SEED GENOMICS

Edited by
Philip W. Becraft

WILEY-BLACKWELL

Seed Genomics

Edited by
Philip W. Becraft
Genetics, Development and Cell Biology Department
Agronomy Department, Iowa State University
Ames, Iowa, USA

WILEY-BLACKWELL
A John Wiley & Sons, Inc., Publication

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

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 is available upon request.

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 design by Matt Kuhns

Set in 10.5/12 pt Times by Aptara® Inc., New Delhi, India
Printed and bound in Malaysia by Vivar Printing Sdn Bhd

1 2013

Seed Genomics

Contributors

Jose M. Barrero	CSIRO Plant Industry, Canberra, ACT, Australia
Philip W. Becraft	Genetics, Development and Cell Biology Department, Agronomy Department, Iowa State University, Ames, Iowa, USA
Judith Burstin	INRA, UMR1347 Agroécologie, Dijon, France
Colin Cavanagh	CSIRO Plant Industry, Canberra, ACT, Australia
Z. Jeffrey Chen	Section of Molecular Cell and Developmental Biology and Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, Texas, USA
Robert L. Fischer	Department of Plant and Microbial Biology, University of California, Berkeley, California, USA
Jennifer M. Frost	Department of Plant and Microbial Biology, University of California, Berkeley, California, USA
Karine Gallardo	INRA, UMR1347 Agroécologie, Dijon, France
Xueying Guan	Section of Molecular Cell and Developmental Biology and Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, Texas, USA
Frank Gubler	CSIRO Plant Industry, Canberra, ACT, Australia
Jeffery L. Gustin	Horticultural Sciences Department, University of Florida, Gainesville, Florida, USA
Tracie A. Hennen-Bierwagen	Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, Iowa, USA
David Holding	Department of Agronomy and Horticulture, Center for Plant Science Innovation, University of Nebraska, and Beadle Center for Biotechnology, Lincoln, Nebraska, USA
Tzung-Fu Hsieh	Department of Plant and Microbial Biology, University of California, Berkeley, California, USA
Christian A. Ibarra	Department of Plant and Microbial Biology, University of California, Berkeley, California, USA

Anna M. G. Koltunow Commonwealth Scientific and Industrial Research Organization (CSIRO) Plant Industry, Urrbrae, South Australia, Australia

D. L. C. Kumari Fonseka The Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, Durham University, Durham, UK

Christine Le Signor INRA, UMR1347 Agroécologie, Dijon, France

Keith Lindsey The Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, Durham University, Durham, UK

Christine J. Lucas Department of Crop Sciences, University of Illinois, Urbana-Champaign, Urbana, Illinois, USA

David W. Meinke Department of Botany, Oklahoma State University, Stillwater, Oklahoma, USA

Joachim Messing Waksman Institute of Microbiology, Rutgers University, Piscataway, New Jersey, USA

Stephen P. Moose Department of Crop Sciences, University of Illinois, Urbana-Champaign, Urbana, Illinois, USA

Anna Mudge The Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, Durham University, Durham, UK

Alan M. Myers Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, Iowa, USA

Mélanie Noguero INRA, UMR1347 Agroécologie, Dijon, France

Odd-Arne Olsen CIGENE/Department of Plant and Environmental Sciences, Norwegian University of Life Sciences, Aas; and Faculty of Education and Natural Sciences, Hedmark University College, Hamar, Norway

Peggy Ozias-Akins Department of Horticulture, College of Agricultural and Environmental Sciences, University of Georgia Tifton Campus, Tifton, Georgia, USA

Martha Schneerman Department of Crop Sciences, University of Illinois, Urbana-Champaign, Urbana, Illinois, USA

A. Mark Settles Horticultural Sciences Department, University of Florida, Gainesville, Florida, USA

Peter R. Shewry Rothamsted Research, Harpenden, Hertfordshire, UK

Juhyun Shin Department of Plant and Microbial Biology, University of California, Berkeley, California, USA

Imran Siddiqi Centre for Cellular and Molecular Biology (CCMB), Hyderabad, India

Richard Thompson INRA, UMR1347 Agroécologie, Dijon, France

Jennifer F. Topping The Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, Durham University, Durham, UK

