitken and colleagues (2008) for detecting and mapping associations with yield components. The whole genome sequence already available for sorghum, maize, and rice should improve these strategies and facilitate the development of accurate candidate markers associated with yield.

Association Mapping

Association studies based on linkage disequilibrium (LD) can also be used to tag QTLs without necessarily calling for detailed linkage information. Unlike linkage analysis based on controlled progenies, LD-based studies do not require segregating populations of known parentage. Linkage, but also selection, or drift, in a population are the main causes for allelic associations to occur at a different frequency from what would be expected if the associations were due to random mating (Flint-Garcia et al. 2003). LD-based studies, or association-mapping studies, are based on existing populations or germplasm collections, which may have major advantages: (1) individuals in such collections may be more or less distantly related and may have accumulated recombination events over many generations and consequently allow for high-resolution mapping, (2) such collections may already be well characterized for a range of interesting traits, and (3) such collections may include a large number of alleles of agronomic value (Morgante and Salamini 2003; Rafalski and Morgante 2004). However, association-mapping studies suffer from certain limitations. There is a higher probability of type I and type II errors compared to classic bi-parental QTL analysis (Breseghello and Sorrells 2006). Type I errors, or detection of false marker-trait associations, may be the result of the genetic structure within the population studied. Type II errors, that is, the probability of missing genuine causal associations, may result from (1) lower associations between markers and genes resulting from the rapid decay of LD, (2) unbalanced design resulting from the presence of alleles at distorted frequencies, and (3) very strict genome-wise significance thresholds resulting from the relative independence of the many markers tested (Carlson et al. 2004). The extent of LD determines whether genome scans or candidate gene association approaches can be used (Nordborg and Tavaré 2002; Flint-Garcia et al. 2003). The potential of LD approaches in sugarcane breeding was highlighted early on (Jannoo et al. 1999). Only a few generations separate modern cultivars from their interspecific founders, limiting the number of meioses and consequently the opportunity for chromosome recombination. Moreover breeding history is characterized by bottlenecks arising from the very limited number of founders (Arceneaux 1967). As a result, despite the relatively large size of the sugarcane genome (about 10 Gb), modern germplasm is thought to encompass a rather modest number of LD blocks. None of the classic measures of LD (D', r^2, d^2) exploiting allele frequency or haplotype frequency can be calculated because of the high polyploidy of sugarcane. However the Fisher exact test probability can be used to test for associations between markers. Using information from a reference map (Hoarau et al. 2001) combined with the study of a small panel of sugarcane cultivars, Raboin and colleagues (2008) assessed LD in modern sugarcanes. As predicted, LD appeared to be more extensive than in numerous other plants, since LD drops sharply only over a distance of 5 cM and instances of LD blocks of 10 to 20 cM are relatively frequent. But many LD blocks may be missed, as the confounding effects of marker dosage due to polyploidy are assumed to mask many instances of linked markers (Costet et al. 2012).

Several authors have applied association mapping in sugarcane relative to disease and insect resistance such as smut, African stalk borer, pachymetra root rot, leaf scald, and Fiji leaf gall (Raboin 2005; Wei et al. 2006; Butterfield 2007), or to yield component traits (Wei et al. 2010). These studies have led to the detection of numerous marker-trait associations, despite the use of a modest number of markers, far from the number required for a "meta" genome of a cultivar panel to be densely scanned. Sugarcane haploid genome size is about 1500 cM. Therefore, considering average LD values in sugarcane, a minimum of 300 to 600 multi-allelic locus-specific markers would be required to achieve a minimum density of one or two markers every

five cM, corresponding to 3,000-7,000 markers to cover the entire polyploid genome (Raboin et al. 2008).

Wei and colleagues (2010) recently published the only association-mapping study related to cane and sucrose yields using a large panel of 480 clones representative of the modern germplasm of current breeding programs. The panel was tested in three different locations using a replicated design. This panel was genotyped with a Diverse Arrays Technology (DArT) microarray (Heller-Uszynska et al. 2010). The authors generated 15,360 DArT score markers and used them as continuous markers, assuming that this quantitative scoring was related to the number of copies of each marker per genotype, which would be a major advantage in the highly polyploid sugarcane. Nevertheless these 15,360 markers corresponded to 1,531 discrete polymorphic markers within a 0.05-0.95 frequency range. Several methods can be used to account for structuration in the panels. A Bayesian clustering model implemented in STRUCTURE software (Pritchard et al. 2000) or kinship analysis inferred from molecular data implemented in SPAGEDI software (Hardy and Vekemans 2002; Yu et al. 2006) are currently used in association-mapping studies in plants but are not suitable for high autopolyploid species. Principal component analysis (Price et al. 2006), genomic control (Devlin and Roeder 1999), or pedigree matrix are suitable alternative methods to account for the structure of the panel in the genomic context of sugarcane. Wei and colleagues (2010) analyzed marker/trait associations with different mixed linear models they adapted from the conventional basic model of Yu and colleagues (2006) to minimize the risk of type I and type II errors. These models combined cofactors representing levels of relatedness between individuals, genotype by environment interactions, and spatial variation within trials. The cofactor used for levels of relatedness between individuals was a pedigree matrix that is very efficient, as it dramatically reduces the number of significant associations. Finally, using the most elaborate model, with $P<0.01$, the authors detected 47 discrete and 352 continuous markers associated with cane yield, and 42 discrete and 377 continuous markers associated with sugar content. Depending on the traits surveyed, these numbers of associations were 3 to 6-fold lower when considering a threshold *P* value of 10^{-3}. Six associations for sucrose content were still perceptible at $P<10^{-4}$, while at this threshold no significant markers would be expected by random chance. The results of this first association study of sugarcane yield components are encouraging, thanks to the large panel surveyed and the phenotypic data that were acquired on a multi-loci experimental basis. However, the repeatability of these marker trait associations and validation of these markers at the scale of a breeding program remain the major concern.

Perspectives

In highly polyploid sugarcane, tagging useful QTLs related to complex traits such as cane or sucrose yield will always be a challenge, irrespective of the strategy used, whether bi-parental QTL or panel association studies. Regarding ploidy levels, one would expect a reduction in the mean size of allele effects within the loci of interest. Analyses show medium to large proportions of trait variation explained by a swarm of barely significant QTLs with small individual effect size (R^2). Moreover QTLs are often specific to a single environment and crop cycle. Validating QTLs across different genetic backgrounds for breeding purposes is a difficult task (Piperidis et al. 2008). In sugarcane, as in other crops, the use of MAS appears to be useful for traits with simple inheritance, such as disease resistance (Daugrois et al. 1996; Raboin et al. 2006; Aljanabi et al. 2007). The advantage of using of MAS for quantitative traits such as yield is more questionable. Up until now, the only known example of using markers for selection in sugarcane is for rust resistance (Costet et al. 2012). Numerous QTL mapping experiments in many species have been published in

the last 20 years. Results of such studies have demonstrated the limited use of QTL mapping for breeding (Xu and Crouch 2008; Jannink et al. 2010). According to the recent review by Brumlop and Flinckh (2011), a total of 83 papers were published between 1995 and 2009 concerning the main areas of applicability for MAS in plant breeding programs or in research projects. This survey included very few studies (only 8 out of 83) that reported successful application of MAS related to yield improvement. The main problems involved in using MAS for quantitative traits are the difficulty of accurately determining the effects of the QTL, and of extrapolating QTL expression from one genetic background to another and from one environment to another. This may explain why molecular markers are not currently used in breeding schemes for the improvement of sugarcane yield. New models need to be developed that take the high complexity of the sugarcane genome into account. Effective incorporation of molecular genetics in breeding programs will also depend on the availability of innovative genotyping technologies yielding higher throughput markers to enable more dense coverage of the large sugarcane genome (10 Gb).

Towards Increasing Throughput Marker Systems

Efforts have to be invested in developing "universal" markers that would enable comparison of the location of QTLs across sugarcane studies and also between sugarcane and related species (sorghum, miscanthus, maize). The achievement of this objective should be supported by the large scale development of markers based on DNA sequences via high throughput genotyping facilities, such as microarray and sequencing technologies. The recent release of the reference sequence of the sorghum genome (Paterson et al. 2009) which has a large degree of synteny with the sugarcane genome (Glaszmann et al. 1997; Ming et al. 1998; Jannoo et al. 2007; Le Cunff et al. 2008; Wang et al. 2010) should also provide a valuable guide for probable gene arrangements in sugarcane, both at the global and local scales. This new resource should facilitate strategies aimed at defining a reference sequence of the sugarcane genome based on bacterial artificial chromosome (BAC) sequencing approaches. Regarding the complexity and the size of the genome, Souza et al. (2011) suggest focusing sequencing on 'euchromatin' BAC regions identified using the sorghum sequence template. Euchromatin regions are believed to be rich in genes and might include most of the recombination scattered across the genome. Conversely 'heterochromatin' is gene-poor, repeat-rich, and recalcitrant to recombination. Therefore, sequencing at least the gene-rich portions of the sugarcane genome should provide a valuable resource for the development of high throughput marker systems in tight linkage disequilibrium with many QTLs of agronomic interest. Souza and colleagues (2011) predict that sequencing about 4,000-5,000 sugarcane BACs could capture much of the euchromatin.

To further develop high-throughput marker systems from this draft sequence of euchromatin, advantage should be taken of next-generation sequencing (NGS) technologies. One of the first investigations of the potential of NGS in sugarcane was conducted by Bundock and colleagues in 2009, using 454 Genome Sequencer FLX. These authors demonstrated that the discovery of reliable single nucleotide polymorphism (SNP) in sugarcane is feasible and would enable the dosage of each allele to be measured. To envision a global NGS-based strategy for SNP detection, we need to improve our understanding of genome organization and evolution linked to polyploidization in order to assess the extent of polymorphism existing among homo(eo)logous loci. Two recent studies (Jannoo et al. 2007; Le Cunff et al. 2008) based on the structural analysis of two series of homo(eo)logous BACs provided a first insight into genome dynamics by revealing perfect colinearity as well as high gene-structure conservation between sugarcane

homo(eo)logous haplotypes. Sugarcane does not appear to have undergone a major reshaping of its genome as a consequence of polyploidization. Additional series of homo(eo)logous BACs are needed to refine these first results.

A recent NGS-based strategy called genotyping-by-sequence (GBS), associated with one step aimed at reducing genome complexity (generating a limited amount of sequence data to be analyzed), is another new way to discover SNP (Baird et al. 2008; Elshire et al. 2011). The concept is based on acquiring the sequence adjacent to a set of particular restriction enzyme recognition sites, rather than randomly sequencing the whole genome. Large amounts of polymorphism data can be generated by massive parallel sequencing. This approach increases the coverage for a given sequenced site, increasing both the confidence in base identity and the likelihood that the same sites will be sequenced in multiple samples. This promising approach, which could allow simultaneous SNP discovery and genotyping, is currently under investigation in sugarcane (Glynn et al. 2011; D'Hont pers. com.).

Model-Assisted Phenotyping

For overcoming the problems caused by genotype by environment interactions in yield traits possible solutions include model-assisted phenotyping. The "gene-to-phenotype" approach connects ecophysiological models to statistical methods for detecting complex traits (Hammer et al. 2004; Chenu et al. 2009; Prudent et al. 2011). This approach should improve the detection of QTLs and our understanding of the genetic architecture of complex biological processes. Phenotypic traits for production potential interact strongly with the environment, and are the result of multiple processes that are difficult to measure and tag at the genetic level. Plant growth modeling can advance our understanding of complex biological systems by formalizing dynamic interactions among several biological processes, such as those related to phenology, morphogenesis, carbon acquisition, and allocation among sinks. Yield formation can thus be described dynamically as a set of interactive equations using only a small number of genotypic parameters. These parameters control plant reaction norms that are at the basis of plant growth response to the environment (Dingkuhn et al. 2005). The different parameters involved in phenotype expression can be considered as synthetic component traits, presumably controlled by fewer genes than the integrative, complex agronomic trait. In this sense, if the models represent relevant biological processes, the model parameters measured can be expected to be closer to gene or QTL effects, in the sense that Genotype x Environment 'noise' is reduced (Hammer et al. 2002; Reymond et al. 2003; Yin et al. 2003; Dingkuhn et al. 2005). In such a heuristic approach, variation in parameters among genotypes can be interpreted as the expression of allelic diversity and analyzed accordingly in QTL or association studies. Model parameter values can be estimated by optimizing a relevant criterion based on deviations of predictions from observations, using target files containing observations on the plant (classical phenotyping) and on the environment (e.g., weather, soil). Traits resulting from ecophysiological models can be used for more detailed investigation of the biological processes involved in the development of yield. Model-assisted phenotyping has already been used for peach (Quilot et al. 2004; Quilot et al. 2005), barley (Yin et al. 1999), maize (Reymond et al. 2003), and rice (Dingkuhn et al. 2006; Luquet et al. 2006; Luquet et al. 2007). Several studies allowed ecophysiological modeling of sugarcane yield elaboration with models like Mosicas (Martiné et al. 2000; Martiné et al. 2001; Martiné 2003; Martiné 2007), Apsim (Keating et al. 1999; Keating et al. 2003), and Canegro (O'Leary 2000). Some authors are beginning to evaluate the feasibility of model-assisted phenotyping in sugarcane (Luquet et al. 2010; Martiné et al. 2010; Nibouche et al. 2010).

