Review

Identifying essential genes in *Arabidopsis thaliana*

David Meinke[1], Rosanna Muralla[1], Colleen Sweeney[1] and Allan Dickerman[2]

[1] Department of Botany, Oklahoma State University, Stillwater, OK 74078, USA
[2] Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Eight years after publication of the *Arabidopsis* genome sequence and two years before completing the first phase of an international effort to characterize the function of every *Arabidopsis* gene, plant biologists remain unable to provide a definitive answer to the following basic question: what is the minimal gene set required for normal growth and development? The purpose of this review is to summarize different strategies employed to identify essential genes in *Arabidopsis*, an important component of the minimal gene set in plants, to present an overview of the datasets and specific genes identified to date, and to discuss the prospects for future saturation of this important class of genes. The long-term goal of this collaborative effort is to facilitate basic research in plant biology and complement ongoing research with other model organisms.

The concept of essential genes

Essential is a common word that does not often require further explanation. Defining essential in the context of plant growth and development, however, can be problematic because one must differentiate between essential cellular processes, essential protein functions, and essential genes, and recognize that what is essential under one set of growth conditions may be dispensable under another. One might even argue that most genes are essential because otherwise they would not be maintained through natural selection. For this article, we have chosen the following definition of an essential gene: one that is required for normal growth and development and is associated with a loss-of-function phenotype in a standard genetic background. Duplicated genes that encode essential proteins with redundant functions and expression patterns are not considered to be essential according to this definition and will not be detected in phenotypic screens unless the appropriate multiple mutants are constructed. Synthetic lethals that exhibit a phenotype only when mutations in distinct but interacting genes are combined will also be missed. The initial list of essential genes presented here therefore represents a valuable though incomplete sampling of the more comprehensive but elusive dataset of *Arabidopsis* genes that individually or collectively encode proteins with essential functions.

The importance of lethals

Lethal mutants are often considered to be of limited value because the mutant tissue needed for analysis cannot be studied or maintained. With *Arabidopsis*, this conclusion is somewhat misguided. First, recessive embryonic lethals can be readily maintained as heterozygotes that repeatedly exhibit the desired phenotype with each new silique (fruit) produced. Second, valuable information can often be obtained by analyzing mutant embryos before the onset of lethality. This approach has revealed important details of the null phenotypes of many essential genes, including the *Arabidopsis DICER* ortholog (*DCL1*) required to generate small RNAs [1]. Third, advances in technology have made possible the analysis of small amounts of plant material. Mutant embryos are therefore not beyond experimental manipulation [2]. And fourth, alternative methods can be used to generate weak alleles of an essential gene [3]. The literature is full of examples in which embryonic and gametophytic lethals have provided valuable information on topics ranging from cell biology and metabolism to signal transduction and developmental interactions. Lethals are therefore a crucial and informative part of the *Arabidopsis* mutant collection.

The systematic identification of genes with essential functions has been described for several different prokaryotes [4–6] and a number of eukaryotes, including *Saccharomyces* [7], *Caenorhabditis* [8], and humans [9]. These studies have provided valuable insights into the minimal gene set required for basic cell functions in a wide range of organisms. An equivalent dataset for *Arabidopsis* would enable comparisons with other model organisms, facilitate the analysis of plant genes with important but otherwise unknown functions, and contribute to our understanding of essential biological processes in flowering plants. The immediate challenge is to devise effective strategies for finding and confirming the identities of essential plant genes. We have chosen to focus here on genes required for seed development, and to some extent gametogenesis, because they represent the most robust dataset of essential plant genes available. A list of 620 *Arabidopsis* genes with mutant phenotypes detected throughout the life cycle was published five years ago [10]. Mutants with a seed phenotype were the most common class represented.

Forward genetic screens

The history of *Arabidopsis* genetics is filled with examples of genetic screens for informative mutant phenotypes. The first publication devoted to lethal mutants appeared 50 years ago [11]. Screens for embryonic lethals were described in detail 5 and 20 years later [12,13], followed by seedling screens for mutants with altered patterns of

Corresponding author: Meinke, D. (david.meinke@okstate.edu).

1360-1385/$ – see front matter doi:10.1016/j.tplants.2008.06.003 Available online 4 August 2008

embryo development [14]. Thousands of embryo-defective (*emb*) mutants have been isolated and characterized in dozens of laboratories over the years. Many of the disrupted genes remain to be identified. Mutants are typically maintained as heterozygotes that produce siliques with 25% defective seeds. Allelism between mutants can be confirmed by crossing heterozygotes and screening the resulting siliques for defective seeds.

The advent of T-DNA insertional mutagenesis enabled large-scale screens of insertion lines for tagged mutants with defects in seed development [15] attributed to known genes [16]. This led to the development of a centralized database (http://www.seedgenes.org) of essential genes associated with a seed phenotype [17]. The December 2007 database release includes more than 350 essential genes and over 600 mutant alleles identified in different laboratories and through a variety of experimental approaches (Table 1). Most of the gene identities resulted from forward screens of T-DNA insertion lines [15]. The frequency of embryo-defective mutants uncovered in these screens is high because many genes are required for normal seed development. However, only ~30% of seed phenotypes observed in T-DNA lines appear to result from stable integration. Considerable effort is therefore required to determine which mutants are tagged with an insert and to identify large numbers of *EMB* genes in these populations [15]. Forward genetic screens also become less efficient as saturation is approached because the mutants uncovered are likely to represent new alleles of known genes. Identifying the full spectrum of essential genes in *Arabidopsis* will therefore require both forward and reverse genetics.

Gametophytic lethals

Forward genetic screens for gametophytic mutants have often focused on reduced transmission of an associated selectable marker [18–20]. Plants heterozygous for a recessive mutation that is lethal to male or female gametophytes (but not both) should transmit the mutant allele to 50% rather than 75% of progeny seeds following self-pollination. This difference can often be detected by scoring the ratio of resistant to sensitive seedlings derived from collections of insertion mutants. A more definitive strategy for the identification of male gametophytic lethals involves the *quartet* mutation, which interferes with separation of pollen tetrads. The advantage to this approach is that all four products of male meiosis remain attached, and hence the desired insertion mutants can be readily identified and distinguished from unwanted chromosomal aberrations [21]. The most extensive single collection of gametophytic mutants includes 130 candidate genes required for female gametophyte development [18]. Most of these genes are represented by single mutant alleles. Another 50 to 60 genes required for gametophyte function have been identified in publications from multiple laboratories over the past decade. Some of these encode proteins that contribute specialized functions required for pollen germination or pollen-tube growth.

