SEED GENOMICS

Edited by

Philip W. Becraft

WILEY-BLACKWELL

Seed Genomics

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

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

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.

 ISBN-13: 978-0-4709-6015-8/2013.

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

1 2013

Contents

Contributors **xi**

Introduction **1**
Philip W. Becraft

Chapter 1 **Large-Scale Mutant Analysis of Seed Development in *Arabidopsis*** **5**
David W. Meinke

Introduction 5
Historical Perspective 5
Arabidopsis Embryo Mutant System 7
Large-Scale Forward Genetic Screens for Seed Mutants 7
Approaches to Mutant Analysis 8
Strategies for Approaching Saturation 10
SeedGenes Database of Essential Genes in *Arabidopsis* 11
Embryo Mutants with Gametophyte Defects 13
General Features of *EMB* Genes in *Arabidopsis* 14
Value of Large Datasets of Essential Genes 15
Directions for Future Research 16
Acknowledgments 17
References 17

Chapter 2 **Embryogenesis in *Arabidopsis*: Signaling, Genes, and the Control of Identity** **21**
D. L. C. Kumari Fonseka, Xiyan Yang, Anna Mudge, Jennifer F. Topping, and Keith Lindsey

Introduction 21
Cellular Events 21
Genes and Signaling – the Global Picture 23
Coordination of Genes and Cellular Processes: a Role for Hormones 25
Genes and Pattern 30
Conclusion and Future Directions 36
References 36

Chapter 3 Endosperm Development **43**
Odd-Arne Olsen and Philip W. Becraft

Introduction 43
Overview of Endosperm Structure and Development 43
Endosperm Cell Fate Specification and Differentiation 48
Genomic Resources 53
Transcriptional Profiling of Endosperm Development 54
Gene Imprinting in Cereal Endosperm 56
Conclusion 57
Acknowledgments 58
References 58

Chapter 4 Epigenetic Control of Seed Gene Imprinting **63**
Christian A. Ibarra, Jennifer M. Frost, Juhyun Shin, Tzung-Fu Hsieh, and Robert L. Fischer

Introduction 63
Genomic Imprinting and Parental Conflict Theory 63
Epigenetic Regulators of *Arabidopsis* Imprinting 65
Mechanisms Establishing *Arabidopsis* Gene Imprinting 69
Imprinting in the Embryo 74
Imprinting in Monocots 75
Evolution of Plant Imprinting 77
Conclusion 78
Acknowledgments 78
References 78

Chapter 5 Apomixis **83**
Anna M. G. Koltunow, Peggy Ozias-Akins, and Imran Siddiqi

Introduction 83
Biology of Apomixis in Natural Systems 84
Phylogenetic and Geographical Distribution of Apomixis 89
Inheritance of Apomixis 90
Genetic Diversity in Natural Apomictic Populations 93
Molecular Relationships between Sexual and Apomictic Pathways 94
Features of Chromosomes Carrying Apomixis Loci and Implications for Regulation of Apomixis 95
Genes Associated with Apomixis 96
Transferring Apomixis to Sexual Plants: Clues from Apomicts 97
Synthetic Approach to Building Apomixis 98
Synthetic Clonal Seed Formation 102
Conclusion and Future Prospects 103
References 103

Chapter 6 High-Throughput Genetic Dissection of Seed Dormancy **111**
Jose M. Barrero, Colin Cavanagh, and Frank Gubler

Introduction 111
Profiling of Transcriptomic Changes 113
Use of New Sequencing Platforms and Associated Techniques to Study Seed Dormancy 114
Visualization Tools 116
Coexpression Studies and Systems Biology Approaches 116
Mapping Populations for Gene Discovery 117
Perspective 118
Acknowledgments 119
References 119

Chapter 7 Genomic Specification of Starch Biosynthesis in Maize Endosperm **123**
Tracie A. Hennen-Bierwagen and Alan M. Myers

Introduction 123
Overview of Starch Biosynthetic Pathway 124
Genomic Specification of Endosperm Starch Biosynthesis in Maize 126
Conclusion 134
References 134

Chapter 8 Evolution, Structure, and Function of Prolamin Storage Proteins **139**
David Holding and Joachim Messing

Introduction 139
Prolamin Multigene Families 139
Endosperm Texture and Storage of Prolamins 143
Conclusion 154
References 154

Chapter 9 Improving Grain Quality: Wheat **159**
Peter R. Shewry

Introduction 159
Grain Structure and Composition 159
End Use Quality 161
Redesigning the Grain 163
Manipulation of Grain Protein Content and Quality 163
Manipulation of Grain Texture 167
Development of Wheat with Resistant Starch 168
Improving Content and Composition of Dietary Fiber 169
Wheat Grain Cell Walls 169
Conclusion 173
Acknowledgments 173
References 173

Chapter 10 Legume Seed Genomics: How to Respond to the Challenges and Potential of a Key Plant Family? 179
Mélanie Noguero, Karine Gallardo, Jérôme Verdier, Christine Le Signor, Judith Burstin, and Richard Thompson

Introduction 179
Development of Genomics Tools 180
Applications of Genomics Tools to Legume Seed Biology 185
Future Challenges 192
References 193