Genomic Selection

Recently a new MAS strategy, Genomic Selection (GS), became popular in animal breeding, after the seminal paper of Meuvissen and colleagues (2001). This strategy simultaneously estimates all marker effects across the entire genome. In this case, instead of many conventional marker-assisted selection (MAS) approaches, there is no defined subset of significant markers. Marker effects are estimated from a training panel using phenotypic and genotypic panel data and pedigree or kinship information. The accuracy of phenotype prediction is tested on a validation panel. The concept and application of GS need to be investigated in plant breeding (Heffner et al. 2009). To date, a few studies have been published on plants, most of which explore the potential benefits of GS from a theoretical viewpoint, based on simulations that are not tailored to any particular biological breeding context, except for a few studies (Bernardo and Yu 2007; Wong and Bernardo 2008). GS's potential resides in the use of all markers as predictors of complex trait performance, thereby capturing more of the variation due to QTLs that are hardly significant. The training population used in GS is generally representative of the genetic diversity used in breeding programs (wide range of allelic diversity and genetic background). Therefore potential applications at the scale of a whole breeding program are expected to be more efficient than any MAS approach derived from QTL studies (Heffner et al. 2009; Heffner et al. 2010; Jannink et al. 2010). GS approaches, which have not yet been tested in sugarcane, appear to be very attractive for breeders, given the high complexity of the sugarcane genome and the complex quantitative nature of the majority of agronomic traits they need to tackle.

Conclusion

Sugarcane is one the most efficient plants for biomass production. However in recent decades, the increase in yield has been slower than in other major crops. The importance of the genotype x environment interaction in yield and sucrose-related traits may explain the slower progress observed in sugarcane breeding as well as the highly quantitative nature of trait determinism. Cultivar improvement has relied largely on traditional breeding methods. These methods involve a lengthy (12–15 years) and expensive process of selection of plants with the desired agronomic traits. Molecular genomics should help improve programs by generating molecular markers that can assist the breeding process or the introgression of new genes into breeding germplasm. However, the development of genomic applications in sugarcane breeding programs has lagged behind many of the diploid plant models. Adapting genomic approaches developed for model plants to sugarcane is rarely straightforward owing to the complication of its high ploidy level. In the last two decades many QTL studies based on bi-parental crosses revealed many QTLs with small effects, while validation of the stability of a QTL's effect across genetic backgrounds, time, and environments remains a challenging task. Conventional MAS approaches are useful only for tagging alleles of major effects related to disease resistance (Al Janabi et al. 2007; Costet et al. 2012), but not for complex traits such as cane yield or quality characters. Moreover traditional marker systems are not powerful or convenient enough to densely scan the sugarcane genome and to develop efficient routine applications.

Given the prohibitive cost of conventional programs, the entire sugarcane research community has a major interest in multidisciplinary studies aimed at unraveling the determinism of complex traits in order to design molecular breeding approaches tailored to 'real life' breeding. The main expectation is more advanced incorporation of association genetics-based approaches in sugarcane improvement. In the context of sugarcane polyploidy, specific scientific creativity will be needed to design a breeder-friendly marker system as well as appropriate bioinformatics pipes adapted to routine genotyping. A major innovation in sugarcane

genetic analyses would be the availability of SNP codominant markers along with the development of statistical tools tailored to infer marker dosage in highly polyploid sugarcane. Such innovative technologies and models would help overcome the poor informative conventional (low-throughput) dominant marker systems that prevent sugarcane breeders from applying deep genetic analysis. Comparative genomics among grasses, based on the dramatic increase in sequencing and bioinformatics capabilities, also presents a powerful opportunity to densely scan regions harboring candidate genes. Finally, research should be inspired by concepts developed for model plants and animal systems, such as genomic selection and model-assisted ecophysiological phenotyping. Genomic selection approaches should improve estimation of the effects of markers, including the swarm of small-effect markers. Plant growth modeling could provide sounder ecophysiological traits and parameters to describe the complex biological process underlying yield and sucrose elaboration. Such models could circumvent traditional problems caused by genotype by environment interaction.

References

Aitken K, Hermann S, Karno K, Bonnett G, McIntyre L, Jackson P, 2008. Genetic control of yield related stalk traits in sugarcane. *Theoretical and Applied Genetics* 117, 1191–1203. http://dx.doi.org/10.1007/s00122-008-0856-6.

Aitken K, Jackson P, McIntyre C, 2005. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo (eo)logous linkage groups in a sugarcane cultivar. *Theoretical and Applied Genetics* 110, 789–801. http://dx.doi.org/10.1007/s00122-004-1813-7

Aitken K, Jackson P, McIntyre C, 2006. Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum* spp.) cultivar × *Saccharum officinarum* population. *Theoretical and Applied Genetics* 112, 1306–1317. http://dx.doi.org/10.1007/s00122-006-0233-2

Alexander AG, 1985. *The energy cane alternative*. Amsterdam: Elsevier Science Publishers BV, 509pp.

Aljanabi S, Parmessur Y, Kross H, Dhayan S, Saumtally S, Ramdoyal K, Autrey L, Dookun-Saumtally A, 2007. Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. *Molecular Breeding* 19, 1–14. http://dx.doi.org/10.1007/s11032-006-9008-3

Alwala S, Kimbeng C, 2010. Molecular genetic linkage mapping in *Saccharum*: Strategies, resources and achievements. In: Henry R, Chittaranjan K (eds), *Genetics, genomics, and breeding of sugarcane*. Enfield NH: CRC Press, Science Publishers, pp. 70–96.

Alwala S, Kimbeng C, Veremis J, Gravois K, 2009. Identification of molecular markers associated with sugar-related traits in a *Saccharum* interspecific cross. *Euphytica* 167, 127–142. http://dx.doi.org/10.1007/s10681-008-9869-0

Anonymous, 2012. REVIEW: 2011 International Sugar Organisation Seminar, London, UK. *International Sugar Journal* 114, 41–43.

Arceneaux G, 1967. Cultivated sugarcanes of the world and their botanical derivation. *Proceeding of the International Society of Sugar Cane Technologists* 12, 844–854.

Arruda P, 2011. Perspective of the sugarcane industry in Brazil. *Tropical Plant Biology* 4, 3–8. http://dx.doi.org/10.1007/s12042-011-9074-5

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA, 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3(10): e3376. http://dx.doi.org/10.1371/journal.pone.0003376

Bernardo R, Yu J, 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Science* 47, 1082–1090. http://dx.doi.org/10.2135/cropsci2006.11.0690

Botha F, 2009. Energy yield and cost in a sugarcane biomass system. *Proceeding of the Australian Society of Sugar Cane Technologists* 31, 1–10.

Brazilian Ministério de Minas e Energia, 2011. Balanço Energetico Nacional 2011, Resultados Preliminares, ano base 2010, p. 49. https://ben.epe.gov.br/downloads/Resultados_Pre_BEN_2011.pdf

Bremer G, 1922. Een cytologisch onderzoek van eenige soorten en soortsbastaarden van het geslacht *Saccharum*. *Arch Suikerindust Nederl Indic Meded Proefsta Java Suikerindust 1-t 12*, (English translation in *Genetica* 5:97–148; 273–326).

Breseghello F, Sorrells ME, 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. http://dx.doi.org/10.1534/genetics.105.044586

Brumlop S, Finckh MR, 2011. *Applications and potentials of marker assisted selection (MAS) in plant breeding,* Final report of the F+E project "Applications and potentials of smart breeding." Federal Agency for Nature Conservation, Bonn, Germany. www.bfn.de/fileadmin/MDB/documents/service/Skript_298.pdf

Bundock P, Eliott F, Ablett G, Benson A, Casu R, Aitken K, Henry R, 2009. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnology Journal* 7, 347–354. http://dx.doi.org/10.1111/j.1467-7652.2009.00401.x

Butterfield M, 2007. *Marker assisted breeding in sugarcane: A complex polyploid*. PhD diss., University of Stellenbosch, 164 pp.

Campo Zabala RR, Pérez Orama SG, 2010. A new nobilisation system in sugarcane (*Saccharum* spp.). *Proc. Int. Soc. Sugar Cane Technol.* 27, 1–4.

Carlson CS, Eberle MA, Kruglyak L, Nickerson DA, 2004. Mapping complex disease loci in whole-genome association studies. *Nature* 429, 446–452. http://dx.doi.org/10.1038/nature02623

Cheavegatti-Gianotto A, de Abreu H, Arruda P, Bespalhok Filho J, Burnquist W, Creste S, di Ciero L, Ferro J, de Oliveira Figueira A, de Sousa Filgueiras T, Grossi-de-Sá M, Guzzo E, Hoffmann H, de Andrade Landell M, Macedo N, Matsuoka S, de Castro Reinach F, Romano E, da Silva W, de Castro Silva Filho M, César Ulian E, 2011. Sugarcane (*Saccharum* X *officinarum*): A reference study for the regulation of genetically modified cultivars in Brazil. *Tropical Plant Biology* 4, 62–89. http://dx.doi.org/10.1007/s12042-011-9068-3

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL, 2009. Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: A 'gene-to-phenotype' modeling approach. *Genetics* 183, 1507–1523.

Cordeiro GM, Eliott F, McIntyre CL, Casu RE, Henry RJ, 2006. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. *Theoretical and Applied Genetics* 113, 331–343. http://dx.doi.org/10.1007/s00122-006-0300-8

Costet L, Le Cunff L, Royeart S, Raboin L-M, Hervouet C, Toubi L, Telismart H, Garsmeur O, Rousselle Y, Pauquet J, Nibouche S, Glaszmann J-C, Hoarau J-Y, D'Hont A, 2012. Haplotype structure around *Bru*1 reveals a narrow genetic basis for brown rust resistance in modern sugarcane cultivars. *Theoretical and Applied Genetics* 125, 825–836. http://dx.doi.org/10.1007/s00122-012-1875-x

Cuadrado A, Acevedo R, De la Espina SMD, Jouve N, De La Torre C, 2004. Genome remodelling in three modern *S. officinarum*— *S. spontaneum* sugarcane cultivars. *Journal of experimental botany* 55, 847–854. http://dx.doi.org/10.1093/jxb/erh093

Da Silva JA, Bressiani JA, 2005. Sucrose synthase molecular marker associated with sugar content in elite sugarcane progeny. *Genetics and Molecular Biology* 28, 294–298. http://dx.doi.org/10.1590/S1415-47572005000200020

Dal-Bianco M, Carneiro MS, Hotta CT, Chapola RG, Hoffmann HP, Garcia AAF, Souza GM, 2011. Sugarcane improvement: How far can we go? *Current Opinion in Biotechnology* 23, 265–270. http://dx.doi.org/10.1016/j.copbio.2011.09.002

Daniels J, Roach B, 1987. Taxonomy and evolution. Chapter 2, in: Heintz D (ed), *Sugarcane improvement through breeding*. Amsterdam: Elsevier, pp. 7–84.

Daugrois J, Grivet L, Roques D, Hoarau J, Lombard H, Glaszmann J, D'Hont A, 1996. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. *Theoretical and Applied Genetics* 92, 1059–1064. http://dx.doi.org/10.1007/BF00224049

De Araujo LM, Moura FBP, 2011. Bioethanol's dirty footprint in Brazil. *Nature* 469, 299–299. http://dx.doi.org/10.1038/469299d

Devlin B, Roeder K, 1999. Genomic control for association studies. *Biometrics* 55, 997–1004. http://dx.doi.org/10.1111/j.0006-341X.1999.00997.x

D'Hont A (2005) Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenetic and Genome Research* 109:27–33. http://dx.doi.org/10.1159/000082378

D'Hont A, Grivet L, Feldmann P, Glaszmann J, Rao S, Berding N, 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics* 250:405–413. http://dx.doi.org/10.1007/BF02174028

D'Hont A, Paulet F, Glaszmann JC, 2002. Oligoclonal interspecific origin of 'North Indian' and 'Chinese' sugarcanes. *Chromosome Research* 10:253–262. http://dx.doi.org/10.1023/A:1015204424287

D'Hont A, Souza GM, Menossi M, Vincentz M, Van-Sluys MA, Glaszmann JC, Ulian E, 2008. Sugarcane: A major source of sweetness, alcohol, and bio-energy, In: Moore PH, Ming R (eds) *Genomics of Tropical Crop Plants*. New York: Springer, pp. 483–513.