Mutations that are 100% lethal to both male and female gametophytes cannot be maintained or studied. The total number of such genes in *Arabidopsis* remains unknown. Several candidates were recently identified among mutants altered in cytosolic aminoacyl-tRNA synthetases [22]. The implication is that a fundamental disruption of translation in the cytosol of developing gametophytes cannot be circumvented. Some mutations disrupt female gametophyte development but not pollen development or tube growth. These often result in heterozygotes with 50% aborted ovules. Other mutations interfere with male transmission but not female transmission. In some cases, both male and female transmission of a mutant allele is reduced but not eliminated. Rare homozygotes that form typically fail to complete embryogenesis.

Some mutants yield heterozygotes with 50% defective seeds regardless of pollen genotype, either because the wild-type allele contributed through the pollen is silenced during seed development or because defects in female gametophyte development become limiting after fertilization [23,24]. Disrupting the second pollen mitosis also generates siliques with 50% aborted seeds following self-pollination [25], although the underlying mechanism is different. Because the distinction between genes required for gametogenesis and those required for embryogenesis is not absolute, and because differences can be complicated by the presence of stored gene products in gametophytes, the most effective strategy for identifying essential genes with functions early in development is to consider embryonic and gametophytic lethals combined. Confirming the identities of large numbers of candidate genes required for gametogenesis nevertheless remains a challenge because allelism tests cannot be performed.

Table 1. The SeedGenes database (December 2007)

Total genes in database:	**358**
Gene identity	
- Confirmed	244
- Not confirmed	104
- Uncertain	5
- Questionable	5
Mutant phenotype	
- Embryo defective	317
- Seed pigment	35
- 50% defective seeds	6
Total mutants in database:	**605**
Mutant phenotype	
- Embryo defective	542
- Seed pigment	57
- 50% defective seeds	6
Nomarski images	
- Available	334
- Not available	271
Initial characterization	
- Meinke laboratory	381
- Other laboratories	224
Pending database additions:	**37 genes**
Double-mutant seed phenotype:	**31 gene pairs**

Informative mutant phenotypes

One strategy for dealing with essential genes is to focus attention on a small number of genes with interesting functions or knockout phenotypes and ignore the rest. This strategy has three major limitations: (i) cellular functions that seem mundane to one investigator may be fascinating to another; (ii) some proteins with unremarkable biochemical activities have surprising developmental func-

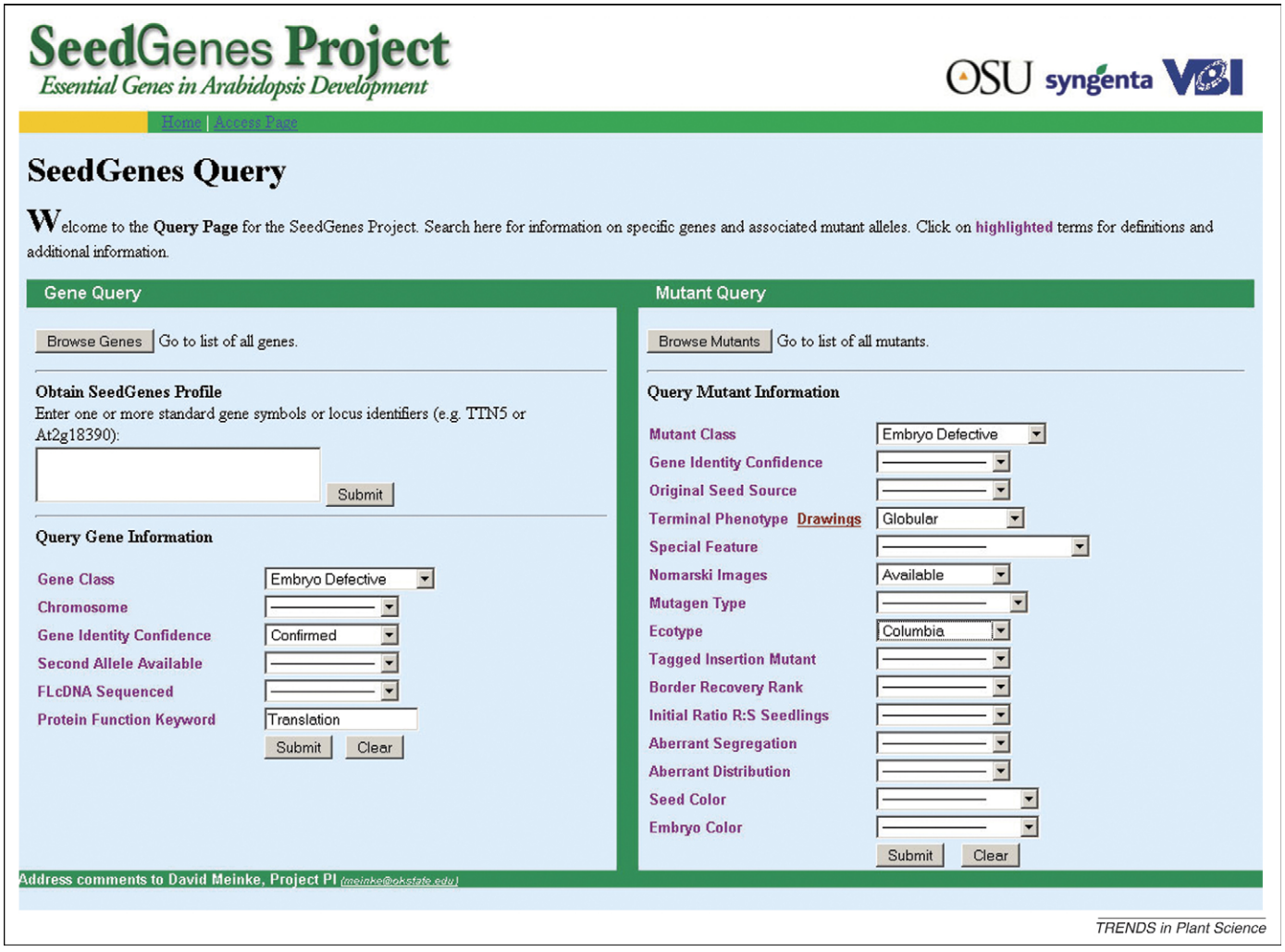

Figure 1. Screen capture of the Query page from the SeedGenes database available at http://www.seedgenes.org. Several query options (e.g. Embryo defective, Translation, Globular) have been selected for illustration. The December 2007 release of the database includes information on ~350 genes required for seed development in *Arabidopsis* and ~600 mutant alleles with known disruptions in these genes.

tions that become apparent after broad genetic screens; and (iii) unusual phenotypes do not always lead to informative gene functions. For example, the origin of twin embryos in *Arabidopsis* [26] is an intriguing developmental phenomenon that is amenable to genetic analysis. The first *twin* mutant cloned, however, was an atypical allele of an aminoacyl-tRNA synthetase [27]. Another *twin* mutant appears to be altered in a ribosomal protein [16]. Meanwhile, the regulatory factors that prevent twinning during normal development remain to be identified. We have long advocated the value of identifying large numbers of genes with a loss-of-function seed phenotype to assess the full range of functions associated with seed development [13]. The challenge then becomes to collect and organize the most relevant data on many different genes and mutant alleles.