Chapter 11 Cotton Fiber Genomics 203
Xueying Guan and Z. Jeffrey Chen

Introduction 203
Cotton Fiber Development 204
Roles for Transcription Factors in Development of *Arabidopsis* Leaf Trichomes, Seed Hairs, and Cotton Fibers 204
Fiber Cell Expansion through Cell Wall Biosynthesis 208
Regulation of Phytohormones during Cotton Fiber Development 209
Cotton Fiber Genes in Diploid and Tetraploid Cotton 210
Roles for Small RNAs in Cotton Fiber Development 211
Conclusion 212
References 213

Chapter 12 Genomic Changes in Response to 110 Cycles of Selection for Seed Protein and Oil Concentration in Maize 217
Christine J. Lucas, Han Zhao, Martha Schneerman, and Stephen P. Moose

Introduction 217
Background on the Illinois Long-Term Selection Experiment 217
Phenotypic Responses to Selection 219
Additional Traits Affected by Selection 220
Unlimited Genetic Variation? 221
Genetic Response to Selection: QTL Mapping in the Crosses of IHP x ILP and IHO x ILO 222
New Mapping Population: Illinois Protein Strain Recombinant Inbreds 223
Characterization of Zein Genes and Their Expression in Illinois Protein Strains 225
Contribution of Zein Regulatory Factor *Opaque2* to Observed Responses to Selection in Illinois Protein Strains 227
Major Effect QTL May Explain IRHP Phenotype 228
Zein Promoter-Reporter Lines to Investigate Regulation of 22-kDa α-Zein Gene Expression in Illinois Protein Strains 229
Regulatory Changes in *FL2-mRFP* Expression When Crossed to Illinois Protein Strains 230
Regulation of *FL2-mRFP* 232
Acknowledgments 233
References 234

Chapter 13 Machine Vision for Seed Phenomics **237**
Jeffery L. Gustin and A. Mark Settles

Introduction 237
High-Energy Imaging: X-ray Tomography and Fluorescence 238
Optical Imaging: Visible Spectrum 240
Resonance Absorption: Infrared Spectrum 242
Resonance Emission: Nuclear Magnetic Resonance 245
Conclusion 246
Acknowledgments 246
References 246

Color plate section found between pages 42 and 43.

Index **253**

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

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.

Sheridan, 1980; Sheridan and Neuffer, 1980). Another 64 *dek* mutants, along with 51 embryo-specific mutants, were later described in genetic stocks known to contain the transposable element, Robertson's Mutator (Clark and Sheridan, 1991; Sheridan and Clark, 1993; Scanlon *et al.*, 1994). Although many additional mutants of this type have likely been encountered in screens of other transposon insertion lines, a global analysis of all disrupted genes associated with kernel phenotypes in maize has not been published. Attention has focused instead on a detailed characterization of selected mutants of particular interest (José-Estanyol *et al.*, 2009), including the *dek1* mutant defective in aleurone cell identity (Becraft *et al.*, 2002, Tian *et al.*, 2007; Yi *et al.*, 2011) and a number of viviparous mutants that exhibit premature germination (Suzuki *et al.*, 2003, 2006, 2008). By contrast, seed mutants in other grasses such as rice (Hong *et al.*, 1996; Kamiya *et al.*, 2003; Kurata *et al.*, 2005) have been examined in much less detail, with most genetic studies focused on other phenotypes of interest.

The isolation and characterization of embryo-lethal mutants of *Arabidopsis* was first described by Andreas Müller in Gatersleben, Germany. Müller (1963) characterized 60 mutants with different embryo phenotypes, including defects in embryo pigmentation, demonstrated that mutant and wild-type seeds could be distinguished in heterozygous siliques, and established the "Müller embryo test" to assess the mutagenic effects of ionizing radiation and chemical treatments in *Arabidopsis*. Although his attention was later directed to other systems, Müller remained particularly interested in fusca mutants, which accumulate anthocyanin during embryo maturation (Miséra *et al.*, 1994). Original stocks of the other mutants identified by Müller (1963) were not maintained.

I started to work on *Arabidopsis* as a graduate student in the laboratory of Ian Sussex at Yale University. My Ph.D. dissertation described the isolation and characterization of six embryo-lethal mutants of *Arabidopsis* and the value of such mutants in the study of plant embryo development (Meinke and Sussex, 1979a, 1979b). This work began at a time when *Arabidopsis* was known more for research in biochemical genetics than in developmental or molecular genetics. After completing a postdoctoral project on soybean seed storage proteins with Roger Beachy at Washington University in St. Louis (Meinke *et al.*, 1981), I moved to Oklahoma State University, where I focused my attention again on embryo mutants of *Arabidopsis*. My initial strategy was to analyze additional mutants isolated following EMS seed mutagenesis (Meinke, 1985; Baus *et al.*, 1986; Heath *et al.*, 1986; Franzmann *et al.*, 1989). Because some mutant seeds were capable of germinating and producing defective seedlings in culture, I adopted the term "embryo defective" (*emb*) rather than "embryo lethal" to describe the expanding collection. This nomenclature has been used ever since, although some *EMB* locus numbers were later replaced with more informative symbols (*sus*, *twn*, *lec*, *bio*, *ttn*) to indicate phenotypes of special interest.