Dias M, da Cunha M, Maciel Filho R, Bonomi A, Jesus C, Rossell C, 2011a. Simulation of integrated first and second generation bioethanol production from sugarcane: Comparison between different biomass pretreatment methods. *Journal of Industrial Microbiology & Biotechnology* 38, 955–966. http://dx.doi.org/10.1007/s10295-010-0867-6

Dias MOS, Junqueira TL, Cavalett Ov, Cunha MP, Jesus CDF, Rossell CEV, Maciel Filho R, Bonomi A, 2011b. Integrated versus stand-alone second generation ethanol production from sugarcane bagasse and trash. *Bioresource technology* 103, 152–161. http://www.sciencedirect.com/science/article/pii/S0960852411014088

Dingkuhn M, Luquet D, Kim HK, Tambour L, Clement-Vidal A, 2006. EcoMeristem, a model of morphogenesis and competition among sinks in rice. 2. Simulating genotype responses to phosphorus deficiency. *Functional Plant Biology* 33, 325–337. http://dx.doi.org/10.1071/FP05267

Dingkuhn M, Luquet D, Quilot B, de Reffye P, 2005. Environmental and genetic control of morphogenesis in crops: Towards models simulating phenotypic plasticity. *Australian Journal of Agricultural Research* 56, 1289–1302. http://dx.doi.org/10.1071/AR05063

Duailibe AK, 2010. Brazil's renewable energy success. *Nature* 468, 1041–1041. http://dx.doi.org/10.1038/4681041b

Edme SJ, Miller JD, Glaz B, Tai PYP, Comstock JC, 2005. Genetic contribution to yield gains in the Florida sugarcane industry across 33 years. *Crop Science* 45, 92–97. http://dx.doi.org/10.2135/cropsci2005.0092

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*. 6(5): e19379 http://dx.doi.org/10.1371/journal.pone.0019379

FAOSTAT, 2012. http://faostat.fao.org/default.aspx

Flint-Garcia SA, Thornsberry JM, Buckler IV ES, 2003. Structure of linkage disequilibrium in plant. *Annual Review of Plant Biology* 54, 357–374. http://dx.doi.org/10.1146/annurev.arplant.54.031902.134907

Garoma T, Ben-Khaled M, Beyene A, 2011. Comparative resource analyses for ethanol produced from corn and sugarcane in different climatic zones. *International Journal of Energy Research* 36, 1065–1076. http://dx.doi.org/10.1002/er.1853

Glaszmann J, Dufour P, Grivet L, D'Hont A, Deu M, Paulet F, Hamon P, 1997. Comparative genome analysis between several tropical grasses. *Euphytica* 96, 13–21. http://dx.doi.org/10.1023/A:1002987620250

Glynn NC, Nipper RW, Comstock JC, 2011 Application of RAD LongRead sequencing for SNP discovery in sugarcane, *Plant Genome Conference Proceedings*, San Diego, p. 184.

Gravois K, Milligan S, 1992. Genetic relationships between fiber and sugarcane yield components. *Crop Science* 32, 62–67. http://dx.doi.org/10.2135/cropsci1992.0011183×003200010014x

Grivet, L, Arruda P. 2002. Sugarcane genomics: Depicting the complex genome of an important tropical crop. *Current Opinion in Plant Biology* 5(2):122–127.

Grivet L, d'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC, 1996. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): Genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* 142, 987–1000.

Grivet L, Daniels C, Glaszmann JC, D'Hont A, 2004. A review of recent molecular genetics evidence for sugarcane evolution and domestication. *Ethnobotany Research & Applications* 2, 9–17. www.ethnobotanyjournal.org/vol2/i1547–3465-02-009.pdf

Grivet L, Glaszmann J, D'Hont A, 2006. Molecular evidence of sugarcane evolution and domestication, in: Motley T, Zerega N, Cross H (eds), *Darwin's harvest: New approaches to the origins, evolution, and conservation of crops*. New York: Columbia University Press, pp. 49–66.

Grivet L, Glaszmann JC, Vincentz M, Da Silva F, Arruda P, 2003. ESTs as a source for sequence polymorphism discovery in sugarcane: example of the *Adh* genes. *Theoretical and Applied Genetics* 106, 190–197. http://dx.doi.org/10.1007/s00122-002-1075-1

Hammer G, Chapman S, Van Oosterom E, Podlich D, 2004. Trait physiology and crop modelling to link phenotypic complexity to underlying genetic systems. *Proceedings of the 4th International Crop Science Congress "New directions for a diverse planet" 26 Sep – 1 Oct 2004, Brisbane, Australia*, pp. 1–15.

Hammer GL, Kropff MJ, Sinclair TR, Porter JR, 2002. Future contributions of crop modelling—from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. *European Journal of Agronomy* 18, 15–31. http://dx.doi.org/10.1016/S1161-0301(02)00093-X

Hardy O, Vekemans X, 2002. SPAGEDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2, 618–620. http://dx.doi.org/10.1046/j.1471-8286.2002.00305.x

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME, 2010. Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science* 50, 1681–1690. http://dx.doi.org/10.2135/cropsci2009.11.0662

Heffner EL, Sorrells ME, Jannink JL, 2009. Genomic selection for crop improvement. *Crop Science* 49, 1–12. http://dx.doi.org/10.2135/cropsci2008.08.0512

Heller-Uszynska K, Uszynski G, Huttner E, Evers M, Carlig J, Caig V, Aitken K, Jackson P, Piperidis G, Cox M, Gilmour R, D'Hont A, Butterfield M, Glaszmann J-C, Kilian A, 2010. Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Molecular Breeding* 1, 37–55. http://dx.doi.org/10.1007/s11032-010-9460-y

Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L, 2001. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. *Theoretical and Applied Genetics* 103, 84–97. http://dx.doi.org/10.1007/s001220000390

Hoarau JY, Grivet L, Offmann B, Raboin LM, Diorflar JP, Payet J, Hellmann M, D'Hont AD, Glaszmann JC, 2002. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.).II. Detection of QTLs for yield components. *Theoretical and Applied Genetics* 105, 1027–1037. http://dx.doi.org/10.1007/s00122-002-1047-5

Hogarth DM, 1977. Quantitative inheritance studies in sugarcane. III. The effect of competition and violation of genetic assumptions on estimation of genetic variance components. *Australian Journal of Agriculture Research* 28, 257–268.

Jackson P, 1994. Genetic relationships between attributes in sugarcane clones closely related to *Saccharum spontaneum*. *Euphytica* 79, 101–108. http://dx.doi.org/10.1007/BF00023581

Jackson P, Hogarth D, 1992. Genotype x environment interactions in sugarcane, 1. Patterns of response across sites and crop-years in north Queensland. *Australian Journal of Agricultural Research* 43, 1447–1459. http://dx.doi.org/10.1071/AR9921447

Jackson PA, 2005. Breeding for improved sugar content in sugarcane. *Field Crops Research* 92, 277–290. http://dx.doi.org/10.1016/j.fcr.2005.01.024

Jannink JL, Lorenz AJ, Iwata H, 2010. Genomic selection in plant breeding: From theory to practice.

Briefings in Functional Genomics 9, 166–177. http://dx.doi.org/10.1093/bfgp/elq001

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A, 2007. Orthologous comparison in a gene rich region among grasses reveals stability in the sugarcane polyploid genome. *The Plant Journal* 50, 574–585.

Jannoo N, Grivet L, David J, D'Hont A, Glaszmann J, 2004. Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. *Heredity* 93, 460–467. http://dx.doi.org/10.1038/sj.hdy.6800524

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC, 1999. Linkage disequilibrium among modern sugarcane cultivars. *Theoretical and Applied Genetics* 99, 1053–1060. http://dx.doi.org/10.1007/s001220051414

Jeswiet J, 1930. The development of selection and breeding of the sugar cane in Java, *Proc. Int. Soc. of Sugarcane Technol*, pp. 44–57.

Jordan DR, Casu RE, Besse P, Carroll BC, Berding N, McIntyre CL, 2004. Markers associated with stalk number and suckering in sugarcane colocate with tillering and rhizomatousness QTLs in sorghum. *Genome* 47, 988–993. http://dx.doi.org/10.1139/g04-040

Kang M, Miller J, 1984. Genotype X Environment interactions for cane and sugar yield and their implications in sugarcane breeding. *Crop Science* 24, 435–440. http://dx.doi.org/10.2135/cropsci1984.0011183×002400030002x.

Kang M, Sosa O, Miller J, 1989. Path analysis for percent fiber and cane and sugar yield in sugarcane. *Crop Science* 29, 1481–1483. http://dx.doi.org/10.2135/cropsci1989.0011183×002900060032x.

Kang M, Tai P, Miller J, 1983. Genetic and phenotypic path analyses and heritability in sugarcane. *Crop Science* 23, 643–647.

Keating B, Robertson M, Muchow R, Huth N, 1999. Modelling sugarcane production systems I. Development and performance of the sugarcane module. *Field Crops Research* 61, 253-271. http://dx.doi.org/10.1016/S0378-4290(98)00167-1

Keating BA, Carberry P, Hammer G, Probert ME, Robertson M, Holzworth D, Huth N, Hargreaves J, Meinke H, Hochman Z, 2003. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* 18, 267–288. http://dx.doi.org/10.1016/S1161-0301(02)00108-9

Lander E, Kruglyak L, 1995. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nature Genetics* 11, 241–247. http://dx.doi.org/10.1038/ng1195-241

Le Cunff L, Garsmeur O, Raboin LM, Pauquet J, Telismart H, Selvi A, Grivet L, Philippe R, Begum D, Deu M, Costet L, Wing R, Glaszmann JC, D'Hont A, 2008. Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (*Bru*1) in highly polyploid sugarcane (2n ∼ 12x ∼ 115). *Genetics* 180, 649–660. http://www.genetics.org/cgi/content/abstract/180/1/649

Lingle SE, Johnson RM, Tew TL, Viator RP, 2010. Changes in juice quality and sugarcane yield with recurrent selection for sucrose. *Field Crops Research* 118, 152–157. http://dx.doi.org/10.1016/j.fcr.2010.05.002

Lingle SE, Viator RP, Johnson RM, Tew TL, Boykin DL, 2009. Recurrent selection for sucrose content has altered growth and sugar accumulation in sugarcane. *Field Crops Research* 113, 306–311. http://dx.doi.org/10.1016/j.fcr.2009.06.015

Luquet D, Dingkuhn M, Kim HK, Tambour L, Clement-Vidal A, 2006. EcoMeristem, a model of morphogenesis and competition among sinks in rice. 1. Concept, validation and sensitivity analysis. *Functional Plant Biology* 33, 309–323. http://dx.doi.org/10.1071/FP05266

Luquet D, Song YH, Elbelt S, This D, Clément-Vidal A, Périn C, Fabre D, Dingkuhn M, 2007. Model-assisted physiological analysis of Phyllo, a rice architectural mutant. *Functional Plant Biology* 34, 11–23. http://dx.doi.org/10.1071/FP06180

Luquet D, Soulié JC, Huet C, Martiné JF, 2010. Modelling plant morphogenesis and source-sink processes to support crop performance phenotyping: Application of Ecomeristem model to sugar cane. In: Wery, J., Shili-Touzi, I., Perrin, A. (eds), *Agro 2010: The XIth ESA Congress*, Montpellier, France. Montpellier: Agropolis International, pp. 579–580.

Machado G, 2001. Sugarcane variety notes, An international directory, 7th. Revision, Piracicaba (Brazil), p. 132.

Manners JM, 2011. Chapter 3, Functional genomics of sugarcane. In: Kader J-C, Delseny M (eds), *Advances in Botanical Research*. New York: Academic Press, pp. 89–168. http://dx.doi.org/10.1016/B978-0-12-385851-1.00003-2

Martiné JF, 2003. Modélisation de la production potentielle de la canne à sucre en zone tropicale, sous conditions thermiques et hydriques contrastées. Applications du modèle. *PhD diss., Institut national agronomique, Paris*, 131.

Martiné JF, 2007. Analysis and forecasting of the sucrose content of sugarcane crops during the harvest period in Reunion Island. *Proceeding of the International Society Sugar Cane Technologists* 27, 607–612.

Martiné JF, Gozé E, Luquet D, Thong-Chane A, Houles X, Soulié JC, Rouan L, Nibouche S, 2010. Exploring the feasibility of sugarcane phenotyping using crop models with contrasted climatic conditions in Réunion island: [Poster Abstract], *Agro2010 the XI ESA Congress, Montpellier, 29-09-2010*, Montpellier, France, pp. 577–578.

Martiné J, Lebret P, Bonhomme R, 2001. Effect of soil water content, intercepted radiation and temperature on sugar cane sucrose balance and partition. *Proceeding of the Australian Society of Sugar Cane Technologists, XXIV Congress, Brisbane, Australia, 17–21 September 2001* 2, 58–64.

Martiné JF, Siband P, Bonhomme R, 2000. Simulation of the maximum yield of sugar cane at different altitudes:

Effect of temperature on the conversion of radiation into biomass. *Agronomie* 19, 3–12.

Matsuoka S, Ferro J, Arruda P, 2009. The Brazilian experience of sugarcane ethanol industry. *In Vitro Cellular & Developmental Biology – Plant* 45, 372–381. http://dx.doi.org/10.1007/s11627-009-9220-z

McIntyre CL, Jackson M, Cordeiro GM, Amouyal O, Hermann S, Aitken K, Eliott F, Henry R, Casu R, Bonnett G, 2006. The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. *Molecular Breeding* 18, 39–50. http://dx.doi.org/10.1007/s11032-006-9012-7

Meuwissen T, Hayes B, Goddard M, 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Ming R, Del Monte TA, Hernandez E, Moore PH, Irvine JE, Paterson AH, 2002a. Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. *Genome* 45, 794–803. http://dx.doi.org/10.1139/g02-042

Ming R, Liu S-C, Lin Y-R, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH, 1998. Detailed Alignment of Saccharum and Sorghum Chromosomes: Comparative Organization of Closely Related Diploid and Polyploid Genomes. *Genetics* 150, 1663–1682. http://www.genetics.org/content/150/4/1663.abstract

Ming R, Liu S-C, Moore PH, Irvine JE, Paterson AH, 2001. QTL analysis in a complex autopolyploid: Genetic control of sugar content in sugarcane. *Genome Research* 11, 2075–2084. http://dx.doi.org/10.1101/gr.198801

Ming R, Wang Y, Draye X, Moore P, Irvine J, Paterson A, 2002b. Molecular dissection of complex traits in autopolyploids: Mapping QTLs affecting sugar yield and related traits in sugarcane. *Theoretical and Applied Genetics* 105, 332–345. http://dx.doi.org/10.1007/s00122-001-0861-5

Morgante M, Salamini F, 2003. From plant genomics to breeding practice. *Current Opinion in Biotechnology* 14, 214–219. http://dx.doi.org/10.1016/S0958-1669(03)00028-4

Nibouche S, Martiné JF, Luquet D, Gozé E, Rouan L, Costet L, D'Hont A, Soulié JC, Thong-Chane A, 2010. The DELICAS projet: Model assisted phenotyping in sugarcane for the identification of marker-trait associations. In: Wery, J., Shili-Touzi, I., Perrin, A. (eds), *Agro 2010: The XIth ESA Congress*, Montpellier, France. Montpellier: Agropolis International, pp. 541–542.