Jérôme Verdier The Samuel Roberts Noble Foundation, Ardmore, Oklahoma, USA

Xiyan Yang The Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, Durham University, Durham, UK

Han Zhao Department of Crop Sciences, University of Illinois, Urbana-Champaign, Urbana, Illinois, USA

Contents

Introduction

Philip W. Becraft

Agrarian civilization arose independently several times around the world. One of the earliest events occurred in the Fertile Crescent encompassing the Tigris and Euphrates river valleys of what is presently southeastern Turkey and northern Syria (Lev-Yadun *et al.*, 2000). This is believed to have occurred as early as 11,000 BP and to have involved cultivation of seven founder crops: einkorn wheat, emmer wheat, barley, lentil, pea, bitter vetch, and chickpea. At roughly the same time, early agriculture was occurring in the Yangtze valley of China centered on rice cultivation (Zhao, 2010) and in Mesoamerica involving primarily maize, beans, and squash (Zizumbo-Villarreal and Colunga-GarcíaMarín, 2010). It is notable how prevalent seeds and grains are among these early crops, and this is no accident but due to their high nutritional content and amenability to long-term storage without spoiling. With the advent of agriculture and the resultant stable food supplies came the ability to form permanent settlements, which led ultimately to the rise of modern civilization.

Amazingly, we remain as dependent as ever on seed crops. According to the Food and Agriculture Organization of the United Nations (FAO Statistical Yearbook 2012; http://www.fao.org/), an estimated 50% of global human dietary calories come directly from cereal grains. This figure represents a decline in recent decades, which is largely attributed to increased consumption of calories from vegetable oils, primarily derived from oilseed crops. Livestock products, including dairy, account for only about 13% of human calories, and much of that is indirectly derived from seed-based feeds. Thus, most human caloric intake derives from seed crops.

Seed science has never faced more important challenges or more exciting opportunities than at the present time. As human populations continue to grow, fuel costs soar, and climate change progresses, agriculture will face ever-increasing pressure to produce more food and biofuel, with lower inputs and under increasingly adverse environmental conditions. It is paramount that research investments be made to keep ahead of these growing challenges. New genomic technologies allow biological systems to be studied on scales and at depths not possible just a few years ago. These technologies are providing new insights into the fundamental biology of seed development and metabolism and leading to new strategies for improving seed traits through biotechnical approaches and breeding.

Seed biology is fascinating and complex. Seeds must survive a highly desiccated state and remain quiescent for an indeterminate period of time, then on sensing favorable environmental conditions, reactivate metabolic processes and initiate germination. Seed development involves the coordinated activities of three genetically distinct entities: the embryo, the endosperm, and the maternal plant. The embryo represents the next plant generation; the endosperm is a support tissue that nourishes

Seed Genomics, First Edition. Edited by Philip W. Becraft.

the embryo and, in some species, the germinating seedling; and the maternal tissues contribute the protective and dispersal functions of the seed coat and pericarp. During the morphogenetic phase, the basic body plan of the embryo is established. During filling, storage products accumulate, and finally during seed maturation, tissues acquire the highly specialized ability to survive seed drying and often develop dormancy to ensure against premature germination. The accumulated storage products include starches, oils, proteins, and minerals. They are required to nourish the germinating seedling until it can become established, produce its own photosynthate, and acquire its own mineral nutrients. These storage products are also what make seeds valuable as crops.

This book contains contributions from internationally renowned scientists who describe the application of genomic analyses to various aspects of seed research and improvement. The primary focus of the book is biological rather than technical, although a wide spectrum of technical approaches and considerations are described throughout. In Chapter 1, David Meinke, one of the pioneers of large-scale seed mutant analysis in *Arabidopsis*, provides a historical perspective on the field and his group's contributions. He discusses the SeedGenes database, which compiles a vast reservoir of community information and data on existing seed mutants and the corresponding genes. This chapter illustrates one of the most important ongoing challenges in the genomics era: storing and managing huge amounts of data and presenting it in a format that is accessible and useful to the research community.