The SeedGenes database

Detailed information on genes required for seed development in *Arabidopsis* is presented in the SeedGenes database (http://www.seedgenes.org) [17]. Individuals working on mutants with a seed phenotype or genes of interest that fail to generate knockout homozygotes are encouraged to consult this database and to cite when appropriate the mutant alleles described therein. Based on past estimates of 500 to 1000 total *EMB* genes in *Arabidopsis* [15,28] and the continued identification of new *EMB* genes throughout the community, the current collection of ~350 *EMB* genes is substantial but not yet close to saturation. Gene identities confirmed through molecular complementation or allelism tests are distinguished in SeedGenes from those that are not confirmed (single mutant alleles) or uncertain because of questionable insert location or conflicting genetic data. This distinction allows both preliminary and confirmed data on candidate genes to be included. Similar standards of identity confirmation need to be maintained while developing additional datasets of genes with knockout phenotypes in the future.

The SeedGenes Query page (Figure 1) generates a list of genes or mutants that match selected criteria. Clicking on a specific gene or mutant directs the user to a Profile page for that locus, with detailed gene information on the left side and mutant information on the right. Links provide additional documentation such as flanking sequence information for insertion mutants, summarized BLAST data, and images of cleared mutant seeds viewed with Nomarski optics. A representative collection of images is presented in Figure 2. Methods used to generate data in SeedGenes are

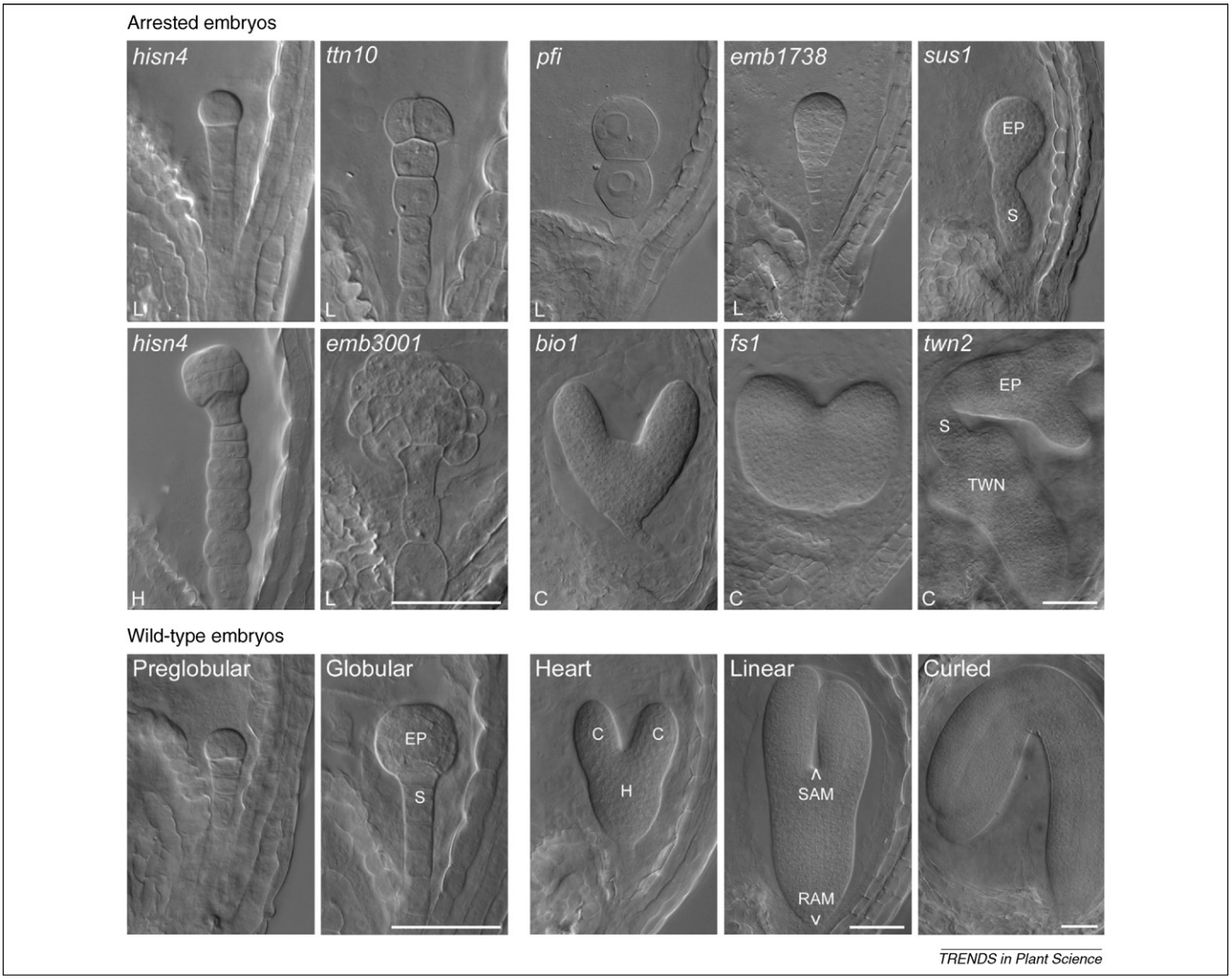

Figure 2. Representative collection of embryo-defective phenotypes found in the SeedGenes database. Defined regions of wild-type embryos include the embryo proper (EP), suspensor (S), cotyledons (C), hypocotyl (H), shoot apical meristem (SAM), and root apical meristem (RAM). Examples of aberrant development include irregular patterns of cell division, altered embryo morphology, enlarged cells (*pfi*), uneven epidermal layer (*emb3001*), giant suspensors (*sus1*), and twin embryos (*twn2*). 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 (differential interference contrast) light microscopy. Arrested (mutant) embryos were obtained from green siliques when wild-type embryos were at the heart (H), linear (L), or curled cotyledon (C) stage, as noted in the lower left of each figure. The six images on the left side are more highly magnified than those at right. Scale bars, 50 μm.

presented in a tutorial section on the website. Other SeedGenes features include lists of genes to be included in future releases, redundant gene pairs that are by definition not essential but generate a double-knockout phenotype, and seed stocks for 1412 additional *emb* mutants derived from Syngenta insertion lines [15] that are either not tagged with T-DNA or unresolved with respect to tagging status. Future database releases will include information on genes required for gametogenesis and summaries of essential gene functions and patterns of expression.