Nordborg M, Tavaré S, 2002. Linkage disequilibrium: What history has to tell us. *Trends in Genetics* 18, 83–90. http://dx.doi.org/10.1016/S0168-9525(02)02557-X

O'Leary GJ, 2000. A review of three sugarcane simulation models with respect to their prediction of sucrose yield. *Field Crops Research* 68, 97–111. http://dx.doi.org/10.1016/S0378-4290(00)00112-X

Pastina M, Malosetti M, Gazaffi R, Mollinari M, Margarido GRA, Oliveira K, Pinto L, Souza A, van Eeuwijk F, Garcia AAF, 2012. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theoretical and Applied Genetics* 124, 835–849. http://dx.doi.org/10.1007/s00122-011-1748-8

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556. http://dx.doi.org/10.1038/nature07723

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL, 1995. The weediness of wild plants: Molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. *Proceedings of the National Academy of Sciences* 92, 6127.

Pinto L, Garcia AAF, Pastina M, Teixeira LHM, Bressiani J, Ulian E, Bidoia MAP, Souza A, 2010. Analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (*Saccharum* spp.) commercial cross. *Euphytica* 172, 313–327. http://dx.doi.org/10.1007/s10681-009-9988-2

Piperidis G, D'Hont A, 2001. Chromosome composition analysis of various *Saccharum* interspecific hybrids by genomic in situ hybridisation (GISH). *Australian Society of Sugar Cane Technologists* 24, 565–566.

Piperidis G, Piperidis N, D'Hont A, 2010. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics* 284, 65–73. http://dx.doi.org/10.1007/s00438-010-0546-3

Piperidis N, Jackson P, D'Hont A, Besse P, Hoarau J-Y, Courtois B, Aitken K, McIntyre C, 2008. Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. *Molecular Breeding* 21, 233–247. http://dx.doi.org/10.1007/s11032-007-9124-8

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D, 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38, 904–909. http://dx.doi.org/10.1038/ng1847

Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. http://www.genetics.org/cgi/content/abstract/155/2/945

Prudent M, Lecomte A, Bouchet JP, Bertin N, Causse M, Génard M, 2011. Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. *Journal of Experimental Botany* 62, 907–919. http://dx.doi.org/10.1093/jxb/erq318

Quilot B, Génard M, Kervella J, Lescourret F, 2004. Analysis of genotypic variation in fruit flesh total sugar content via an ecophysiological model applied to peach. *Theoretical and Applied Genetics* 109, 440–449. http://dx.doi.org/10.1007/s00122-004-1651-7

Quilot B, Kervella J, Génard M, Lescourret F, 2005. Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. *Journal of experimental botany* 56, 3083–3092. http://dx.doi.org/10.1093/jxb/eri305

Raboin L-M, Pauquet J, Butterfield M, D'Hont A, Glaszmann J-C, 2008. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theoretical and Applied Genetics* 116, 701–714. http://dx.doi.org/10.1007/s00122-007-0703-1

Raboin L, Oliveira K, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau J, D'Hont A, 2006. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: Identification of a gene controlling stalk colour and a new rust resistance gene. *Theoretical and Applied Genetics* 112, 1382–1391. http://dx.doi.org/10.1007/s00122-006-0240-3

Raboin LM, 2005. Génétique de la résistance au charbon de la canne à sucre causé par *Ustilago scitaminea*: Caractérisation de la diversité génétique du pathogène, cartographie de QTL dans un croisement bi-parental et étude d'associations dans une population de cultivars modernes. ENSAM, Montpellier, France, p. 119.

Rafalski A, Morgante M, 2004. Corn and humans: Recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* 20, 103–111. http://dx.doi.org/10.1016/j.tig.2003.12.002

Reffay N, Jackson PA, Aitken KS, Hoarau JY, D'Hont A, Besse P, McIntyre CL, 2005. Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. *Molecular Breeding* 15, 367–381. http://dx.doi.org/10.1007/s11032-004-7981-y

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F, 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiology* 131, 664–675. http://dx.doi.org/10.1104/pp.013839

Roach B, 1989. Origin and improvement of the genetic base of sugarcane. *Proceeding of Australian Society of Sugar Cane Technologists* 11, 34–47.

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D'Hont A, 2003. Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Molecular Genetics and Genomics* 269, 406–419. http://dx.doi.org/10.1007/s00438-003-0849-8

Sills GR, Bridges W, Al-Janabi SM, Sobral BWS, 1995. Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes & Jesw. ex Grassl). *Molecular Breeding* 1, 355–363. http://dx.doi.org/10.1007/BF01248413

Simmonds NW, 1979. The impact of plant breeding on sugarcane yields in Barbados. *Tropical Agriculture, Trinidad* 56, 289–300.

Souza GM, Berges H, Bocs S, Casu R, D'Hont A, Ferreira JE, Henry R, Ming R, Potier B, Van Sluys MA, 2011. The sugarcane genome challenge: Strategies for sequencing a highly complex genome. *Tropical Plant Biology* 4, 145–156. http://dx.doi.org/10.1007/s12042-011-9079-0

Tew TL, Cobill RM, 2008. Genetic Improvement of Sugarcane (Saccharum spp.) as an energy crop. In: Vermerris W (ed) *Genetic improvement of bioenergy crops*. New York: Springer, pp. 273–294. http://dx.doi.org/10.1007/978-0-387-70805-8_9

Vettore AL, Da Silva FR, Kemper EL, Souza GM, Da Silva AM, Ferro MIT, Henrique-Silva F, Giglioti AA, Lemos MVF, Coutinho LL, 2003. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Research* 13, 2725–2735. http://dx.doi.org/10.1101/gr.1532103

Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM, 2010. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. *Plant Biotechnology Journal* 8, 263–276. http://dx.doi.org/10.1111/j.1467-7652.2009.00491.x

Walker DA, 2009. Biofuels, facts, fantasy, and feasibility. *Journal of Applied Phycology* 21, 509–517. http://dx.doi.org/10.1007/s10811-009-9446-5

Wang J, Roe B, Macmil S, Yu Q, Murray J, Tang H, Chen C, Najar F, Wiley G, Bowers J, Van Sluys M-A, Rokhsar D, Hudson M, Moose S, Paterson A, Ming R, 2010. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* 11, 261. http://www.biomedcentral.com/1471–2164/11/261

Wei X, Jackson P, McIntyre C, Aitken K, Croft B, 2006. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theoretical and Applied Genetics* 114, 155–164. http://dx.doi.org/10.1007/s00122-006-0418-8

Wei X, Jackson PA, Hermann S, Kilian A, Heller-Uszynska K, Deomano E, 2010. Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome* 53, 973–981. http://dx.doi.org/10.1139/G10-050

Wong C, Bernardo R, 2008. Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. *Theoretical and Applied Genetics* 116, 815–824. http://dx.doi.org/10.1007/s00122-008-0715-5

Wu K, Burnquist W, Sorrells M, Tew T, Moore P, Tanksley S, 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* 83, 294–300. http://dx.doi.org/10.1007/BF00224274

Xu Y, Crouch JH, 2008. Marker-assisted selection in plant breeding: From publications to practice. *Crop Science* 48, 391–407. http://dx.doi.org/10.2135/cropsci2007.04.0191.

Yin X, Kropff MJ, Stam P, 1999. The role of ecophysiological models in QTL analysis: The example of

specific leaf area in barley. *Heredity* 82, 415–421. http://dx.doi.org/10.1038/sj.hdy.6885030

Yin X, Stam P, Kropff MJ, Schapendonk AHCM, 2003. Crop Mmdeling, QTL mapping, and their complementary role in plant breeding. *Agronomy Journal* 95, 90–98. http://dx.doi.org/10.2134/agronj2003.9000

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES, 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38, 203–208. http://dx.doi.org/10.1038/ng1702

Appendix I – Contributors

Koh Aoki
Kazusa DNA Research Institute
Japan

Osaka Prefecture University
Japan

Raman Babu
International Maize and Wheat Improvement Center (CIMMYT)
Mexico

Julia Bailey-Serres
Department of Botany and Plant Sciences
Center for Plant Cell Biology
University of California
United States

J. H. Chin
International Rice Research Institute (IRRI)
Philippines

Bertrand C. Y. Collard
International Rice Research Institute (IRRI)
Philippines

L. Costet
CIRAD
France

Santosh P. Deshpande
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
India

Shalabh Dixit
International Rice Research Institute (IRRI)
Philippines

Majid R. Foolad
Department of Plant Science and The Intercollege Graduate Degree Programs in Plant Biology and Genetics
The Pennsylvania State University
United States

Enrico Francia
Department of Life Sciences
University of Modena and Reggio Emilia
Italy

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

Masanobu Fukami
Chiba Prefectural Agriculture and Forestry Research Center
Japan

Galiba Gabor
Agricultural Institute
Agricultural Research Center
Hungarian Academy of Sciences
Hungary

Research Institute of Chemical and Process Engineering
Faculty of Information Technology
University of Pannonia
Hungary

R. Gamuyao
International Rice Research Institute (IRRI)
Philippines

M. Gouy
eRcane
CIRAD
UMR PVBMT
Université de la Réunion
UMR PVBMT
France

Claudia T. Guimarães
Embrapa Maize and Sorghum
Brazil

S. M. Haefele
International Rice Research Institute (IRRI)
Philippines

Makoto Hasegawa
Chiba Prefectural Agriculture and Forestry Research Center
Japan

C. Tom Hash
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
India

Amelia Henry
International Rice Research Institute (IRRI)
Philippines

Sigrid Heuer
International Rice Research Institute (IRRI)
Philippines

J. Y. Hoarau
CIRAD
UMR AGAP
France

Kiyosumi Hori
National Institute of Agrobiological Sciences
Japan

Abdelbagi M. Ismail
International Rice Research Institute (IRRI)
Philippines

Sachiko Isobe
Kazusa DNA Research Institute
Japan

Jana Kholova
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
India

Matthew P. Kinkade
Campbell Soup Company
Vegetable R&D – Campbell's Seeds
United States

Hiroyuki Kiyoshima
Chiba Prefectural Agriculture and Forestry Research Center
Japan

Leon V. Kochian
Robert W. Holley Center for Agriculture and Health
USDA – ARS
Cornell University
United States

Gabor Kocsy
Agricultural Institute
Agricultural Research Center
Hungarian Academy of Sciences
Hungary

Padmalatha Koilkonda
Kazusa DNA Research Institute
Japan

Arvind Kumar
International Rice Research Institute (IRRI)
Philippines

Atsushi Kurabayashi
Kazusa DNA Research Institute
Japan

Chikara Kuwata
Chiba Prefectural Agriculture and Forestry Research Center
Japan

David J. Mackill
International Rice Research Institute (IRRI)
Philippines

MARS Food Global
Dept. of Plant Science
University of California
United States

Jurandir V. Magalhaes
Embrapa Maize and Sorghum
Brazil

Lyza G. Maron
Robert W. Holley Center for Agriculture and Health
USDA – ARS
Cornell University
United States

Justyna Milc
Department of Life Sciences
University of Modena and Reggio Emilia
Italy

S. Nibouche
CIRAD
UMR PVBMT
France

Natalia Palacios
International Maize and Wheat Improvement Center (CIMMYT)
Mexico

Nicola Pecchioni
Department of Life Sciences
University of Modena and Reggio Emilia
Italy

Miguel A. Piñeros
Embrapa Maize and Sorghum
Brazil

J. Damien Platten
International Rice Research Institute (IRRI)
Philippines

B. M. Prasanna
International Maize and Wheat Improvement Center (CIMMYT)
Kenya

Punna Ramu
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
India

Shigemi Sasamoto
Kazusa DNA Research Institute
Japan

Endang M. Septiningsih
International Rice Research Institute (IRRI)
Philippines

Kenta Shirasawa
Kazusa DNA Research Institute
Japan

Eric J Stockinger
Department of Horticulture and Crop Science
The Ohio State University/Ohio Agricultural Research and Development Center
United States