A different approach to genetic analysis of plant embryo development was first described 20 years ago in a publication from Gerd Jürgens' laboratory in Germany (Mayer *et al.*, 1991). Rather than attempt to analyze every mutant defective in embryo development, the Jürgens group focused attention on a small number of mutants with defective seedlings that appeared to result from alterations in embryo pattern formation. As described elsewhere in this book, several of these mutants uncovered important cellular pathways associated with plant embryo development, although in many cases, the gene products were unexpected and did not appear to support the original hypothesis, based on work with *Drosophila*, that embryo patterning mutants should identify transcription factors that regulate developmental decisions. Whereas my approach was to "cast a wide net" and explore interesting stories based on the analysis of many different types of mutants, the Jürgens group focused on a limited set of phenotypes defined by a handful of genes with multiple alleles and identified gene networks associated with those phenotypes. In retrospect, both of these approaches were required to develop a comprehensive picture of the genetic control of plant embryo development.

Table 1.1 Experimental Features That Make *Arabidopsis thaliana* an Attractive System for Large-Scale Mutant Analysis of Seed Development

Arabidopsis Feature	Relevance of Feature to Genetic Analysis of Seed Development
Self-pollinated flowers	Crosses not required to maintain *emb* mutants and most genetic stocks
Indeterminate inflorescences	Mature plants contain large numbers of siliques at different stages of development, arranged in a predictable progression along each stem; facilitates identification of embryos at desired stage of development
Transparent seed coat	Wild-type seeds at the cotyledon stage are green and can be readily distinguished from unfertilized ovules and aborted seeds
Spontaneous seed abortion rare	Facilitates identification of mutant seeds in heterozygous siliques
Small seed size at maturity	Embryos within immature seeds are readily observed with Nomarski (DIC) light microscopy; optical sectioning through immature seeds possible
Siliques contain 50–60 seeds	Segregation of normal and mutant seeds readily observed in 1 silique
Short pollen-tube growth path	Facilitates recovery of mutants defective in both embryo and gametophyte development

Arabidopsis Embryo Mutant System

The advantages of *Arabidopsis* as a model system for research in plant biology are well known (Redéi, 1975; Meyerowitz and Somerville, 1994; Meinke *et al.*, 1998; Koornneef and Meinke, 2010). Important features that make *Arabidopsis* suitable for large-scale mutant analysis of seed development have also been described (Meinke, 1994). Several of these features are highlighted in Table 1.1. Recessive embryo-defective mutants are maintained as heterozygotes, which typically produce 25% mutant seeds after self-pollination. Because each silique contains 50–60 total seeds and multiple siliques are arranged in a developmental progression along the length of each stem, mutant seeds at many different stages of development can be found on a single plant at maturity. Mutant and normal seeds can be readily distinguished, based on size, color, and embryo morphology, by screening immature siliques under a dissecting microscope. Mutant embryos that have reached an advanced globular stage can be removed with fine-tipped forceps and examined further; embryos arrested at earlier stages of development are best observed under a compound microscope equipped with Nomarski (differential interference contrast [DIC]) optics. After seed mutagenesis, siliques of chimeric M_1 plants can be screened to identify flowers that arose from the mutant sector (Meinke and Sussex, 1979a, 1979b). Mature siliques derived from this sector are harvested to collect dry seeds. After germination, heterozygous and wild-type plants often segregate in a 2:1 ratio. If insertion lines are involved and the disrupted *EMB* gene is associated with a selectable marker, the appropriate selection agent can be used to identify heterozygous plants at the seedling stage, provided that there are no additional inserts located elsewhere in the genome. With EMS mutants, heterozygous plants cannot be distinguished from wild-type plants until selfed siliques have matured and are screened for defective seeds. When plants segregating for an *emb* mutation are crossed for allelism tests, parental heterozygotes must be identified before the cross can be performed, which limits the time available for crosses to be completed. Allelic mutants that fail to complement result in siliques with 25% mutant seeds; mutants disrupted in different genes typically produce siliques with all normal seeds.

Large-Scale Forward Genetic Screens for Seed Mutants

In contrast to screens for most visible phenotypes in *Arabidopsis*, which involve the identification of homozygotes in a second (M_2) generation following seed mutagenesis, forward genetic screens for embryo-defective mutations can be performed directly on M_1 plants. This approach was used

to isolate most of the original *emb* mutants analyzed in my laboratory (Meinke, 1985). When Ken Feldmann developed a method for *Agrobacterium*-mediated seed transformation in *Arabidopsis* and began to grow large populations of T-DNA insertion lines at DuPont in the late 1980s (Feldmann, 1991), two different groups were involved in screening the populations for embryo-defective mutations. One group, comprising investigators associated with Robert Goldberg's laboratory at UCLA (Yadegari *et al.*, 1994), screened half of the plants; members of my laboratory screened the other half (Errampalli *et al.*, 1991; Castle *et al.*, 1993). The same strategy was used with a second population of plants that Feldmann made available several years later at the University of Arizona. My approach to the analysis of these populations was first to determine which mutants were tagged with T-DNA and which were not tagged. About two thirds of the lines that segregated for an embryo-defective mutation were not amenable to rapid gene identification because they fell into the second category. The method used to resolve tagging status involved transplanting kanamycin-resistant seedlings derived from selfed heterozygotes to soil and determining whether all of those plants produced siliques with 25% mutant seeds, as expected if a single T-DNA insert was present and disrupted an *EMB* gene. When additional inserts were involved, we identified subfamilies in future generations that contained a single insert and then proceeded with the analysis described above. For mutants examined in my laboratory, the original *emb1* to *emb69* alleles were identified after EMS (or in some cases x-ray) seed mutagenesis, *emb71* to *emb180* mutants involved the DuPont collection, and the *emb200* series was reserved for the Arizona collection. Most of these *EMB* loci are listed in Meinke (1994) and in the "Archival Data on Meinke Lab Mutants" link at the SeedGenes website devoted to genes with essential functions during seed development in *Arabidopsis* (www.seedgenes.org). In some cases, the gene responsible for the mutant phenotype has since been identified. In many cases, however, the association between mutant phenotype and gene function remains to be determined.