Chapters 2 and 3 provide detailed accounts of the processes of embryogenesis and endosperm development, emphasizing their genetic regulation. The embryo produces the next generation of sporophyte plant. Embryogenesis begins with a single-celled zygote and through processes of pattern formation and morphogenesis produces an embryo containing the basic body plan that is perpetuated throughout the life of the plant. The endosperm derives from a second fertilization event and serves as a support tissue to nourish the embryo during early embryogenesis. In species with persistent endosperm, such as cereals, the endosperm also nourishes the germinating seedling until it can become established. In addition to their biological significance, both structures serve as reservoirs for seed storage compounds, which are of value to humans. Both chapters highlight the complexities of these systems, illustrating the power of single-gene mutant analyses and their inherent limitations and the need for systems biology approaches that fully integrate data to understand the interacting networks that simultaneously occur at different levels (e.g., transcriptomic, proteomic, metabolomic).

Endosperm also exhibits gene imprinting, whereby maternally inherited versus paternally inherited alleles show differential expression because of epigenetic regulation. The adaptive functions and molecular mechanisms of this phenomenon are presented in Chapter 4. It appears to be involved in regulating nutrient allocation to developing seeds with implications for seed yield as well as maintaining genome integrity by suppressing transposon activity during reproduction. One exciting aspect of imprinting is that some of the molecular machinery appears to be involved in repressing seed development until triggered by fertilization, which could relate to apomixis. Apomixis is the fertilization-independent formation of seeds that retain the identical genetic constitution of the mother plant. As discussed in Chapter 5, apomixis has enormous economic potential because of the possibility of fixing hybrid vigor, and more recent progress suggests it might soon be possible to engineer apomixis into sexual crop species.

Seeds occupy a critical phase in the plant life cycle, and seed dormancy controls the timing of germination to maximize the likelihood that seedlings will be met with favorable conditions to establish, grow, and complete their reproductive cycle. The many mechanisms of dormancy allow different species to exist in their respective ecological niches by synchronizing germination to the various limiting conditions present in different environments (e.g., temperature or moisture).

Dormancy is also a critical agronomic trait; inadequate dormancy can result in crop yield losses owing to preharvest sprouting, whereas overly dormant seeds might fail to germinate when planted resulting in poor stand establishment. Chapter 6 discusses ongoing approaches to dissect the complex regulation of seed dormancy.

As mentioned, the major value of seeds to humans comes from the storage compounds they accumulate, primarily proteins, oils, and starch as well as minerals and secondary metabolites. In addition to nourishing the germinating seedling, these compounds contribute to the nutritional value of seeds for human or livestock consumption, providing energy and protein as well as other dietary benefits such as antioxidants and fiber. These compounds have found increasing use more recently in industrial applications, including biofuels, plastics, and more. Not only is the yield of these various compounds important but also the quality. The biochemical differences in seed composition impact the end use of seeds by affecting things such as baking characteristics of flour, flavor or heat tolerance of oils, or the digestibility of starch. Chapters 7–10 discuss starches, proteins, and oils, including their metabolism and factors that affect their accumulation and quality for various end uses. A common theme for all these compounds is the surprising complexity in their metabolism and how subtle structural variation can influence their physical properties. For example, starch with nothing but polymers of glucose subunits connected by α1-4 or α1-6 glycoside bonds shows dramatic differences in things such as gelling properties and digestibility, depending on the particular arrangement of the bonds and molecular packing into granules. There is a large repertoire of enzymes, not fully understood, that confer these molecular properties to the starch molecule. Proteins and oils are similarly diverse and complex. Genetic and genomic studies, including comparative genomics of different species, are lending insights to how variation in such properties are controlled and how these storage systems evolved.

In addition to the storage compounds that accumulate in seeds, another valuable seed product is cotton fiber, which is important in the textile industry and for other uses. Chapter 11 describes genomic studies in cotton where the most important seed trait is fiber. Ongoing studies seek to understand the genomic underpinnings controlling fiber quality and yield. This also serves as a model for studying processes of plant cell growth and cell wall deposition. Studies on the establishment of fiber cell fate specification provide an excellent example of translational research where basic research in *Arabidopsis* trichome development directly contributes to the understanding of an economically important trait. Cotton is also a model polyploid system for studying the negotiations and accommodations that occur between independent genomes when they are combined.