Features of known *EMB* genes

Genes required for seed development are randomly distributed throughout the genome and encode proteins with a wide range of biochemical and cellular functions. Basal processes such as DNA replication, RNA processing, and protein synthesis, which should become critical during embryo development, are prevalent among the seed phenotype class [16]. Transcription factors and components of signaling pathways are under-represented and are more common among genes with knockout phenotypes that affect later stages of development [16]. Surprisingly, the functional classifications of genes required for seed development overlap to a great extent with those required for gametogenesis. Whether a particular gene disruption results in embryo or gametophytic lethality therefore appears to be determined in part by the availability of stored gene products in gametophytes and not simply by the specific protein function involved [22]. Identifying these stored transcripts will be difficult because the current methods used to assess transcript diversity in gametophytes do not distinguish between post-meiotic transcripts derived from haploid spores and transcripts produced by pre-meiotic sporocytes or surrounding diploid cells. With respect to the intracellular compartmentalization of protein products, many *EMB* genes encode essential chloroplast proteins. A functional chloroplast is therefore required for normal embryo development in *Arabidopsis*.

By contrast, a complete loss of mitochondrial function often seems to be associated with zygotic or gametophytic lethality [22].

Overall, *EMB* genes identified to date are somewhat larger than average, consistent with their frequent isolation through forward genetics, and have fewer paralogs than average, which reduces the likelihood that a redundant gene will mask the loss-of-function phenotype [16]. Most *EMB* genes appear to be expressed at multiple stages of the life cycle. Thus, a seed phenotype does not necessarily indicate embryo-specific expression but rather the developmental stage when a loss of gene product first becomes limiting. With some notable exceptions [29,30], most *emb* mutants appear to be nulls, not simply weak alleles of genes required for gametogenesis. Weak alleles of some *EMB* genes exhibit distinctive phenotypes during later stages of development, consistent with their known patterns of expression throughout the life cycle. Examples include changes in flowering time (*FY*), ozone response (*VTC1*), root meristem development (*RML1*), and ovule morphology (*SIN1*).

Examples of recent gene identifications

The identities of 41 *EMB* genes [31–64] were published between January 2006 and February 2008 (Table 2). Most of these resulted from reverse screens of insertion mutants disrupted in known genes of interest. Another 33 *EMBs* were added to SeedGenes following reverse analysis of candidate essentials in the Meinke laboratory. During this same period, seven gametophytic lethals were identified through forward genetics [65,66] and 10 through reverse genetics [29,35,67–74]. Four other genes were found to be essential for gametophyte recognition during fertilization [75–79]. *EMBs* therefore continue to represent the predominant class of essential genes in *Arabidopsis*. Double knockouts with a lethal phenotype also increased dramatically, with gene identities published for 12 embryonic and eight gametophytic double mutants [35,80–97].

Several of the knockouts listed in Table 2 are noteworthy because they reveal overlaps between gene identification programs in different laboratories. The first *ise2* allele (*emb25*) was isolated following chemical mutagenesis and mapped relative to visible markers at the bottom of chromosome 1 [28]. A second allele was later identified by screening for mutants with altered size exclusion of plasmodesmata during embryogenesis [38]. Map-based cloning revealed that this locus (At1g70070) corresponded to *PDE317,* originally named for the pigment phenotype of a weak insertion mutant from the Syngenta collection [15]. In another case, a forward genetic screen for mutants with enhanced gene silencing [36] resulted in the identification of weak alleles of two *EMB* genes (At1g32490 and At5g23880) that encode proteins involved in RNA processing. Recent cloning of the *ICU2* locus [67] revealed that distinctive leaf and floral phenotypes of the original mutant allele resulted from a partial loss of function of At5g67100, which encodes the catalytic subunit of DNA polymerase α. Null alleles of this locus are gametophytic lethal. Another locus (At2g41500) was identified as essential based on genetic screens in three different laboratories: a forward screen for female gametophytic mutants [43], a reverse screen for knockouts of genes involved in RNA processing [44], and a reverse screen of *Arabidopsis* orthologs of known essential genes [16].

Several different strategies involving multiple laboratories have therefore resulted in the identification of a modest number of *EMB* genes over the past two years. By contrast, 160 *EMB* genes were first identified through one forward screen at Syngenta [15] and incorporated into the SeedGenes database between March 2002 and September 2004. We are not aware of current attempts elsewhere to identify large numbers of *EMB* genes through a similar approach. Most remaining *EMB* genes will therefore in all likelihood be uncovered gradually through reverse genetics. The question then becomes which target genes represent the most promising candidates for analysis.

Strategies for approaching saturation

The following sections highlight the reverse genetic strategies being used to identify candidate *EMB* genes in *Arabidopsis*. The long-term goal is not to screen every knockout available for a seed phenotype, which is both inefficient and unrealistic, but rather to focus attention on those genes most likely to be required for seed development. Promising insertions in genes of interest can be identified by querying the Salk database [98] of insertion mutants at http://signal.salk.edu. Although compiling a definitive list of non-essential genes that lack a knockout phenotype would also be helpful and informative, the inherent complexities of existing mutant collections make this goal impractical at present.

Orthologs of essential genes in other organisms

One source of candidate *EMB* genes is non-redundant orthologs of essential genes identified in other model organisms. From an initial dataset of 240 such genes identified in *Arabidopsis* [16], we selected 74 candidates for a study that involved screening 215 Salk insertion lines for the presence of a seed phenotype that correlated with the insert. Although this sample proved to be enriched for essentials, with 19 *EMB* genes identified, the failure to confirm some insertions and the high frequency of background mutations with a seed phenotype complicated the analysis. Several *EMBs* found with this approach also corresponded to genes identified elsewhere through concurrent screens. We therefore conclude that a 'shared essentials' approach to saturation is robust but may duplicate ongoing work in other laboratories.

Shared processes, pathways, and protein interactors

An alternative approach is to focus on candidates that share something in common with a known *EMB*. We pursued a 'shared process' approach with aminoacyl-tRNA synthetases (AARSs) when it became apparent that several *EMB* genes identified through forward genetic screens were required for the aminoacylation of tRNAs during translation [22]. The resulting screen of 50 insertion lines identified three additional *EMB* genes required for translation in chloroplasts and nine *OVA* genes with an aborted ovule phenotype that results from loss of translation in mitochondria. Candidates for future screens utilizing this

Table 2. Recent publications on *EMB* gene identification in *Arabidopsis* (January 2006 – February 2008)[a]