Shigeru Suzuki
Chiba Prefectural Agriculture and Forestry Research Center
Japan

Satoshi Tabata
Kazusa DNA Research Institute
Japan

Michael J. Thomson
International Rice Research Institute (IRRI)
Philippines

Hisano Tsuruoka
Kazusa DNA Research Institute
Japan

Roberto Tuberosa
Department of Agricultural Sciences
University of Bologna
Italy

Vincent Vadez
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
India

Rajeev K. Varshney
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
India

Tsuyuko Wada
Kazusa DNA Research Institute
Japan

M. Wissuwa
Crop, Environment and Livestock Division
Japan International Center for Agricultural Sciences (JIRCAS)
Japan

Masahiro Yano
National Institute of Agrobiological Sciences
Japan

Appendix II – Reviewers

Nourollah Ahmadi
CIRAD
Montpellier
France

Neeraja Naga Chirravori
Directorate of Rice Research
Hyderabad
India

Nick Collins
University of Adelaide
Glen Osmond
Australia

P. Janila
ICRISAT
Hyderabad
India

L. Krishnamurthy
ICRISAT
Hyderabad
India

Sanjeet Kumar
AVRDC
Shanhua
Taiwan

Sylvia Moraes de Sousa
Embrapa Maize and Sorghum
Sete Lagoas
Brazil

Peggy Ozias-Akins
University of Georgia
Tifton
USA

Prathima PT
Sugarcane Breeding Institute
Coimbatore
India

Sujay Rakshit
Directorate of Sorghum Research
Hyderabad
India

S. Robin
TNAU
Coimbatore
India

R. M. Sundaram
Directorate of Rice Research
Hyderabad
India

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

Index

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

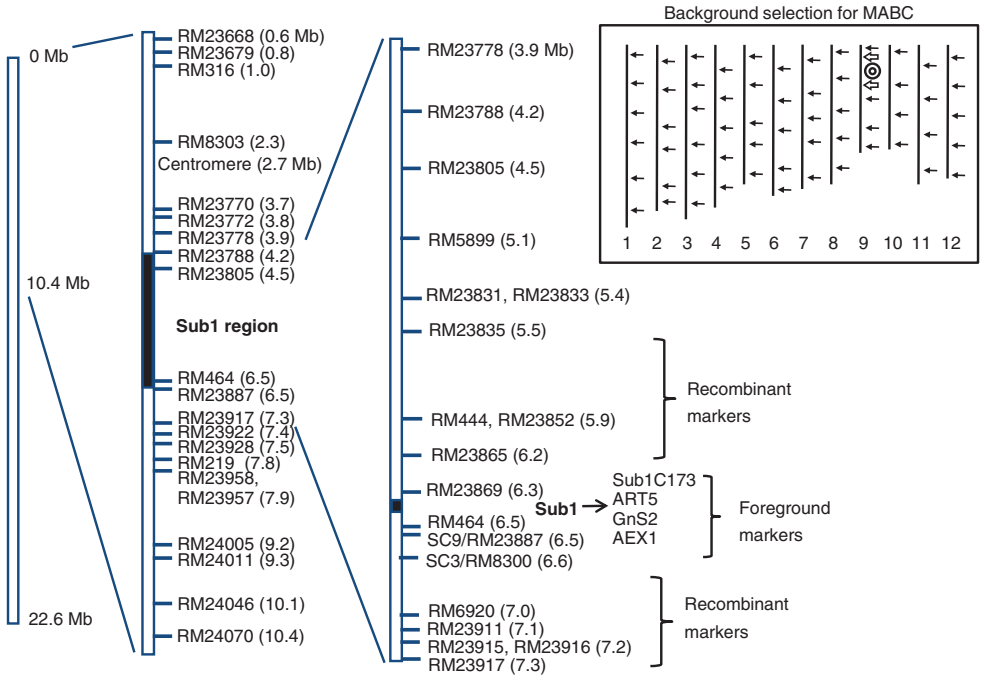


Plate 2.1. Marker-assisted backcrossing (MABC) for *SUB1*. A number of markers have been identified as useful for foreground markers (used to retain the tolerant *SUB1* allele) and recombinant markers (flanking markers used to select for a small *SUB1* introgression). Once foreground and recombinant selection have reduced the population size, background selection is performed to eliminate donor introgressions across the rest of the genome and return to the recurrent parent genome (see inset).

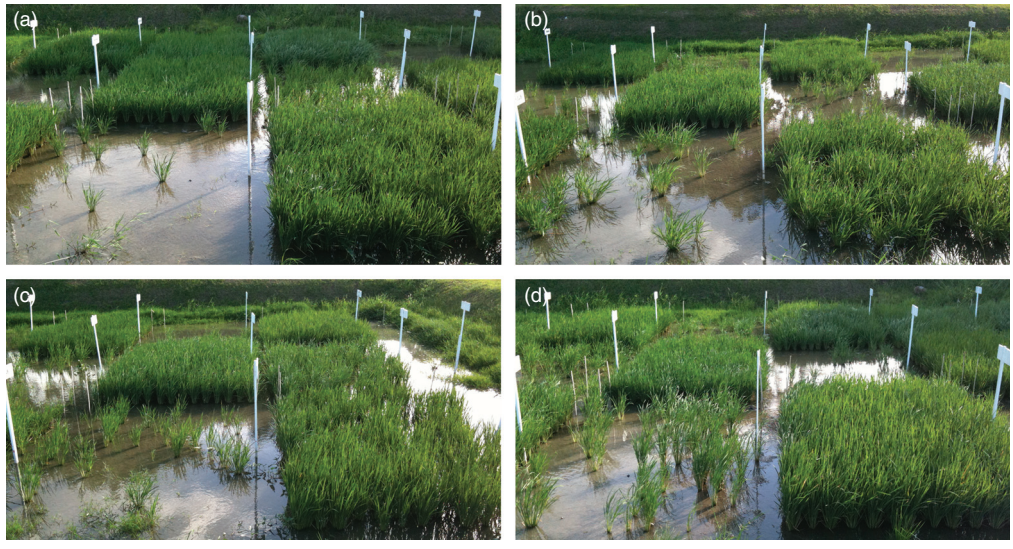


Plate 2.2. Photographs of Sub1 variety compared with the original variety (foreground) at 2 months after 16 days of submergence in an IRRI field trial. (a) Swarna vs Swarna-Sub1; (b) BR11 vs BR11-Sub1; (c) Samba Mahsuri vs Samba Mahsuri-Sub1; and (d) Ciherang vs Ciherang-Sub1.

Translational Genomics for Crop Breeding, Volume II: Abiotic Stress, Yield and Quality.
Edited by Rajeev K. Varshney and Roberto Tuberosa.

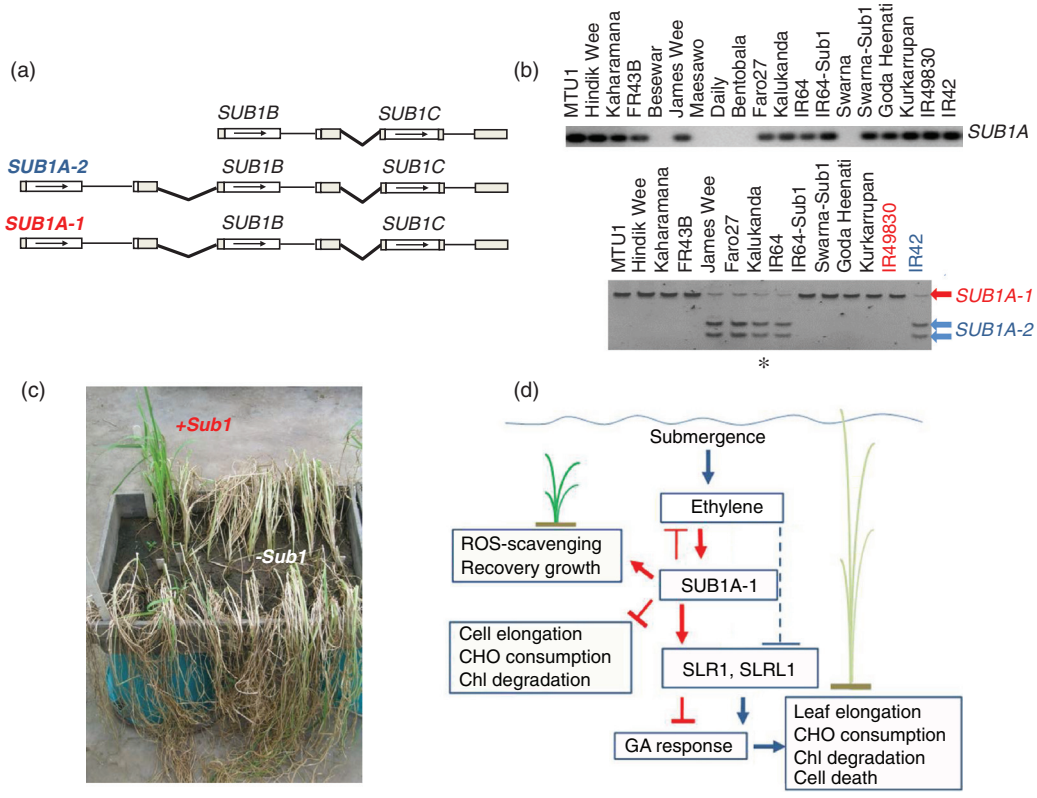


Plate 2.3. ***SUB1*-specific markers and the molecular basis of submergence tolerance.** The *SUB1* locus contains either two (*SUB1B*, *SUB1C*) or three (*SUB1A*, *SUB1B*, *SUB1C*) ERF genes (**a**). Two different indica-/aus-specific *SUB1A* alleles are known, of which the *SUB1A-1* allele is present in submergence-tolerant varieties. Molecular markers distinguish between varieties with and without *SUB1A* (**b**, top panel) and between the *SUB1A-1* and *SUB1A-2* allele (**b**, bottom panel). The visible effect of the *SUB1* locus is the suppression of growth during submergence and plant recovery within about 2 weeks after de-submergence (**c**). Under submergence, ethylene induces the gibberellic acid (GA)-dependent escape response, which is suppressed by *SUB1A-1*-mediated maintenance of GA repression via the GA repressor proteins SLR1 and SLRL1 (**d**). ROS, reactive oxygen species; CHO, carbohydrate, CHL, chlorophyll, GA, gibberellic acid, SLR1, SLENDER RICE GRAIN 1, SLRL1, SLENDER RICE LIKE 1.

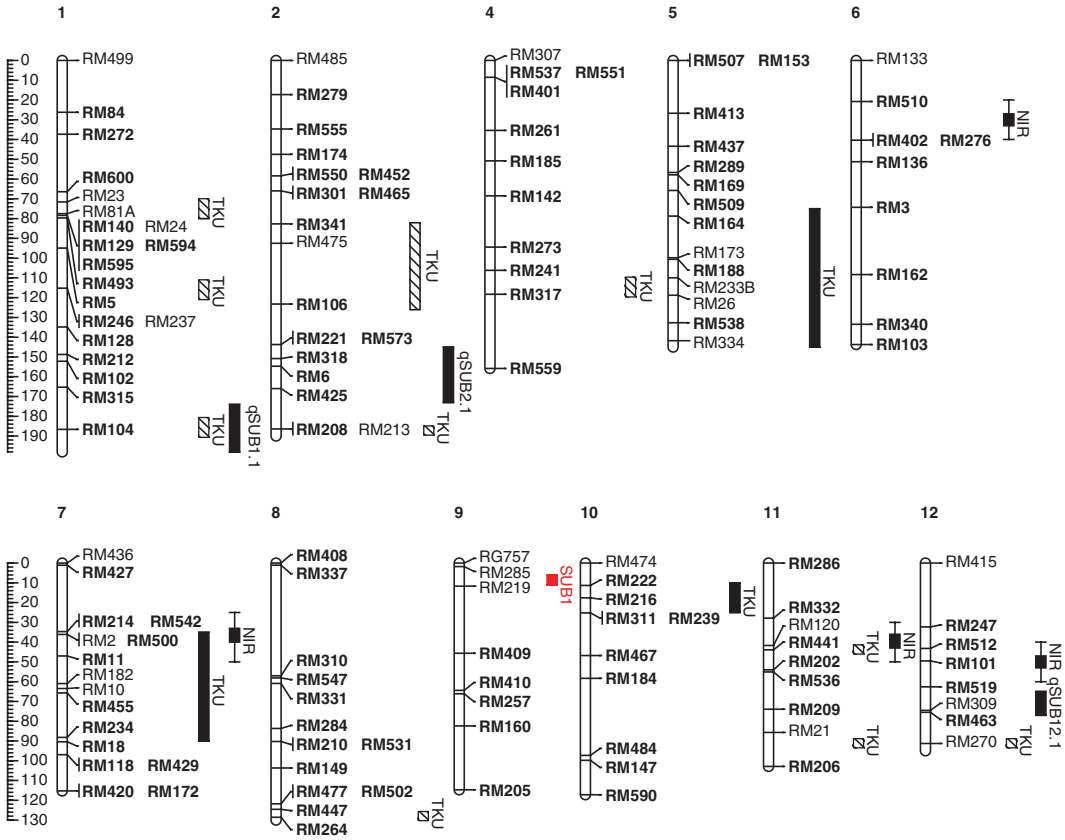


Plate 2.4. Consensus QTL map for submergence tolerance. Linkage maps are derived from "anchor" SSR markers (i.e., common to both maps) from the "Cornell SSR 2001" and "CIAT SSR 2006" (www.gramene.org; Temnykh et al., 2000; Orjuela et al., 2010). Chromosome 3 is not shown because no QTLs have been detected on this chromosome. Anchor SSRs are indicated in bold and larger font; other markers are from the Cornell SSR 2001 map only, using genetic map distances. *SUB1* is shown as a red QTL bar on chromosome 9. Approximate QTL positions are represented from Toojinda et al. (2003; "TKU"), Nandi et al. (1997; "NIR"), and Septiningsih et al. (2012). Filled QTL bars indicate QTLs that were detected in multiple populations and/or component traits, whereas hatched QTL bars indicate minor QTLs detected in only single populations or using a single component trait. Box plot QTL bars are indicated for QTLs from Nandi et al. (1997). QTL positions could only be estimated due to the low number of SSRs used in this study. Possible QTLs overlapping between studies were detected on chromosomes 7 and 11. The figure was produced using MapChart 2.1 (Voorrips, 2002).

