SEED GENOMICS

Edited by
Philip W. Becraft

WILEY-BLACKWELL

Contents

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.

other expert assistance is required, the services of a competent professional should be sought.

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

Disclaimer

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.	The Integrative Cell Biology Laboratory, School of Biological and

Kumari Fonseka	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 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 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.

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.

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

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