Locus[b]	Symbol	Alias[c]	Phenotype	Mutagen	Predicted gene function	Ref(s)
At1g03360	*RRP4*	-	Embryo	T-DNA	RNA processing; exosome subunit	[31]
At1g09770	*AtCDC5*	-	Embryo	T-DNA	DNA-binding protein; cell cycle control	[32]
At1g10270	*GRP23*	-	Embryo	TN; T-DNA	Putative transcriptional regulator	[33]
At1g23400	*AtCAF2*	-	Embryo	T-DNA	Chloroplast intron splicing factor	[34]
At1g31860	*HISN2*	-	Embryo	T-DNA	Histidine biosynthesis	[35]
At1g32490	*ESP3*	*EMB2733*	Silencing; embryo	EMS; T-DNA	RNA helicase; mRNA splicing	[36]
At1g62750	*SCO1*	-	Embryo	T-DNA	Plastid translation elongation factor	[37]
At1g64790	*ILA*	-	Embryo	T-DNA	Translational activator	[74]
At1g70070	*ISE2*	*EMB25; PDE317*	Embryo; pigment	EMS; T-DNA	RNA helicase; plasmodesmata function	[38]
At1g74960	*FAB1*	-	Embryo	T-DNA	Ketoacyl-acyl carrier protein synthase	[39]
At2g01350	*QPT1*	-	Embryo	TN	NAD biosynthesis	[40]
At2g17510	*RRP44A*	*EMB2763*	Embryo	T-DNA	RNA processing; exosome subunit	[31]
At2g21470	*SAE2*	*EMB2764*	Embryo	T-DNA	SUMO activating enzyme	[41]
At2g36230	*HISN3*	*APG10*	Embryo; pigment	T-DNA; TN	Histidine biosynthesis	[35]
At2g38670	*PECT1*	-	Embryo	EMS	Phosphatidylethanolamine biosynthesis	[42]
At2g41500	*LIS; PRP4*	*EMB2776*	FEG; embryo	EMS; T-DNA	snRNP; mRNA splicing	[43,44]
At3g14230	*AtRAP2.2*	-	Embryo	T-DNA	Transcription factor	[45]
At3g19770	*AtVPS9A*	-	Embryo	T-DNA	Rab5 guanine exchange factor	[46]
At3g55610	*AtP5CS2*	-	Embryo	T-DNA	Proline biosynthesis	[47]
At3g57150	*AtCBF5*	-	Embryo	T-DNA	Nucleolar protein; RNA processing	[48]
At3g57870	*SCE1*	*EMB1637*	Embryo	T-DNA	SUMO conjugating enzyme	[41]
At4g00220	*JLO*	-	Embryo	T-DNA; TN	LOB domain protein	[49]
At4g03240	*AtFH*	-	Embryo	T-DNA	Frataxin; biosynthesis of Fe-S proteins	[50,51]
At4g21800	*QQT2*	-	Embryo	T-DNA	ATP/GTP binding protein; microtubules	[52]
At4g22970	*AESP*	-	Embryo	T-DNA	Separase; sister chromatid separation	[53]
At4g26500	*AtSufE*	*EMB1374*	Embryo	T-DNA	Fe-S cluster protein	[54]
At4g26900	*HISN4*	-	Embryo	T-DNA	Histidine biosynthesis	[35]
At4g31780	*MGD1*	*EMB2797*	Embryo	T-DNA	MGDG (galactolipid) synthesis	[55]
At4g32720	*AtLA1*	-	Embryo	T-DNA	RNA binding protein	[56]
At4g33495	*RPD1*	-	Root; embryo	T-DNA	Unknown	[57]
At4g36480	*LCB1*	*EMB2779*	Embryo	T-DNA	Sphingolipid biosynthesis	[58]
At5g14760	*AO*	-	Embryo	T-DNA	NAD biosynthesis	[40]
At5g22370	*QQT1*	*EMB1705*	Embryo	T-DNA	ATP binding protein; microtubules	[52]
At5g23880	*ESP5*	*EMB1265*	Silencing; embryo	EMS; T-DNA	mRNA cleavage and polyadenylation	[36]
At5g48600	*SMC4*	-	Embryo	T-DNA	Chromosome condensation	[59]
At5g48840	*AtPTS*	-	Embryo	T-DNA	Pantothenate synthetase	[60]
At5g49160	*MET1*	-	Embryo	EMS	Methyltransferase; DNA methylation	[61]
At5g50210	*QS*	-	Embryo	T-DNA	NAD biosynthesis	[40]
At5g52920	*PKP1*	-	Embryo	T-DNA	Plastidic pyruvate kinase B1 subunit	[62]
At5g57600	*BIO3*	-	Embryo	T-DNA	Biotin synthesis; bifunctional enzyme	[63]
At5g59440	*ZEUS1*	-	Embryo	T-DNA	Thymidylate kinase; DNA replication	[64]

[a]Green symbols denote essential genes revealed through forward genetics and red symbols through reverse genetics. Blue symbols represent genes with other phenotypes revealed through forward genetics and embryo phenotypes noted through reverse genetics. Abbreviations: FEG, female gametophyte; TN, transposon; EMS, ethyl methanesulfonate; MGDG, monogalactosyldiacylglycerol; T-DNA, transferred DNA from *Agrobacterium tumefaciens*.

[b]Excludes 33 other *EMBs* identified in the Meinke laboratory and listed at http://www.seedgenes.org during this time, either in the central database (At2g17250, At2g38770, At2g43650, At2g45000, At3g13200, At4g03430, At4g11820, At5g05560, At5g14800, At5g15540, At5g27740, At5g63960) or linked to the 'pending additions' page (At1g04950, At1g07320, At1g24706, At1g28395, At1g49870, At2g02150, At2g18290, At2g31060, At2g32590, At2g39080, At2g43650, At3g10220, At3g17300, At3g23110, At3g46960, At4g27010, At4g29910, At4g36690, At5g05680, At5g15920, At5g56290).

[c]All *emb* and *pde* mutants listed here were characterized in the Meinke laboratory. The *apg10* mutant was described elsewhere [104].

'shared process' approach include genes associated with DNA replication, RNA processing, and ribosome assembly.

A related strategy is to focus on basic metabolic pathways required for seed development. The histidine pathway is of particular interest because it involves multiple steps and has long been a paradigm for gene regulation in bacteria. After forward genetics revealed that disruption of histidine biosynthesis resulted in embryo lethality [16], we took a reverse genetic approach that culminated in the identification of four additional *EMB* genes, one gametophytic lethal, and one double-knockout lethal [35]. A similar approach was utilized to disrupt an intermediate step in biotin biosynthesis [63]. This candidate gene (At5g57600) ultimately defined a bifunctional locus (*BIO3–BIO1*) that undergoes differential splicing and includes another gene (At5g57590) required for biotin biosynthesis [15]. Although these examples illustrate the success of a 'shared pathway' approach to *EMB* gene identification, the complexity and redundancy of metabolic pathways in plants may limit the strategy overall.

Proteins that form complexes with known *EMB* gene products represent another source of candidates. The assumption here is that disrupting any portion of the complex will result in lethality. This 'shared interactors' approach was recently used to demonstrate that the *QQT2* gene product, known to interact with another protein (QQT1; EMB1705) identified through forward genetics, is also required for completion of embryo development [52]. Protein interactors that form complexes conserved throughout eukaryotes can be discovered most readily [99,100], although some of these candidates will duplicate those identified with other methods. Realizing the full

potential of this approach to identify plant-specific complexes with proteins of unknown functions will require a more complete characterization of the protein interactome in *Arabidopsis*.