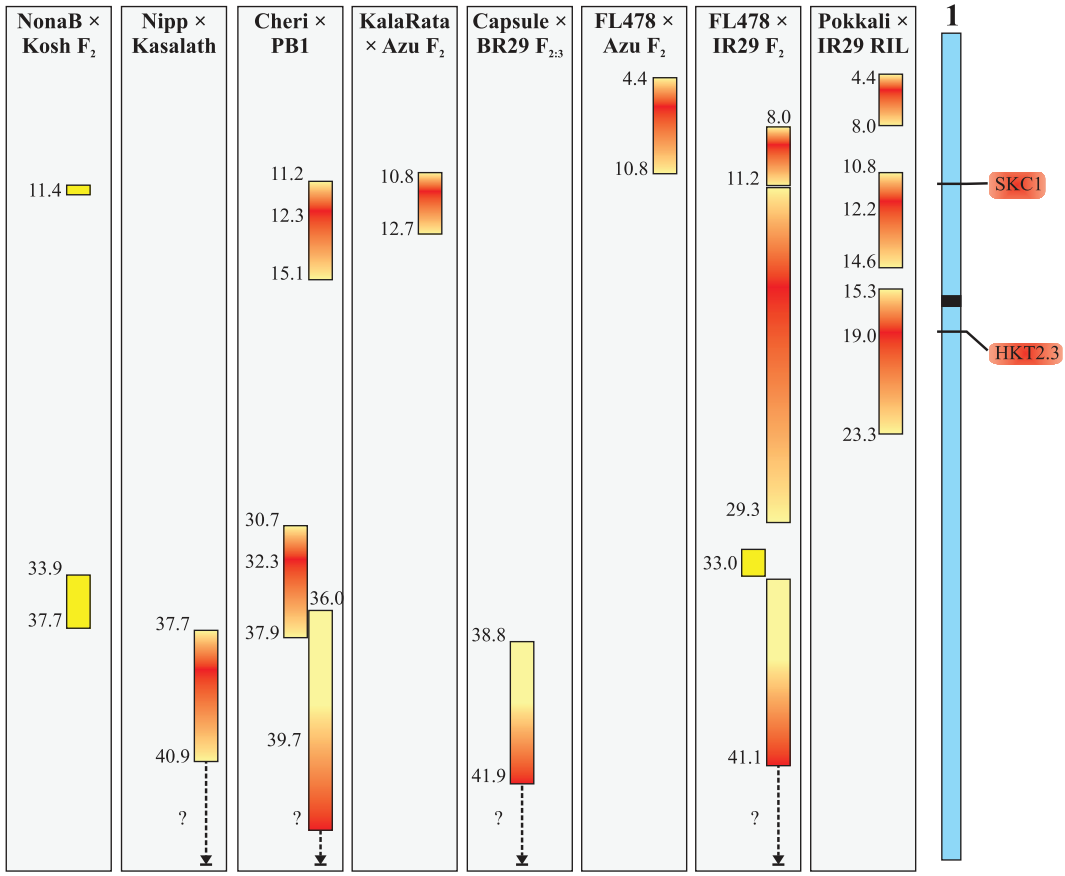


Plate 3.1. Comparison of QTLs on (a) chromosome 1 and (b) chromosome 3 identified in various mapping populations.

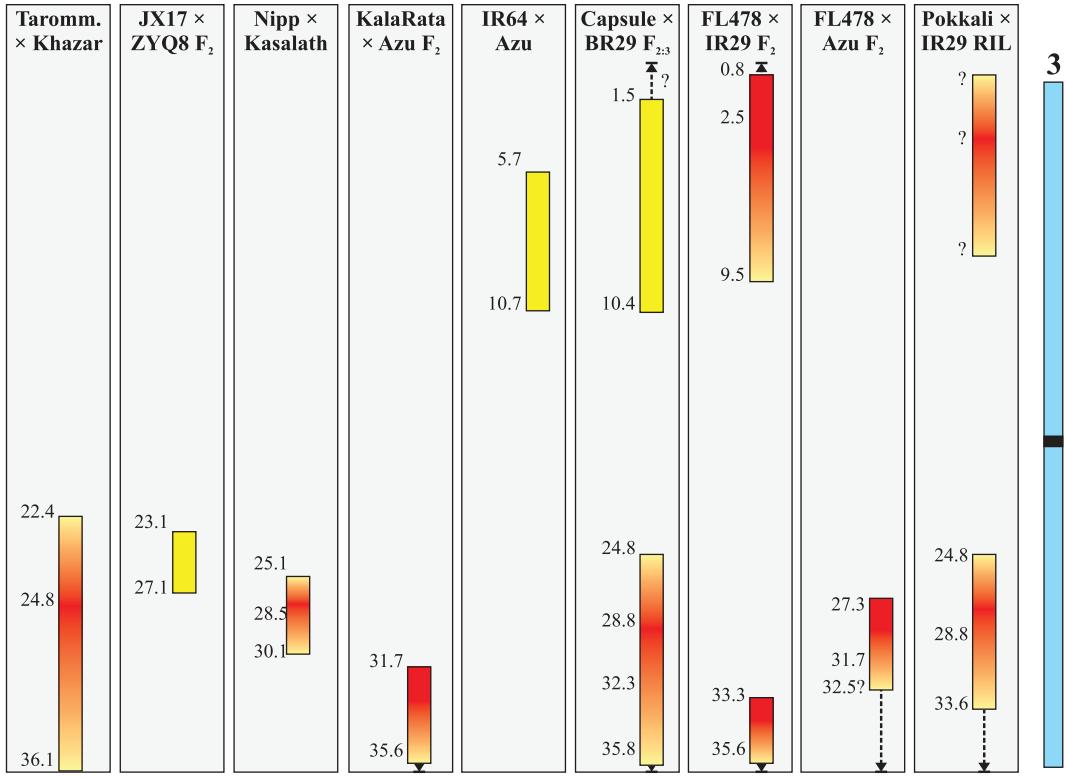


Plate 3.1. (*Continued*)

Plate 4.3. A semi-automated instrument rack for monitoring canopy temperature and NDVI of +QTL and –QTL lines under drought in the rainout shelter at IRRI.

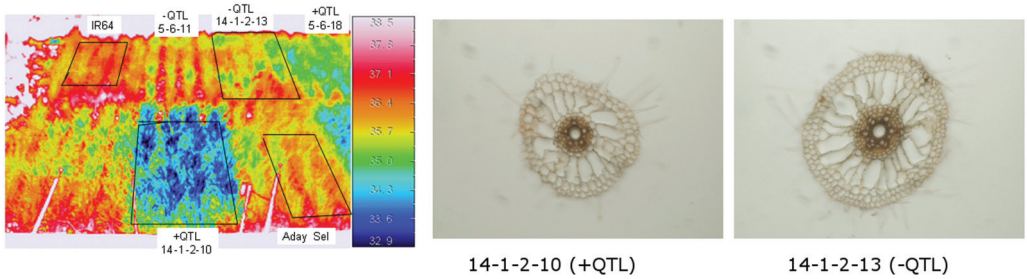


Plate 4.4. The IR64 x Aday Sel NILs showed large differences in canopy temperature under severe drought and smaller root and xylem vessel diameters.

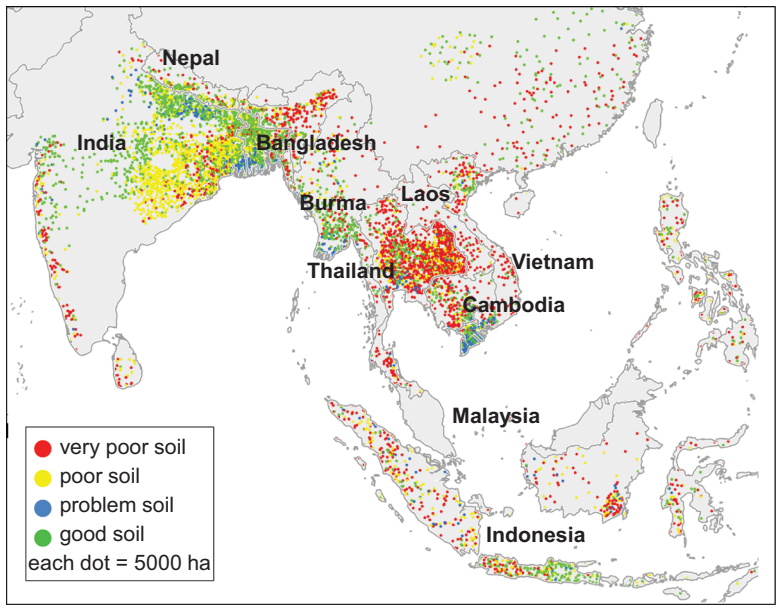


Plate 5.1. Soil quality in rain-fed lowland rice systems. Soils in rain-fed (intermediate and shallow) rice systems in Asia are often constrained by abiotic stresses and nutrient deficiency (Haefele and Hijmans 2007; Haefele and Hijmans 2009). Note: Rain-fed upland rice areas are not included in this map.

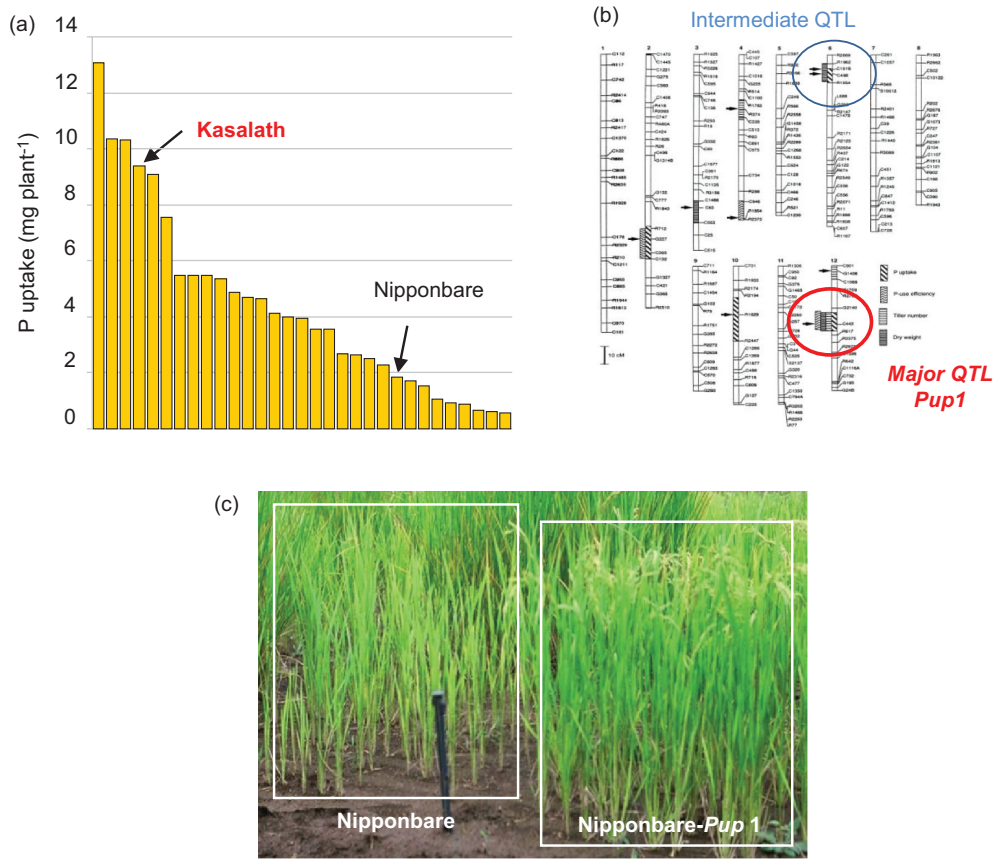


Plate 5.2. Mapping of the *Pup1* major QTL. A screening of 30 diverse rice accessions under P-deficient field conditions was conducted to identify genotypes with high P-uptake ability (**a**). The data were derived from Wissuwa and Ae (2001a). QTL mapping was conducted using a Kasalath x Nipponbare mapping population and *Pup1* was identified as a large-effect QTL on Chr. 12 (**b**). An intermediate-effect QTL is located on Chr. 6 (Wissuwa et al. 1998; see text for details). Nipponbare and a tolerant near-isogenic line with the *Pup1* QTL were grown in a P-deficient field in Tsukuba, Japan (**c**).