A major breakthrough in forward genetic analysis of seed development occurred in the late 1990s, when David Patton and Eric Ward at Ciba-Geigy, which later became Syngenta (Research Triangle Park, NC), embarked on a large-scale, forward genetic screen for essential genes of *Arabidopsis*. The rationale was that some essential gene products identified through such efforts might represent promising targets for novel herbicides. Over the next 15 years, in close collaboration with my laboratory, >120,000 T-DNA insertion lines were screened for seed phenotypes, including embryo and seed pigment defects, >1600 promising mutants were isolated and characterized, ~440 tagged mutants were identified, and ~200 gene identities were revealed (McElver *et al.*, 2001). Of equal importance, Syngenta ultimately agreed to make most of these gene identities public, after they had been evaluated in house (Tzafrir *et al.*, 2004). This provided the foundation for a large-scale NSF 2010 project in my laboratory that established, in collaboration with Allan Dickerman at the Virginia Bioinformatics Institute, a comprehensive database of all known essential genes required for seed development in *Arabidopsis* (Tzafrir *et al.*, 2003). Results of the "SeedGenes" project are described later in this chapter.

Approaches to Mutant Analysis

The belief that lethal mutants are not useful or informative because they cannot be analyzed in detail is misguided. Sometimes the terminal phenotype alone is sufficient to offer valuable insights. Abnormal suspensor (*sus*) and twin (*twn*) mutants provided early support for the idea that the embryo proper normally restricts the developmental potential of the suspensor (Marsden and Meinke, 1985; Schwartz *et al.*, 1994; Vernon and Meinke, 1994). The leafy cotyledon (*lec*) mutant phenotype (Meinke, 1992; Meinke *et al.*, 1994) indicated that the default state for cotyledons is a leaflike

structure, consistent with the evolution of cotyledons as modified leaves. The initial steps in mutant analysis often involve determination of segregation ratios of mutant seeds in heterozygous siliques and the use of dissecting microscopes and light microscopy to characterize the terminal embryo phenotype. A reduced frequency or unusual distribution of mutant seeds in heterozygous siliques frequently indicates an additional role of the disrupted gene in male gametophyte development (Meinke, 1982; Muralla *et al.*, 2011). A mixture of aborted seeds and unfertilized ovules often indicates a role in female gametophyte development (Berg *et al.*, 2005). Examination of mutant seeds with Nomarski optics or sectioned material with light or electron microscopy can reveal unexpected features, such as incomplete cell walls in the *cyt1* mutant (Nickle and Meinke, 1998) and giant endosperm nuclei and enlarged embryo cells in titan (*ttn*) mutants (Liu and Meinke, 1998; Liu *et al.*, 2002; Tzafrir *et al.*, 2002). Light microscopy, in combination with gel electrophoresis of seed proteins, can also be used to determine whether mutant embryos unable to complete morphogenesis continue to differentiate at the cellular level (Heath *et al.*, 1986; Patton and Meinke, 1990; Yadegari *et al.*, 1994).

The original idea behind examining the response of mutant embryos in culture was to determine whether mutant seedlings or callus tissue could be produced for further analysis and to search for auxotrophic mutants that survived on enriched media containing the required nutrient (Baus *et al.*, 1986). This approach resulted in the successful identification of the first auxotrophic mutant (*bio1*) known to be associated with embryo lethality in *Arabidopsis*, and it helped to explain the scarcity of plant auxotrophs identified at the seedling stage (Schneider *et al.*, 1989). Another mutant (*bio2*) was later found to be blocked at a different step in the same pathway (Patton *et al.*, 1998). The chromosomal deletion associated with this mutant allele also contributed, by chance, to the identification of a closely linked gene (*FPA*) involved in floral induction (Schomburg *et al.*, 2001). Reverse genetic analysis later revealed that *BIO1* is part of a complex (*BIO3–BIO1*) locus that encodes a fusion protein responsible for two sequential steps in biotin synthesis (Muralla *et al.*, 2008). Embryo culture experiments also enabled further characterization of mutants with late defects in embryo development (Vernon and Meinke, 1995) and resulted in the identification of two mutants with especially striking phenotypes: *lec1* (Meinke, 1992) and *twn1* (Vernon and Meinke, 1994).