Genes expressed in embryos or gametophytes

Another strategy is to select promising candidates from lists of genes known to be expressed in embryos or female gametophytes. The obvious limitation to this approach is the technical challenge of identifying rare transcripts in small samples. One idea has been to analyze mutant ovules lacking a megagametophyte [74,101] and look for missing transcripts present in the wild-type ovule. One study found several known essentials among megagametophyte transcripts identified in this manner but analyzed only two essential genes not previously described [74]. Problems were also encountered with chromosomal translocations, which complicate the analysis of T-DNA mutants [28]. Another study found several known essential genes among transcripts identified using laser microdissection of developing embryos [102]. With continued technical advances in mRNA isolation and characterization, these combined strategies may focus attention on the most promising candidates for future analysis.

Absence of knockout homozygotes

We recently examined 130 candidate genes associated with insertion mutants that fail to generate knockout homozygotes. This strategy was designed to complement a genome-wide effort to identify a knockout homozygote for every *Arabidopsis* gene (see http://signal.salk.edu). We expected that this approach would be straightforward because most genes that fail to yield knockout homozygotes should be essential. However, only 25% of the genes screened to date appear promising. The high frequency of false positives reflects a variety of problems, including sampling and genotyping errors, PCR primers that amplify a second locus, inserts missing from plants sampled, difficulty predicting allele severity, and inherent complexities of T-DNA lines. The apparent absence of knockout homozygotes in genome-wide screens is therefore not yet a reliable indicator of genes with essential functions.

Future directions

We have attempted to document here both the remarkable progress made in identifying essential genes of *Arabidopsis* and the potential limitations of different strategies for reaching saturation. Although T-DNA insertional mutagenesis made possible the large-scale identification of essential genes, the unpredictable nature of existing collections of insertion lines and the molecular complexities of many insertion sites appear to define the greatest obstacle to future saturation. Nevertheless, we believe that by pursuing the different strategies described here, incorporating future advances in targeted gene inactivation, examining genes without insertions [103] as candidates for dual gametophytic lethality [22], and assembling information from multiple laboratories on genes that fail to produce viable knockout homozygotes, it should be possible to make continued progress towards defining a comprehensive dataset of essential genes in a model plant.

Acknowledgements

We thank current and past members of the Meinke laboratory (OSU), the Dickerman laboratory (VBI), and the Patton laboratory (Syngenta) for many valuable contributions to the SeedGenes project, which has been supported by the National Science Foundation (NSF) 2010 program.

References

1 Schauer, S.E. *et al.* (2002) *DICER-LIKE1*: blind men and elephants in *Arabidopsis* development. *Trends Plant Sci.* 7, 487–491
2 Sauer, M. and Friml, J. (2004) *In vitro* culture of *Arabidopsis* embryos within their ovules. *Plant J.* 40, 835–843
3 Henikoff, S. *et al.* (2004) TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol.* 135, 630–636
4 Ji, Y. *et al.* (2001) Identification of critical *Staphylococcal* genes using conditional phenotypes generated by antisense RNA. *Science* 293, 2266–2269
5 Akerley, B.J. *et al.* (2002) A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 966–971
6 Kobayashi, K. *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4678–4683
7 Giaever, G. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391
8 Kamath, R.S. *et al.* (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231–237
9 Silva, J.M. *et al.* (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science* 319, 617–620
10 Meinke, D.W. *et al.* (2003) A sequence-based map of *Arabidopsis* genes with mutant phenotypes. *Plant Physiol.* 131, 409–418
11 Langridge, J. (1958) A hypothesis of developmental selection exemplified by lethal and semi-lethal mutants of *Arabidopsis*. *Aust. J. Biol. Sci.* 11, 58–68
12 Müller, A.J. (1963) Embryonentest zum Nachweis rezessiver Letalfaktoren bei *Arabidopsis thaliana*. *Biol. Zentralbl.* 82, 133–163
13 Meinke, D.W. and Sussex, I.M. (1979) Embryo-lethal mutants of *Arabidopsis thaliana*: a model system for genetic analysis of plant embryo development. *Dev. Biol.* 72, 50–61
14 Mayer, U. *et al.* (1991) Mutations affecting body organization in the *Arabidopsis* embryo. *Nature* 353, 402–407
15 McElver, J. *et al.* (2001) Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* 159, 1751–1763
16 Tzafrir, I. *et al.* (2004) Identification of genes required for embryo development in *Arabidopsis*. *Plant Physiol.* 135, 1206–1220
17 Tzafrir, I. *et al.* (2003) The *Arabidopsis* SeedGenes project. *Nucleic Acids Res.* 31, 90–93
18 Pagnussat, G.C. *et al.* (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* 132, 603–614
19 Procissi, A. *et al.* (2001) Five gametophytic mutations affecting pollen development and pollen tube growth in *Arabidopsis thaliana*. *Genetics* 158, 1773–1783
20 Lalanne, E. *et al.* (2004) Analysis of transposon insertion mutants highlights the diversity of mechanisms underlying male progamic development in *Arabidopsis*. *Genetics* 167, 1975–1986
21 Johnson, M.A. *et al.* (2004) *Arabidopsis hapless* mutations define essential gametophytic functions. *Genetics* 168, 971–982
22 Berg, M. *et al.* (2005) Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in *Arabidopsis*. *Plant J.* 44, 866–878
23 Leroy, O. *et al.* (2007) Polycomb group proteins function in the female gametophyte to determine seed development in plants. *Development* 134, 3639–3648
24 Gehring, M. *et al.* (2006) *DEMETER* DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124, 495–506
25 Iwakawa, H. *et al.* (2006) *Arabidopsis CDKA;1*, a *cdc2* homologue, controls proliferation of generative cells in male gametogenesis. *Plant J.* 45, 819–832
26 Vernon, D.M. and Meinke, D.W. (1994) Embryogenic transformation of the suspensor in *twin*, a polyembryonic mutant of *Arabidopsis*. *Dev. Biol.* 165, 566–573
27 Zhang, J.Z. and Somerville, C.R. (1997) Suspensor-derived polyembryony caused by altered expression of valyl-tRNA