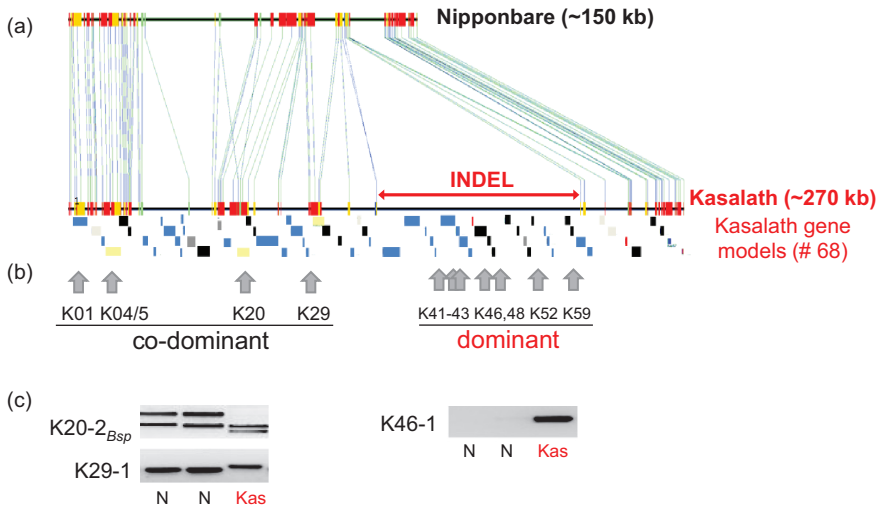


Plate 5.3. Genomic sequence of the *Pup1* region and gene-specific markers. The *Pup1* genomic region derived from sequencing of Kasalath BAC clones was aligned with the corresponding region in the Nipponbare reference genome (**a**). Some regions show partial sequence similarity (indicated by vertical lines). A large insertion/deletion (INDEL) specific to Kasalath *Pup1* is indicated. Sixty-eight Kasalath gene models (indicated by different-size blocks) were predicted in silico and validated gene models were targeted for the design of allele-specific codominant and dominant markers (**b**). Three ideal *Pup1* markers were identified and are recommended for breeding applications (**c**). Details and references are given in the text.

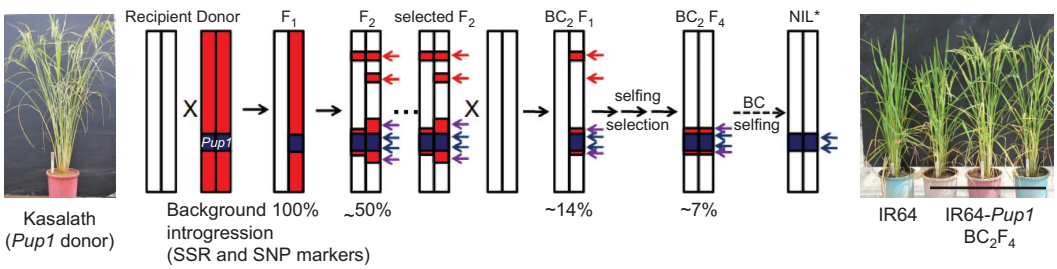


Plate 5.4. Marker-assisted breeding of Pup1 varieties. The tolerant rice variety Kasalath (photo at the left) or Nipponbare-*Pup1* lines were used as donors for the *Pup1* QTL located on chromosome 12 (shown as red bars; *Pup1* QTL indicated in blue). After crossing with an intolerant recipient variety without *Pup1* (Chr. 12 indicated as white bars), F_1 progenies and the following backcross (BC) and selfing generations are selected using *Pup1* foreground markers (blue arrows), flanking markers (purple arrows), and background markers (red arrows). Representative IR64-*Pup1* plants at the BC_2F_4 generation are shown in the photo to the right. Remaining donor introgression in the different generations was determined using SNP markers and indicated as percentage (all chromosomes). *The introgressions in the schemes do represent the actual data; NILs without any remaining background introgressions are hypothetical and probably not attainable.

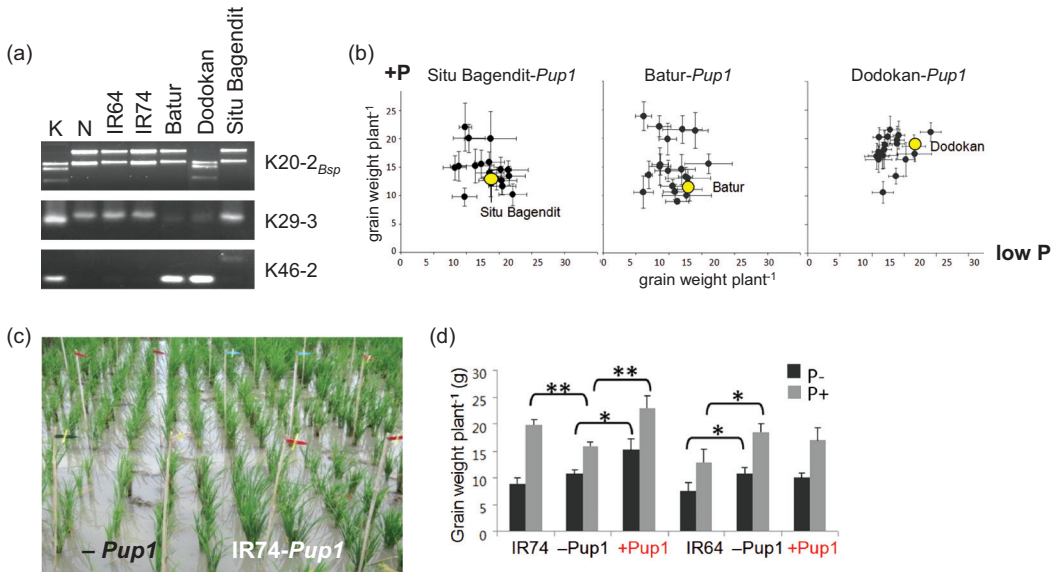


Plate 5.5. Pup1 breeding lines. The recipient parents selected for *Pup1* introgression were genotyped with three *Pup1* markers (**a**). IR64, IR74, and Situ Bagendit have intolerant Nipponbare-type (N) alleles for all three markers. In contrast, Dodokan and Batur have some Kasalath-type (K) alleles, including marker K46-2, which targets *OsPSTOL1*. The Indonesian Pup1 breeding lines (BC_2F_4) were evaluated in an upland field experiment in West Java (Indonesia) under high- and low-P conditions (**b**). The IR64-*Pup1* and IR74-*Pup1* breeding lines were tested under irrigated field conditions in P-deficient soils showing vigorous growth (**c**) and yield advantage (**d**) of IR74-*Pup1* lines.

Plate 6.1. Phylogenetic analysis of MATE-type transporters for all plant MATE transporters that have been functionally characterized to date. Plant members colored in red represent MATE's that have been shown to mediate citrate transport. The asterisks indicate members mediating citrate release in response to aluminum stress. The tree was built using protein sequences with Geneious Tree Builder software.

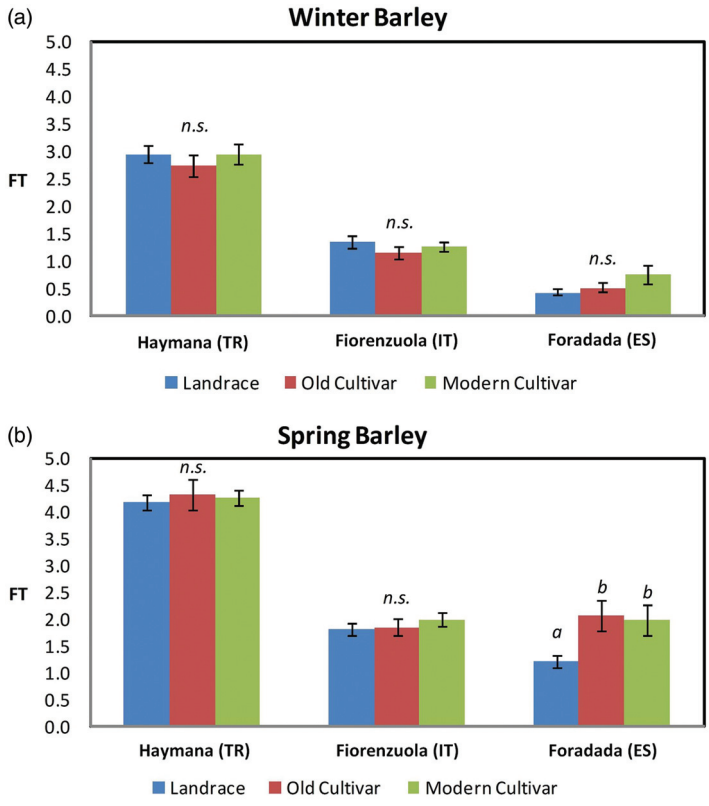


Plate 7.1. Average values of frost tolerance, FT, expressed as visual score ranging from 0 = very tolerant (no visible frost injury) to 5 = very susceptible (>80% plants killed), for A. 91 winter, and B. 59 spring barleys. Genotypes are divided into ancient local landraces, old cultivars (released before 1980), and modern cultivars (released after 1980). Data were recorded in winter 2004/2005 from field trials in three locations in the Mediterranean basin. Vertical bars represent standard error, while letters indicate the result of multiple comparison tests (LSD and Tukey's Procedure, $p < 0.05$).

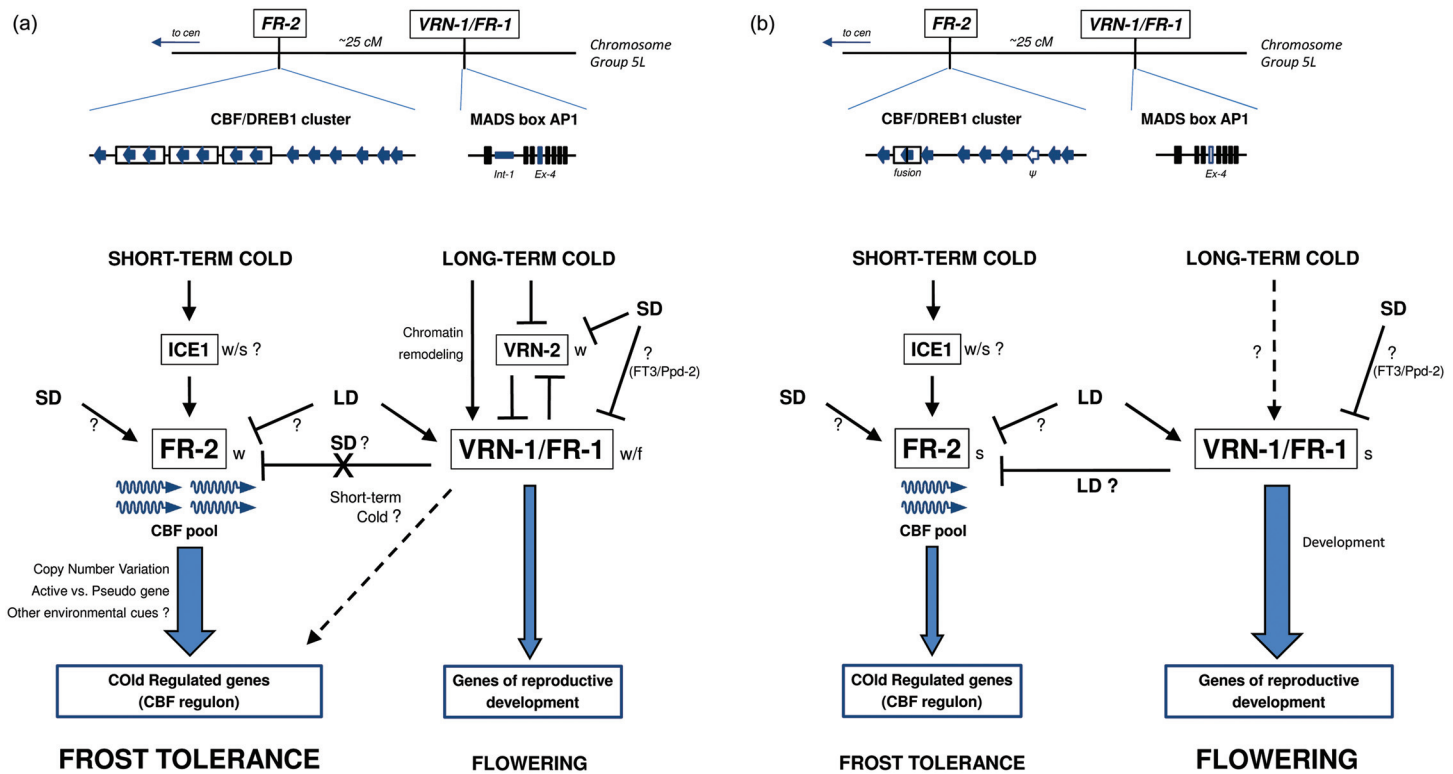


Plate 7.2. Hypothetical model for the action of *VRN-1/FR-1* and *FR-2* in response to different environmental stimuli. The genetic structure of the two loci cis-acting on the long arms of homoeologous group 5 chromosomes in Triticeae is shown separately for winter/facultative (**a**) and spring (**b**) genotypes. SD, short-day conditions; LD, long-day conditions; w/f/s, winter/facultative/spring alleles.

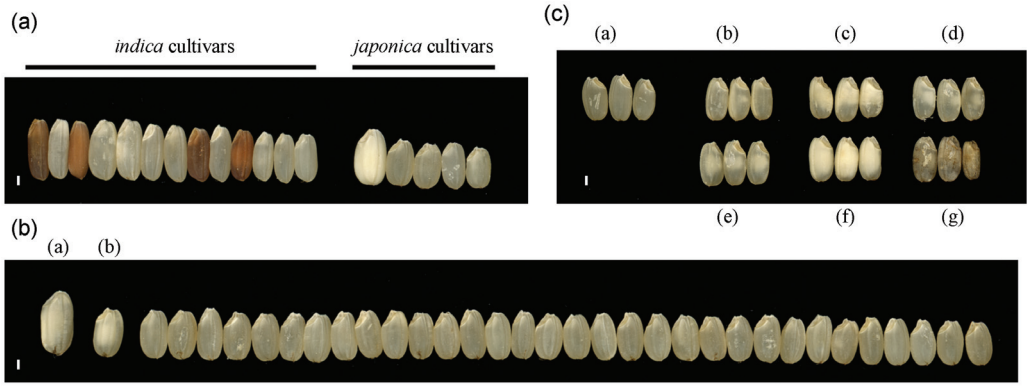


Plate 9.1. (a) Husked rice grains of *indica* (long-grain) and *japonica* (short-grain) rice cultivars. (b) Grain-size variation in Japanese *japonica* rice cultivars of non-glutinous rice: (a) a large-grain cultivar, Oochikara; (b) a Japanese brewing cultivar, Gohyakumangoku. Rice grains were supplied by Dr. Ebana of the National Institute of Agrobiological Sciences Genebank. (c) Appearance of chalkiness in grains of (a) normal type (no chalkiness), (b) white-back type, (c) white-based type, (d) white-belly type, (e) white-core type, (f) milky-white type, and (g) abortive type. *White scale bars* represent 1 mm.