Another approach to mutant analysis that can occur independently of gene isolation involves mapping the chromosomal locations of *EMB* genes relative to morphological or molecular markers. We used this approach to enhance the classical genetic map of *Arabidopsis* and to facilitate the identification of potential mutant alleles suitable for genetic complementation tests (Patton *et al.*, 1991; Franzmann *et al.*, 1995). More recently, we performed allelism tests between mapped (but not cloned) mutants and cloned (but not mapped) mutants to identify new alleles of cloned *EMB* genes and reveal the identities of mapped *EMB* genes (Meinke *et al.*, 2009b). Classical genetic mapping with *emb* mutants is enhanced by the fact that heterozygous plants can be identified by screening selfed siliques for the presence of defective seeds. After completion of the genome sequence, the classical genetic map was found to have many regions inconsistent with the known order of genes along the chromosome. This finding led to the establishment of a sequence-based map of genes with mutant phenotypes to document the confirmed locations of genes previously found on the classical genetic map (Meinke *et al.*, 2003). An expanded version of this dataset, which includes 2400 *Arabidopsis* genes with a loss-of-function mutant phenotype, was recently published (Lloyd and Meinke, 2012).

Because of their small size, *Arabidopsis* seeds were once viewed as inaccessible to biochemical or molecular analysis. With continued technological advances and the development of alternative methods for detecting trace substances of interest, some of these barriers have been removed. One example from my laboratory involved the use of sensitive microbiological assays to demonstrate that

arrested embryos from the *bio1* mutant of *Arabidopsis* contain reduced levels of biotin (Shellhammer and Meinke, 1990). A more recent example is the demonstration that arrested embryos from the *sus1* mutant contain altered profiles of miRNAs, consistent with the known function of the *SUS1/DCL1* gene (Schauer *et al.*, 2002) in promoting the formation of miRNAs (Nodine and Bartel, 2010). Although seed size continues to be an issue for some *Arabidopsis* experiments, all plant embryos begin as single cells, which means that analyzing trace materials during early embryo development will continue to present unique challenges, regardless of the final size of the embryo at maturity.

Ultimately, the most powerful approach to the large-scale analysis of mutants defective in seed development involves identifying the disrupted genes. Although some *EMB* genes have been identified through map-based cloning, most were identified by amplifying genomic sequences flanking insertion sites in T-DNA tagged mutants. Overall, 80% of the *emb* mutants found in the SeedGenes database were generated with T-DNA insertional mutagenesis compared with 9% with transposable elements and 9% with EMS. Advances in polymerase chain reaction (PCR)–based strategies for insertion site recovery played a critical role in identifying large numbers of genes required for seed development in *Arabidopsis*. For *EMB* genes analyzed in my laboratory, *EMB* numbers 1000 through 2750 denote genes uncovered through forward genetic screens of Syngenta insertion lines; *EMB* numbers 2761 through 2820 indicate genes uncovered through reverse genetic screens, often involving Salk insertion lines; *EMB* numbers 3002 to 3013 correspond to French insertion lines (Devic, 2008); and *EMB* numbers 3101 to 3147 correspond to lines first tested in the laboratory of Kazuo Shinozaki at the Riken Plant Science Center in Japan (Bryant *et al.*, 2011).

Strategies for Approaching Saturation

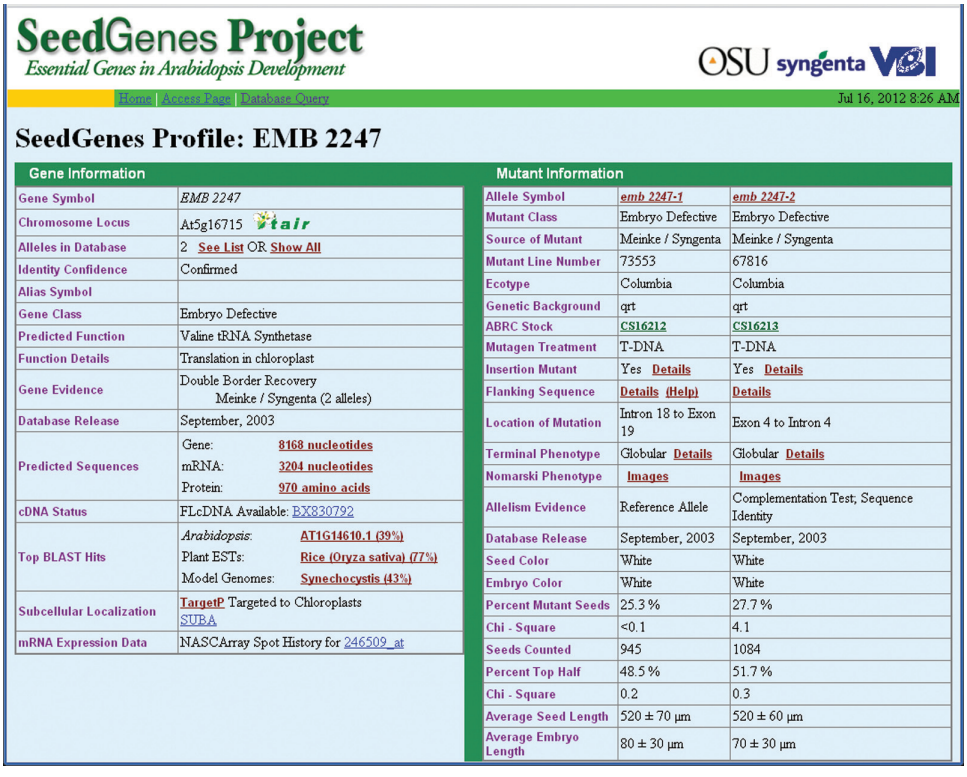
Forward genetics eventually becomes an inefficient strategy for identifying *EMB* genes because many of the new mutants examined represent alleles of known *EMB* genes. This trend has already been observed in *Arabidopsis*, with duplicate mutant alleles frequently encountered in mapped populations (Franzmann *et al.*, 1995; Meinke *et al.*, 2009b) and sequenced insertion lines (McElver *et al.*, 2001; Tzafrir *et al.*, 2004). A substantial number of mutants analyzed in detail in other laboratories have also turned out to be allelic to mutants first identified in my laboratory. About 8 years ago, we began to explore reverse genetic strategies for approaching saturation by focusing on *EMB* gene candidates not found through forward genetics. Promising candidates included *Arabidopsis* orthologs of known essential genes in other model organisms (Tzafrir *et al.*, 2004); genes encoding proteins that function in a shared biosynthetic pathway (Muralla *et al.*, 2007, 2008), cellular process (Berg *et al.*, 2005), or intracellular compartment (Bryant *et al.*, 2011) as a known EMB protein; and genes encoding a protein interactor of a known *EMB* gene product. We also analyzed hundreds of insertion lines that appeared from other studies (O'Malley and Ecker, 2010) to lack insertion homozygotes (Meinke *et al.*, 2008), which we reasoned might indicate embryo or gametophyte lethality. Although dealing with insertion lines on a large scale can be problematic, dozens of additional *EMB* genes were identified through a combination of these approaches. Reverse genetics was also used to find second alleles of genes first identified through forward genetics. When accompanied by genetic complementation tests, these additional alleles confirmed that the gene responsible for the mutant phenotype had been identified. The most difficult problem with Salk insertion lines (Alonso *et al.*, 2003) was reduced expression of the kanamycin-resistance marker, which meant that efficient methods developed for demonstrating close linkage between the disrupted gene and mutant phenotype based on selection, transplantation, and screening protocols (McElver *et al.*, 2001) were replaced by PCR genotyping, which is more expensive and subject to errors. In a substantial number of cases,