synthetase in the *twn2* mutant of *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7349–7355
28 Franzmann, L.H. *et al.* (1995) Saturating the genetic map of *Arabidopsis thaliana* with embryonic mutations. *Plant J.* 7, 341–350
29 Persson, S. *et al.* (2007) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15566–15571
30 Teng, C. *et al.* (2008) Serine palmitoyltransferase, a key enzyme for *de novo* synthesis of sphingolipids, is essential for male gametophyte development in *Arabidopsis*. *Plant Physiol.* 146, 1322–1332
31 Chekanova, J.A. *et al.* (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell* 131, 1340–1353
32 Lin, Z. *et al.* (2007) AtCDC5 regulates the G2 to M transition of the cell cycle and is critical for the function of *Arabidopsis* shoot apical meristem. *Cell Res.* 17, 815–828
33 Ding, Y.H. *et al.* (2006) *Arabidopsis GLUTAMINE-RICH PROTEIN23* is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. *Plant Cell* 18, 815–830
34 Asakura, Y. and Barkan, A. (2006) *Arabidopsis* orthologs of maize chloroplast splicing factors promote splicing of orthologous and species-specific group II introns. *Plant Physiol.* 142, 1656–1663
35 Muralla, R. *et al.* (2007) Genetic dissection of histidine biosynthesis in *Arabidopsis*. *Plant Physiol.* 144, 890–903
36 Herr, A.J. *et al.* (2006) Defective RNA processing enhances RNA silencing and influences flowering of *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14994–15001
37 Ruppel, N.J. and Hangarter, R.P. (2007) Mutations in a plastid-localized elongation factor G alter early stages of plastid development in *Arabidopsis thaliana*. *BMC Plant Biol.* 7, 37
38 Kobayashi, K. *et al.* (2007) *INCREASED SIZE EXCLUSION LIMIT2* encodes a putative DEVH box RNA helicase involved in plasmodesmata function during *Arabidopsis* embryogenesis. *Plant Cell* 19, 1885–1897
39 Pidkowich, M.S. *et al.* (2007) Modulating seed β-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4742–4747
40 Katoh, A. *et al.* (2006) Early steps in the biosynthesis of NAD in *Arabidopsis* start with aspartate and occur in the plastid. *Plant Physiol.* 141, 851–857
41 Saracco, S.A. *et al.* (2007) Genetic analysis of SUMOylation in *Arabidopsis*: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant Physiol.* 145, 119–134
42 Mizoi, J. *et al.* (2006) Defects in CTP:PHOSPHORYLETHANOLAMINE CYTIDYLYLTRANSFERASE affect embryonic and postembryonic development in *Arabidopsis*. *Plant Cell* 18, 3370–3385
43 Groß-Hardt, R. *et al.* (2007) *LACHESIS* restricts gametic cell fate in the female gametophyte of *Arabidopsis*. *PLoS Biol.* 5, e47
44 Raab, S. and Hoth, S. (2007) A mutation in the *AtPRP4* splicing factor gene suppresses seed development in *Arabidopsis*. *Plant Biol.* 9, 447–452
45 Welsch, R. *et al.* (2007) Transcription factor RAP2.2 and its interacting partner SINAT2: stable elements in the carotenogenesis of *Arabidopsis* leaves. *Plant Physiol.* 145, 1073–1085
46 Goh, T. *et al.* (2007) VPS9a, the common activator for two distinct types of Rab5 GTPases, is essential for the development of *Arabidopsis thaliana*. *Plant Cell* 19, 3504–3515
47 Székely, G. *et al.* (2008) Duplicated P5CS genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis. *Plant J.* 53, 11–28
48 Lermontova, I. *et al.* (2007) *Arabidopsis* CBF5 interacts with the H/ACA snoRNP assembly factor NAF1. *Plant Mol. Biol.* 65, 615–626
49 Borghi, L. *et al.* (2007) *Arabidopsis JAGGED LATERAL ORGANS* is expressed in boundaries and coordinates *KNOX* and *PIN* activity. *Plant Cell* 19, 1795–1808
50 Busi, M.V. *et al.* (2006) Deficiency of *Arabidopsis thaliana* frataxin alters activity of mitochondrial Fe-S proteins and induces oxidative stress. *Plant J.* 48, 873–882
51 Vazzola, V. *et al.* (2007) Knockout of frataxin gene causes embryo lethality in *Arabidopsis*. *FEBS Lett.* 581, 667–672
52 Lahmy, S. *et al.* (2007) QQT proteins colocalize with microtubules and are essential for early embryo development in *Arabidopsis*. *Plant J.* 50, 615–626
53 Liu, Z. and Makaroff, C.A. (2006) *Arabidopsis* separase AESP is essential for embryo development and the release of cohesion during meiosis. *Plant Cell* 18, 1213–1225
54 Xu, X.M. and Møller, S.G. (2006) AtSufE is an essential activator of plastidic and mitochondrial desulfurases in *Arabidopsis*. *EMBO J.* 25, 900–909
55 Kobayashi, K. *et al.* (2007) Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 17216–17221
56 Fleurdépine, S. *et al.* (2007) A *bona fide* La protein is required for embryogenesis in *Arabidopsis thaliana*. *Nucleic Acids Res.* 35, 3306–3321
57 Konishi, M. and Sugiyama, M. (2006) A novel plant-specific family gene, *ROOT PRIMORDIUM DEFECTIVE 1*, is required for the maintenance of active cell proliferation. *Plant Physiol.* 140, 591–602
58 Chen, M. *et al.* (2006) The essential nature of sphingolipids in plants as revealed by the functional identification and characterization of the *Arabidopsis* LCB1 subunit of serine palmitoyltransferase. *Plant Cell* 18, 3576–3593
59 Siddiqui, N.U. *et al.* (2006) Disruption of the *Arabidopsis* SMC4 gene, *AtCAP-C*, compromises gametogenesis and embryogenesis. *Planta* 223, 990–997
60 Jonczyk, R. *et al.* (2008) Pantothenate synthetase is essential but not limiting for pantothenate biosynthesis in *Arabidopsis*. *Plant Mol. Biol.* 66, 1–14
61 Xiao, W. *et al.* (2006) DNA methylation is critical for *Arabidopsis* embryogenesis and seed viability. *Plant Cell* 18, 805–814
62 Andre, C. *et al.* (2007) A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in *Arabidopsis*. *Plant Cell* 19, 2006–2022
63 Muralla, R. *et al.* (2008) A bifunctional locus (*BIO3–BIO1*) required for biotin biosynthesis in *Arabidopsis*. *Plant Physiol.* 146, 60–73
64 Ronceret, A. *et al.* (2008) The first zygotic division in *Arabidopsis* requires *de novo* transcription of thymidylate kinase. *Plant J.* 53, 776–789
65 Portereiko, M.F. *et al.* (2006) *NUCLEAR FUSION DEFECTIVE1* encodes the *Arabidopsis* RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. *Plant Physiol.* 141, 957–965
66 Portereiko, M.F. *et al.* (2006) *AGL80* is required for central cell and endosperm development in *Arabidopsis*. *Plant Cell* 18, 1862–1872
67 Barrero, J.M. *et al.* (2007) *INCURVATA2* encodes the catalytic subunit of DNA polymerase α and interacts with genes involved in chromatin-mediated cellular memory in *Arabidopsis thaliana*. *Plant Cell* 19, 2822–2838
68 Andersen, S.U. *et al.* (2007) The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc *in vitro* and *TSO1* is required for both male and female fertility in *Arabidopsis thaliana*. *J. Exp. Bot.* 58, 3657–3670
69 Fujiki, Y. *et al.* (2007) An *Arabidopsis* homolog of yeast *ATG6/VPS30* is essential for pollen germination. *Plant Physiol.* 143, 1132–1139
70 León, G. *et al.* (2007) Mitochondrial complex II is essential for gametophyte development in *Arabidopsis*. *Plant Physiol.* 143, 1534–1546
71 Babiychuk, E. *et al.* (2008) Allelic mutant series reveal distinct functions for *Arabidopsis* cycloartenol synthase 1 in cell viability and plastid biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3163–3168
72 Schnurr, J.A. *et al.* (2006) UDP-sugar pyrophosphorylase is essential for pollen development in *Arabidopsis*. *Planta* 224, 520–532
73 Jiang, L. *et al.* (2007) The *Arabidopsis* cohesin protein SYN3 localizes to the nucleolus and is essential for gametogenesis. *Plant J.* 50, 1020–1034
74 Johnston, A.J. *et al.* (2007) Genetic subtraction profiling identifies genes essential for *Arabidopsis* reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. *Genome Biol.* 8, R204
75 Mori, T. *et al.* (2006) *GENERATIVE CELL SPECIFIC 1* is essential for angiosperm fertilization. *Nat. Cell Biol.* 8, 64–71