One of the most exciting areas of crop genomic science is at the interface with crop breeding. After all, the ultimate goal of plant genomics research is for crop improvement. The Illinois Long-Term Selection Project is a unique resource where a single starting maize population has been subjected to >110 cycles of continuous selection for seed traits including protein and oil content. These selection schemes have been reversed for several subpopulations, lines have been crossed to create mapping populations, germplasm has contributed to breeding programs, and, more recently, genomic analyses have been applied to these populations. As described in Chapter 12, this has provided new insights into genome-level responses to long-term selection, which will have bearing on one of the great questions pondered by plant breeders (or probably more often by nonbreeders): "When will the genetic variation run out?"

Finally, phenotypic analysis is often cited as the bottleneck to high-throughput studies. In closing, Chapter 13 discusses various spectral imaging technologies that are being combined with computer algorithms to develop high-throughput, automated systems for analyzing seed traits. As described, these approaches afford the opportunity to gather much more information in a single measurement than is possible with manual techniques and to do it more quickly and more accurately. Some of these imaging techniques provide three-dimensional spatial information as well as compositional information. Furthermore, the data are preserved and can often be mined for additional information as new computer algorithms are developed. This area holds tremendous promise for future advancement as new imaging technologies are developed and applied to the analysis of seed traits. When combined with genomic studies, basic research on seed biology and breeding for improved seed traits can be greatly accelerated, and genetic potentials can be realized.

I thank the authors for their outstanding contributions. Their efforts make readily accessible an enormous amount of information, some of which was previously unpublished. I greatly enjoyed working on this project and found each of the chapters exciting and educational. I hope you find it valuable, too.

References

Lev-Yadun S, Gopher A, Abbo S (2000). The cradle of agriculture. Science 288: 1602–1603.

Zhao Z (2010). New data and new issues for the study of origin of rice agriculture in China. Archaeological and Anthropological Sciences 2: 99–105.

Zizumbo-Villarreal D, Colunga-GarcíaMarín P (2010). Origin of agriculture and plant domestication in West Mesoamerica. Genetic Resources and Crop Evolution 57: 813–825.

1 Large-Scale Mutant Analysis of Seed Development in *Arabidopsis*

David W. Meinke

Introduction

With advances in DNA sequencing and reports of sequenced genomes appearing at an accelerating rate, one can easily forget an important principle that first guided research in molecular plant biology 25 years ago – that genomics and proteomics are most powerful when focused on model genetic organisms. It is therefore fitting that a book devoted to seed genomics should include several chapters on the use of genetic analysis to address fundamental questions in seed biology. My objective in this chapter is not to detail all of the seed mutants analyzed to date or to describe all of the biological questions that have been addressed with these mutants. Instead, I have chosen to focus on my own professional journey, spanning the past 35 years, to isolate and characterize large numbers of embryo-defective (*emb*) mutants in the model plant, *Arabidopsis thaliana*. This choice is justified by a quick look at the numbers involved. More embryo mutants have been isolated and characterized in *Arabidopsis*, and their genes identified, than in all other angiosperms combined. Any discussion of the strategies, procedures, and conclusions drawn from the analysis of large numbers of mutants defective in seed development must therefore focus on what has been accomplished in *Arabidopsis*. This work has been performed over several decades by dozens of individuals in my laboratory, along with scores of investigators throughout the *Arabidopsis* community. The results summarized in this chapter are a testament to their combined efforts and insights. Readers unfamiliar with basic features of seed development are referred to Chapters 2 and 3 of this book.

Historical Perspective

Mutants defective in seed development have long played an important role in genetic analysis (Meinke, 1986, 1995) – from Mendel's wrinkled seed phenotype in pea, which results from transposon inactivation of a starch-branching enzyme (Bhattacharyya *et al.*, 1990), to studies by early plant geneticists on germless (embryo-specific) and defective kernel (*dek*) mutants of maize (Demerec, 1923; Mangelsdorf, 1923; Emerson, 1932) and the nature of embryo-endosperm interactions during seed development (Brink and Cooper, 1947). Large-scale mutant analysis of seed development in maize began in the late 1970s with the isolation and characterization of several hundred *dek* mutants generated following ethyl methanesulfonate (EMS) pollen mutagenesis (Neuffer and

Seed Genomics, First Edition. Edited by Philip W. Becraft.