the predicted insert could not be found or did not cosegregate with the phenotype. Similar problems with large-scale screens of Salk insertion lines have been described elsewhere (Ajjawi *et al.*, 2010). Populations of insertion lines with a more consistent selectable marker, including the GABI (Rosso *et al.*, 2003) and Riken (Kuromori *et al.*, 2004) collections, were more efficiently analyzed, but unexplained results were still encountered, and decisions had to be made about whether to resolve the ambiguities or move ahead with additional candidates. Some *EMB* candidates confirmed with reverse genetics also turned out to be the subject of ongoing studies in other laboratories, which meant that unwanted duplication of effort was involved. Because of these added complications, we eventually abandoned reverse genetic analysis on a large scale and began to focus instead on further analysis of the existing collection of *EMB* genes.

SeedGenes Database of Essential Genes in *Arabidopsis*

One goal of my NSF 2010 project was to establish a public database that summarized information on genes required for seed development in *Arabidopsis*. The resulting database (www.seedgenes.org) was first released in 2002 and has since been updated multiple times. Allan Dickerman at the Virginia Bioinformatics Institute assisted with construction of the database and oversees its maintenance. The most recent (eighth) database release (December 2010) includes information on 481 genes and 888 mutants. Over 60% of the mutants have been analyzed in my laboratory. Information about the remaining mutants was extracted from the literature. Three classes of mutants are included in the database: embryo defectives, mutants with a pigment-defective embryo (albino, pale green, fusca) of normal morphology, and mutants that produce 50% rather than 25% defective seeds after self-pollination. On entering the database, users encounter the "Access Page," which provides links to lists of genes and mutants found in the database, supplemental and archival datasets, additional information on mutant collections, a tutorial on analyzing embryo-defective mutants, and details on project objectives and participants. The linked "Query Page" is divided into two different parts: gene information and mutant information. Users can browse a list of all genes or mutants, determine which genes of interest are included in the database, and generate lists of genes or mutants that match desired criteria. Database terms are linked to a glossary that provides further details. Each gene is associated with a "Profile" page, which summarizes relevant gene information on the left side of the page and mutant information on the right side. Figure 1.1 shows an example of a Profile page. From this page, users can link to further details on insertion sites for Syngenta mutants, phenotype details and Nomarski images for mutants analyzed in my laboratory, relevant sequence information, and top BLAST hits. Dividing the database into distinct but connected sections for gene and mutant information was critical for data management and represents a key design feature that could be used to develop similar databases for other model plants.