76 von Besser, K. *et al.* (2006) *Arabidopsis HAP2* (*GCS1*) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development* 133, 4761–4769
77 Boisson-Dernier, A. *et al.* (2007) The peroxin loss-of-function mutation *abstinence by mutual consent* disrupts male–female gametophyte recognition. *Curr. Biol.* 18, 63–68
78 Chen, Y.H. *et al.* (2007) The central cell plays a critical role in pollen tube guidance in *Arabidopsis*. *Plant Cell* 19, 3563–3577
79 Escobar-Restrepo, J.M. *et al.* (2007) The FERONIA receptor-like kinase mediates male–female interactions during pollen tube reception. *Science* 317, 656–660
80 Hust, B. and Gutensohn, M. (2006) Deletion of core components of the plastid protein import machinery causes differential arrest of embryo development in *Arabidopsis thaliana*. *Plant Biol.* 8, 18–30
81 Chandler, J.W. *et al.* (2007) The AP2 transcription factors DORNROSCHEN and DORNROSCHEN-LIKE redundantly control *Arabidopsis* embryo patterning via interaction with PHAVOLUTA. *Development* 134, 1653–1662
82 Tilton, G.B. *et al.* (2006) Plant coenzyme A biosynthesis: characterization of two pantothenate kinases from *Arabidopsis*. *Plant Mol. Biol.* 61, 629–642
83 Nodine, M.D. and Tax, F.E. (2008) Two receptor-like kinases required together for the establishment of *Arabidopsis* cotyledon primordia. *Dev. Biol.* 314, 161–170
84 Titiz, O. *et al.* (2006) *PDX1* is essential for vitamin B6 biosynthesis, development and stress tolerance in *Arabidopsis*. *Plant J.* 48, 933–946
85 Rubio, S. *et al.* (2006) An *Arabidopsis* mutant impaired in coenzyme A biosynthesis is sugar dependent for seedling establishment. *Plant Physiol.* 140, 830–843
86 Kovacheva, S. *et al.* (2007) Further *in vivo* studies on the role of the molecular chaperone, Hsp93, in plastid protein import. *Plant J.* 50, 364–379
87 Azevedo, C. *et al.* (2006) Role of *SGT1* in resistance protein accumulation in plant immunity. *EMBO J.* 25, 2007–2016
88 Ge, C. *et al.* (2006) *BUD2*, encoding an *S*-adenosylmethionine decarboxylase, is required for *Arabidopsis* growth and development. *Cell Res.* 16, 446–456
89 Goritschnig, S. *et al.* (2007) The ubiquitin pathway is required for innate immunity in *Arabidopsis*. *Plant J.* 49, 540–551
90 Tsuwamoto, R. *et al.* (2008) *GASSHO1* and *GASSHO2* encoding a putative leucine-rich repeat transmembrane-type receptor kinase are essential for normal development of the epidermal surface in *Arabidopsis* embryos. *Plant J.* 54, 30–42
91 Drakakaki, G. *et al.* (2006) *Arabidopsis* reversibly glycosylated polypeptides 1 and 2 are essential for pollen development. *Plant Physiol.* 142, 1480–1492
92 Su, P.H. and Li, H.M. (2008) *Arabidopsis* stromal Hsp70s are essential for plant development and important for thermotolerance of germinating seeds. *Plant Physiol.* 146, 1231–1241
93 Pastuglia, M. *et al.* (2006) Gamma-tubulin is essential for microtubule organization and development in *Arabidopsis*. *Plant Cell* 18, 1412–1425
94 Pérez-Pérez, J.M. *et al.* (2008) Specialization of CDC27 function in the *Arabidopsis thaliana* anaphase-promoting complex (APC/C). *Plant J.* 53, 78–89
95 Chaiwongsar, S. *et al.* (2006) The protein kinase genes *MAP3Kε1* and *MAP3Kε2* are required for pollen viability in *Arabidopsis thaliana*. *Plant J.* 48, 193–205
96 Doelling, J.H. *et al.* (2007) The ubiquitin-specific protease subfamily UBP3/UBP4 is essential for pollen development and transmission in *Arabidopsis*. *Plant Physiol.* 145, 801–813
97 Lee, Y.R. *et al.* (2007) Two *Arabidopsis* phragmoplast-associated kinesins play a critical role in cytokinesis during male gametogenesis. *Plant Cell* 19, 2595–2605
98 Alonso, J.M. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657
99 Geisler-Lee, J. *et al.* (2007) A predicted interactome for *Arabidopsis*. *Plant Physiol.* 145, 317–329
100 Cui, J. *et al.* (2007) AtPID: *Arabidopsis thaliana* protein interactome database – an integrative platform for plant systems biology. *Nucleic Acids Res.* 36, D999–D1008
101 Steffen, J.G. *et al.* (2007) Identification of genes expressed in the *Arabidopsis* female gametophyte. *Plant J.* 51, 281–292
102 Spencer, M.W. *et al.* (2007) Transcriptional profiling of the *Arabidopsis* embryo. *Plant Physiol.* 143, 924–940
103 Li, Y. *et al.* (2006) Analysis of T-DNA insertion site distribution patterns in *Arabidopsis thaliana* reveals special features of genes without insertions. *Genomics* 87, 645–652
104 Noutoshi, Y. *et al.* (2005) *ALBINO AND PALE GREEN 10* encodes BBMII isomerase involved in histidine biosynthesis in *Arabidopsis thaliana*. *Plant Cell Physiol.* 46, 1165–1172