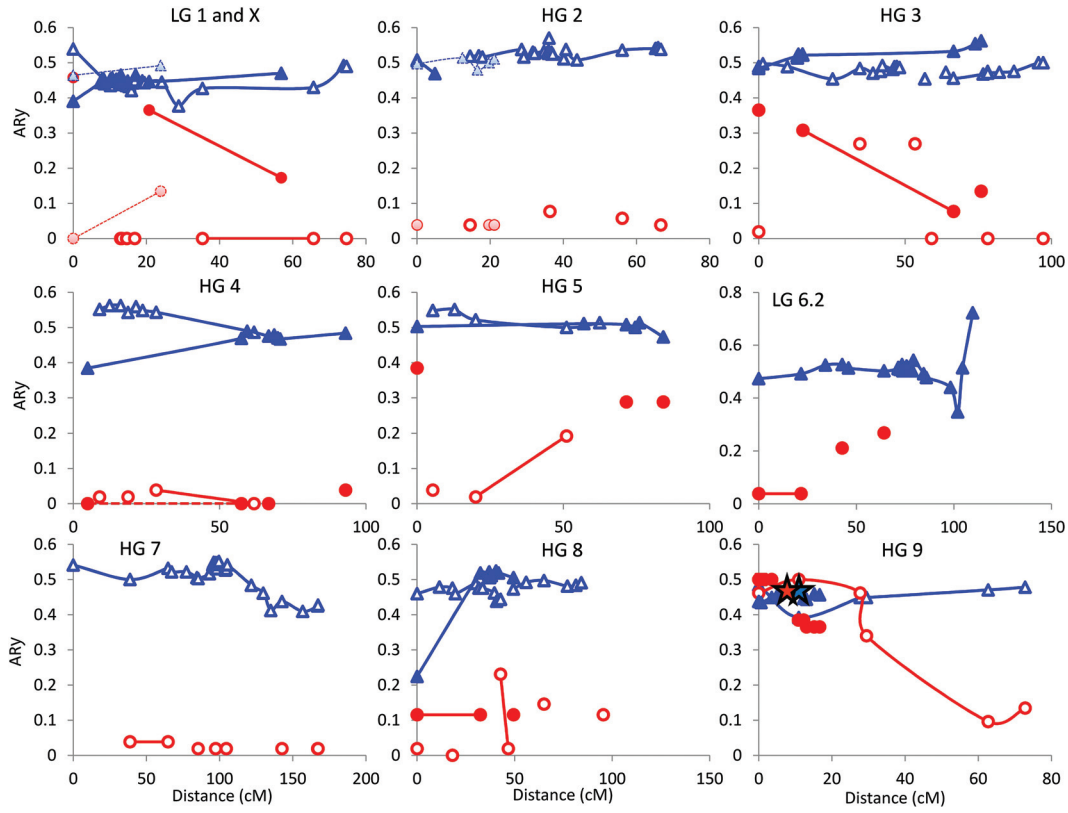


Plate 11.5. Average ratios of the 'YI-0311' allele (AR*y*) on loci in the 186 F_2 and the 26 BC_2F_1 plants. AR*y* at a given locus was calculated by using the equation: AR*y* = (*h*+*y*×2) / (*x*+*h*+*y*) ×2, where *h*, *x*, and *y* are the numbers of homozygous alleles of 'Nakateyutaka,' 'YI-0311,' and heterozygous allele in the F_2 or BC_2F_1 populations, respectively. The graphs are classified by homeologous linkage groups (HGs) 1 to 9. LGX is shown with HG1. The X and Y axes in the graphs correspond to genetic distance on each linkage group (cM) and AR*y*, respectively. Triangles show the F_2 population, while circles show the BC_2F_1 population. Open and solid symbols indicate homeologous LG with appendix 1 (ex. LG1.1), and appendix 2 (ex. LG1.2), respectively. Light-colored symbols represent LG with appendix 3 (LG2.3) or LGX. LG2.2, LG6.1, and LG7.2 were not included in the figures because of incomplete LG development or lower number of mapped markers. Red and blue stars indicate the *ahFAD2A* and *ahFAD2B* loci, respectively.

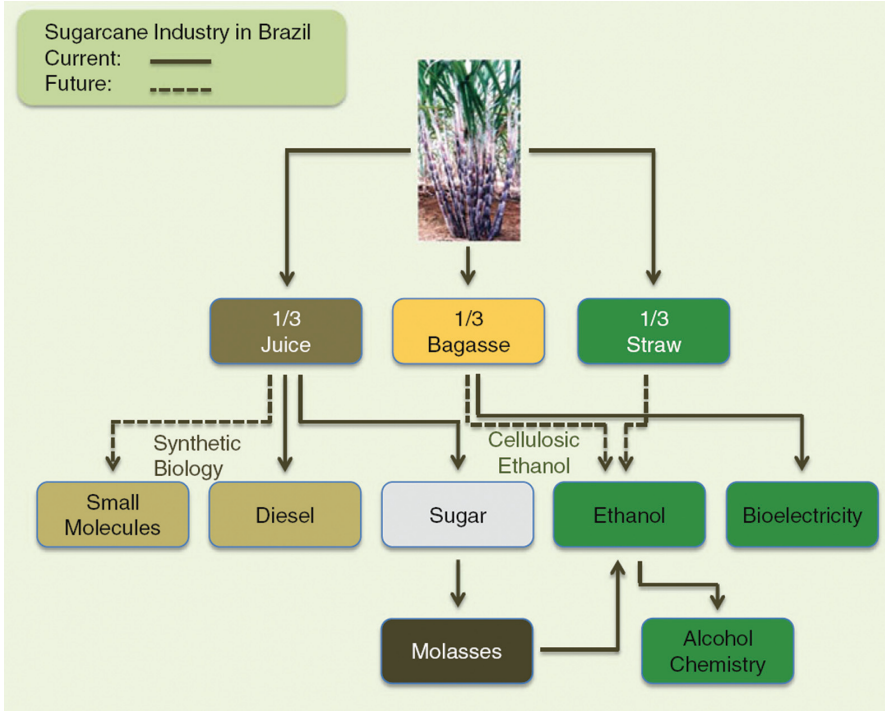


Plate 13.1. Current and future prospects for the sugarcane industry in Brazil (according to Arruda 2011; with kind permission from Springer Science and Business Media).

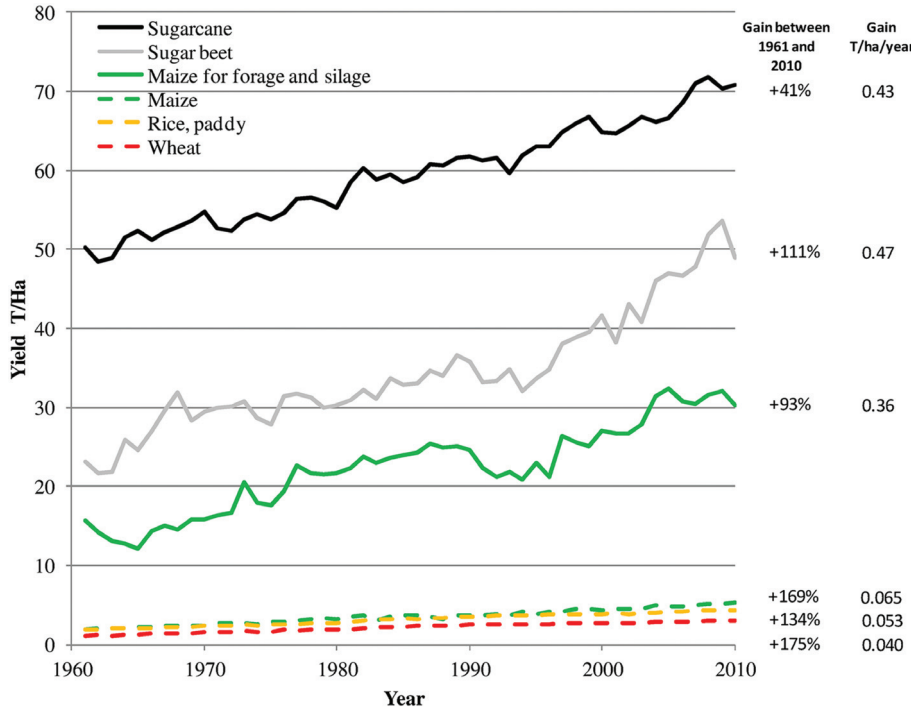


Plate 13.2. Changes in sugarcane yield between 1961 and 2010. Comparison with other grass crops (FAOSTAT 2012).

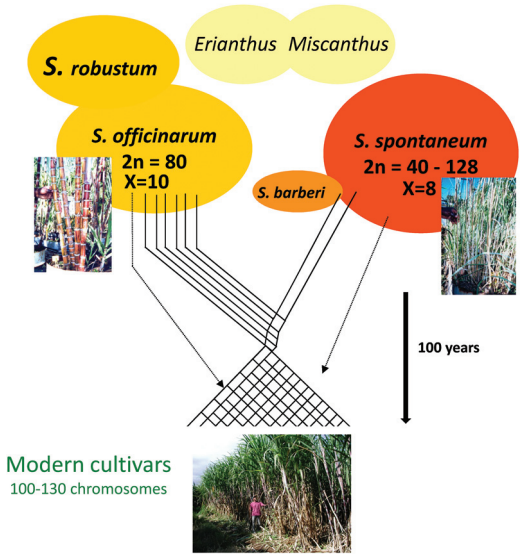


Plate 13.3. Schematic breeding history of sugarcane, showing a second significant step in yield improvement in the 20th century. Modern cultivars (*Saccharum* spp.) rely on a few crosses undertaken 100 years ago between *S. officinarum* (the high-sucrose domesticated species) and *S. spontaneum* (a wild species without sucrose), involving a limited number of founders. These first interspecific hybrids were backcrossed with the high-sucrose species, and subsequently complex recurrent intercrossing of the best products gave rise to modern cultivars.

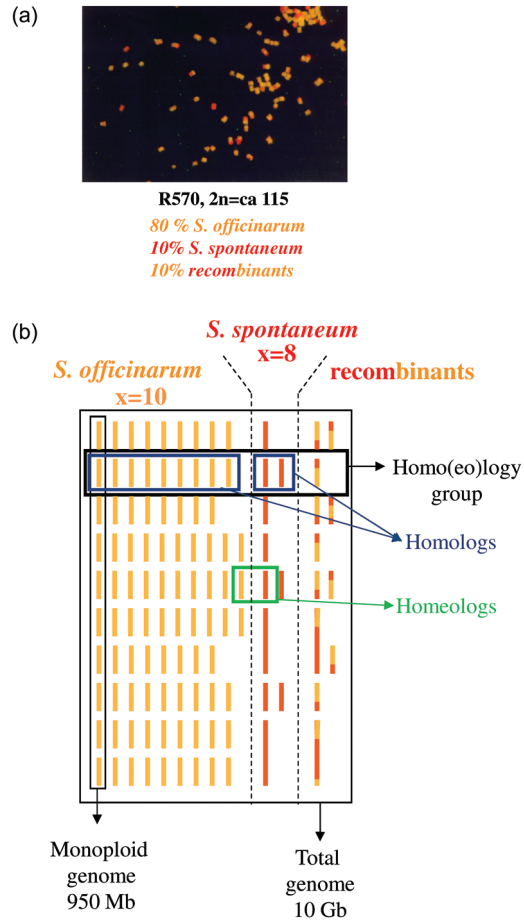


Plate 13.5. Double structure of the genome of modern sugarcane cultivars. About 80% of the chromosome is inherited from *S. officinarum,* 10% from *S. spontaneum* and 10% are recombinant.

(a) *Genomic in situ hybridization* (GISH) using labeled total DNA of *S. officinarum* (yellow fluorescence) and of *S. spontaneum* (red fluorescence) on chromosome preparations of the cultivar R570 (from D'Hont et al. 1996, with kind permission from Springer Science and Business Media).

(b) Schematic representation of the genome of modern interspecific cultivars deduced from molecular cytogenetic data and mapping works. Modern cultivars are highly polyploid and aneuploid with about 120 chromosomes. Colored bars correspond to chromosomes. Chromosome colors are functions of their origins: yellow for *S. officinarum* and red for *S. spontaneum*. Chromosomes from the same row are homologous (or homeologous). Chromosomes of the *S. officinarum* and *S. spontaneum* part of the genome of the modern cultivars are distributed in 10 (X=10) and 8 (X=8) homology groups respectively. (Modified from Grivet and Arruda 2002).