Deciding how to present phenotype information in the database was a major challenge, in part because the project evolved over a period of years and involved multiple student assistants with varying degrees of expertise. For Syngenta lines and mutants that my laboratory analyzed in some detail, we established a standardized set of terminal phenotypes (Figure 1.2) based on embryo morphology as visualized under a dissecting microscope. We also captured Nomarski images of mutant embryos inside the developing seed (Figure 1.3). Details of these methods are given at the tutorial section of the SeedGenes website. Although this approach provided insights into the stage of developmental arrest and the diversity of embryo phenotypes observed, subtle differences in cell division patterns and defects that first distinguished mutant from wild-type embryos were generally not recorded. The SeedGenes database should therefore be viewed as a broad community resource



SeedGenes Project
Essential Genes in Arabidopsis Development

OSU syngenta VBI

Home | Access Page | Database Query

Jul 16, 2012 8:26 AM

SeedGenes Profile: EMB 2247

Gene Information	
Gene Symbol	*EMB 2247*
Chromosome Locus	At5g16715 tair
Alleles in Database	2 See List OR Show All
Identity Confidence	Confirmed
Alias Symbol	
Gene Class	Embryo Defective
Predicted Function	Valine tRNA Synthetase
Function Details	Translation in chloroplast
Gene Evidence	Double Border Recovery Meinke / Syngenta (2 alleles)
Database Release	September, 2003
Predicted Sequences	Gene: 8168 nucleotides mRNA: 3204 nucleotides Protein: 970 amino acids
cDNA Status	FLcDNA Available: BX830792
Top BLAST Hits	*Arabidopsis*: AT1G14610.1 (39%) Plant ESTs: Rice (Oryza sativa) (77%) Model Genomes: Synechocystis (43%)
Subcellular Localization	TargetP Targeted to Chloroplasts SUBA
mRNA Expression Data	NASCArray Spot History for 246509_at

Mutant Information		
Allele Symbol	*emb 2247-1*	*emb 2247-2*
Mutant Class	Embryo Defective	Embryo Defective
Source of Mutant	Meinke / Syngenta	Meinke / Syngenta
Mutant Line Number	73553	67816
Ecotype	Columbia	Columbia
Genetic Background	qrt	qrt
ABRC Stock	CS16212	CS16213
Mutagen Treatment	T-DNA	T-DNA
Insertion Mutant	Yes Details	Yes Details
Flanking Sequence	Details (Help)	Details
Location of Mutation	Intron 18 to Exon 19	Exon 4 to Intron 4
Terminal Phenotype	Globular Details	Globular Details
Nomarski Phenotype	Images	Images
Allelism Evidence	Reference Allele	Complementation Test; Sequence Identity
Database Release	September, 2003	September, 2003
Seed Color	White	White
Embryo Color	White	White
Percent Mutant Seeds	25.3 %	27.7 %
Chi - Square	<0.1	4.1
Seeds Counted	945	1084
Percent Top Half	48.5 %	51.7 %
Chi - Square	0.2	0.3
Average Seed Length	520 ± 70 µm	520 ± 60 µm
Average Embryo Length	80 ± 30 µm	70 ± 30 µm

Figure 1.1 Screen capture of the Profile page for a representative gene (*EMB2247*) included in the SeedGenes database of essential genes in *Arabidopsis* (www.seedgenes.org). In this case, both mutant alleles were identified through a forward genetic screen of T-DNA insertion lines generated at Syngenta. Underlined and colored terms shown here link to other pages, primarily within the SeedGenes database.

Figure 1.2 Classification system for terminal phenotypes of mutant embryos removed from seeds before desiccation. For most of the Syngenta mutants analyzed in the early stages of our NSF 2010 project, we dissected 100 mutant seeds from siliques that contained normal seeds at a mature green stage of development. We then attempted to place each mutant embryo into one of the phenotype classes shown here. For stage "X," no mutant embryo was found on dissection; this usually meant that the mutant embryo was arrested before a late globular stage. Results of these phenotype screens are found by clicking on the "Details" link in the "Terminal Phenotype" section of the SeedGenes Profile page.

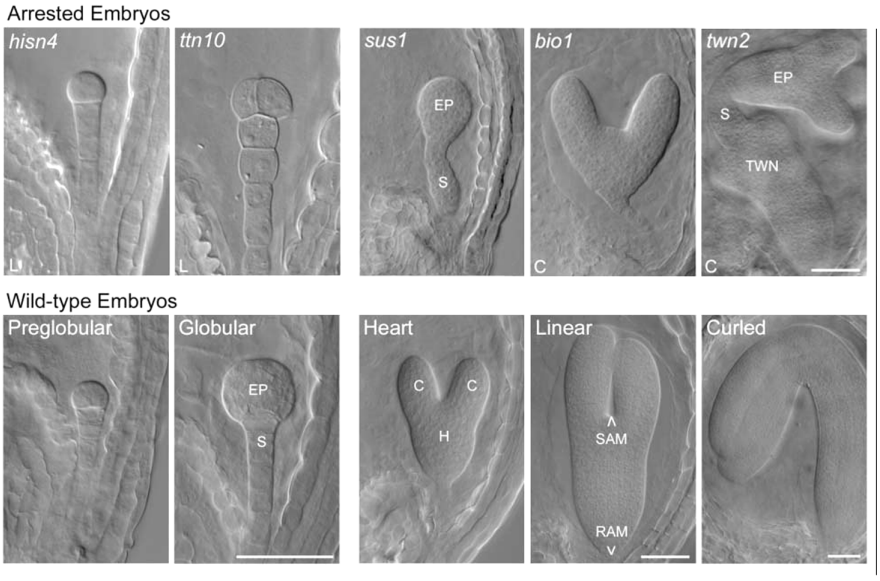


Figure 1.3 Representative collection of embryo-defective phenotypes found in the SeedGenes database. Regions of wild-type embryos include the embryo proper (EP), suspensor (S), cotyledons (C), hypocotyl (H), shoot apical meristem (SAM), and root apical meristem (RAM). Examples of aberrant development include irregular patterns of cell division, altered embryo morphology, giant suspensors, and twin embryos. The second (TWN) embryo in *twn2* arises from the suspensor (S) of the first embryo (EP). Seeds were removed from immature siliques and visualized with Nomarski (DIC) light microscopy. Arrested (mutant) embryos were obtained from heterozygous siliques at the linear (L) or curled cotyledon (C) stage of seed development. The four images on the left side are more highly magnified than the images on the right. Scale bars, 50 μm. (Modified from Meinke *et al.* [2008].)

and a starting point for additional studies rather than a definitive source of detailed phenotype information on individual mutants of interest.

Embryo Mutants with Gametophyte Defects

The more we began to characterize mutants defective in embryo development, the more important it became to distinguish between embryo and gametophyte mutants. Some gametophyte mutants of *Arabidopsis* are leaky, resulting in embryo lethality whenever fertilization takes place. In addition, some embryo mutants exhibit reduced transmission of the mutant allele and noticeable defects in gametophyte development. This raises a fundamental question: How can mutant (*emb*) gametophytes survive if an essential function required throughout the life cycle is disrupted? In other words, why do some essential gene disruptions in *Arabidopsis* result in gametophyte lethality, whereas others lead to embryo lethality?

To address this question, we first had to establish a comprehensive dataset of gametophyte essential genes in *Arabidopsis* that could be compared with the embryo dataset presented at SeedGenes. My laboratory recently created such a dataset, further edited the SeedGenes collection of *EMB* genes, and established several different categories of embryo and gametophyte mutants to facilitate comparative studies (Muralla *et al.*, 2011). The edited *EMB* dataset, which excluded six problematic SeedGenes loci, provides detailed information, including terminal phenotype classes, for 396 *EMB* genes in *Arabidopsis*. This dataset includes 352 "true *EMB*" genes required for seed development but without

a known gametophyte defect and 44 genes assigned to the EMG (Embryo-Gametophyte) subclass of embryo and gametophyte loci, which produce at least 10% defective seeds following self-pollination of heterozygotes and have a reduced frequency of mutant seeds overall, too few mutant seeds at the base of the silique, or an excessive number of aborted ovules, all of which indicate a secondary role in male or female gametophyte function. Genes assigned to the GEM (Gametophyte-Embryo) subclass of gametophyte mutants have a more significant defect in gametophyte function, with heterozygotes known or predicted to produce 2%–10% mutant seeds. The GAM (Gametophyte) subclass of mutants is characterized by even more severe defects in gametophyte function, with $<$2% mutant seeds expected from selfed heterozygotes. Other gametophyte mutants have more variable or less well-defined defects or give rise to viable homozygotes.

To examine the functional differences among these mutant classes, we compared 70 GAM genes with reduced transmission efficiency, 352 true *EMB* loci, and 72 EMG and GEM genes with defects in both embryo and gametophyte development (Muralla *et al.*, 2011). The difference between embryo and gametophyte mutants could not be explained based on protein function alone, although distinctive features of each dataset were identified. Two alternative explanations for how mutants defective in embryo development might survive gametophyte development were also discounted because neither genetic redundancy nor residual protein function in weak mutant alleles appeared to explain the different phenotypes observed. Instead, we proposed that residual gene products derived from transcription in heterozygous microsporocytes and megasporocytes often enable mutant gametophytes to survive the loss of an essential gene product and participate in fertilization, after which time the gene disruption eventually limits embryo growth and development (Muralla *et al.*, 2011).

General Features of *EMB* Genes in *Arabidopsis*

The first question about *EMB* genes that needs to be addressed concerns how many such loci are present in the genome. Our best estimate, based on the frequency of seed mutants and duplicate mutant alleles uncovered in mutagenesis experiments, is 750–1000 genes required for seed development in *Arabidopsis* (Meinke *et al.*, 2009b), which corresponds to about 3% of all protein-coding sequences. The current collection of 400 *EMB* genes likely represents at least 40% saturation, sufficient to begin evaluating salient features. Most *EMB* genes are not embryo-specific in their pattern of expression. Embryo development is simply the stage of development when the loss of gene product first becomes critical. Consistent with this idea, weak alleles of many *EMB* genes exhibit phenotypes later in plant development (Muralla *et al.*, 2011). *EMB* genes are widely distributed throughout the five chromosomes and are more likely than the genome as a whole to be present in a single copy. When functionally redundant genes encode a protein required for embryo development, the mutant phenotype is observed only in double or multiple mutants. Examples of such double mutant phenotypes have increased in recent years, reflecting greater emphasis on the use of reverse genetics to study essential cellular functions.

A second question about *EMB* genes concerns the stage of development reached by mutant embryos before seed desiccation. We recently summarized this information for 352 "true" *EMB* genes without evidence of gametophyte defects (Muralla *et al.*, 2011). This analysis updated information published before, using smaller datasets (Tzafrir *et al.*, 2004; Devic, 2008). Based on phenotype data for the strongest allele, 16% of gene disruptions cause embryo development to become arrested at a preglobular stage; 10%, at a preglobular to globular stage; 29%, at the globular stage; and 9% at the transition (heart) stage. Several examples are shown in Figure 1.3. Mutant embryos in 31%