Abstract
The SeedGenes database (www.seedgenes.org) contains information on more than 400 genes required for embryo development in Arabidopsis. Many of these EMBRYO-DEFECTIVE (EMB) genes encode proteins with an essential function required throughout the life cycle. This raises a fundamental question. Why does elimination of an essential gene in Arabidopsis often result in embryo lethality rather than gametophyte lethality? In other words, how do mutant (emb) gametophytes survive and participate in fertilization when an essential cellular function is disrupted? Furthermore, why do some mutant embryos proceed further in development than others? To address these questions, we first established a curated dataset of genes required for gametophyte development in Arabidopsis based on information extracted from the literature. This provided a basis for comparison with EMB genes obtained from the SeedGenes dataset. We also identified genes that exhibited both embryo and gametophyte defects when disrupted by a loss-of-function mutation. We then evaluated the relationship between mutant phenotype, gene redundancy, mutant allele strength, gene expression pattern, protein function, and intracellular protein localization to determine what factors influence the phenotypes of lethal mutants in Arabidopsis. After removing cases where continued development potentially resulted from gene redundancy or residual function of a weak mutant allele, we identified numerous examples of viable mutant (emb) gametophytes that required further explanation. We propose that the presence of gene products derived from transcription in diploid (heterozygous) sporocytes often enables mutant gametophytes to survive the loss of an essential gene in Arabidopsis. Whether gene disruption results in embryo or gametophyte lethality therefore depends in part on the ability of residual, parental gene products to support gametophyte development. We also highlight here 70 preglobular embryo mutants with a zygotic pattern of inheritance, which provide valuable insights into the maternal-to-zygotic transition in Arabidopsis and the timing of paternal gene activation during embryo development.


Editor: Gordon Chua, University of Calgary, Canada

Received July 22, 2011; Accepted November 7, 2011; Published December 2, 2011

Copyright: © 2011 Muralla et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by National Science Foundation grant DBI-0618166 through the Arabidopsis 2010 Program, and by the Oklahoma Center for the Advancement of Science and Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: david.meinke@okstate.edu

Introduction
Essential genes have long played an important role in microbial and medical genetics. In recent years, several factors have contributed to the establishment of large datasets of essential genes in microorganisms: the failure of mutant cells to proliferate is readily detected in high-throughput gene disruption experiments; the products of essential genes define promising targets for novel, antimicrobial compounds relevant to human health; and lethal mutations in essential genes help to define the minimal gene set required to sustain a living cell [1,2]. Many laboratories have contributed to the large-scale identification of essential genes in bacteria [3–6] and yeast [7–10], and to the establishment of a database of essential genes for comparative studies [11]. Related efforts with Drosophila [12–14], C. elegans [15], mouse [16,17], and humans [18] have generated a wealth of information on the diversity of gene products required for viability in multicellular eukaryotes. Recent evolutionary studies have also explored the relationship between gene function, duplication, and essentiality in animal systems [19–23].

Our desire to establish a comparable dataset of essential genes in a model plant [24,25] arose from a longstanding interest in the isolation and characterization of embryo-defective (emb) mutants of Arabidopsis. [26–28]. Identifying the genes responsible for these mutant phenotypes was facilitated by advances in T-DNA insertional mutagenesis, which enabled large-scale, forward genetic screens for tagged, loss-of-function mutants defective in seed development [29,30]. Eventually, this led to the establishment of the SeedGenes database (www.seedgenes.org), which presents detailed information on genes required for embryo development in Arabidopsis [24]. The December 2010 database release includes 402 EMB genes with an essential cellular function required to produce a normal, mature embryo. Many EMB proteins are likely to be required throughout the life cycle. Mutants exhibit defects in embryogenesis because that is when the absence of a functional gene product first becomes critical.

Other research groups have pursued a complementary approach to the analysis of essential genes in Arabidopsis through the isolation and characterization of mutants defective in gametophyte development [31–33]. Identification of these mutants has been facilitated by screening for insertion lines that exhibit reduced transmission of an associated selectable marker. Although hundreds of mutants defective in gametophyte development have
been found over the years, the identities of genes responsible for these mutant phenotypes have often not been confirmed. Because large deletions and translocations are common in T-DNA and transposon insertion lines of Arabidopsis [36–39], and these chromosomal aberrations often result in lethality before fertilization [35], large datasets of gametophyte essentials that include candidate genes represented by a single mutant allele should be viewed with caution.

In order to characterize the cellular disruptions that lead to defects in embryo rather than gametophyte development, we first established a curated dataset of gametophyte essentials that could be compared with the embryo essentials found in the SeedGenes database. We then used this dataset in combination with an updated SeedGenes dataset to determine what factors most often distinguish mutants defective in embryo development from those defective in gametophyte development. Specifically, we wanted to understand how mutants defective in embryo development are able to survive gametophyte development when an essential function required throughout the life cycle is disrupted, and conversely, why some mutants exhibit gametophyte lethality instead. In addition, we sought to explain why some mutant embryos reach a later stage of development than others, and what general conclusions might be drawn about the relationship between gene function and mutant phenotype.

This analysis led us to conclude that pre-meiotic gene expression in diploid microsporocytes and megasporocytes is likely to be an important feature of reproductive development in Arabidopsis. Although past work has elegantly demonstrated the diversity of transcripts found in male and female gametophytes [40–44], the approaches pursued have not generally distinguished between transcripts produced before meiosis and those produced afterward. We propose that gene products derived from transcription of the wild-type allele in heterozygous sporocytes enable some mutant gametophytes to survive the loss of an essential gene, and that embryo lethality results when these residual products are exhausted. In Arabidopsis, pollen tubes travel short distances in order to reach the ovule, which increases the likelihood that male gametophytes lacking an essential gene product can survive and participate in fertilization. By contrast, equivalent gene disruptions in plants with long pollen tubes, most notably maize, are more likely to result in male gametophyte lethality.

This raises another important question: how can mutant embryos disrupted in an essential gene survive beyond the globular stage of development, given that levels of gene products stored in mutant gametophytes are insufficient to support unlimited growth? Several additional factors, including gene redundancy, residual gene function, delayed activation of specialized pathways, and diffusion of products from maternal tissues likely influence the stage of development reached by mutant embryos before seed desiccation. In addition, some gene products may perform a more limited role in an essential cellular process than others, and this can allow further development than might otherwise be expected. We conclude that the datasets of genes required for embryo and gametophyte development presented here provide a useful resource for comparative studies with other multicellular eukaryotes and valuable insights into the molecular basis of reproductive lethality in a model plant.

**Results and Discussion**

**The SeedGenes dataset of EMB genes required for embryo development**

Based on the abundance of embryo-defective mutants and the frequency of duplicate alleles in mutagenesis experiments, Arabidopsis appears to contain 750 to 1000 EMB genes located throughout the genome [20,45]. Table S1 presents an updated, comprehensive dataset of 396 EMB loci in Arabidopsis. Many of the EMB gene identities revealed through work in our laboratory have not been published before. The current dataset, which excludes six atypical or problematic SeedGenes loci, likely represents about 40% saturation. This is sufficient to begin addressing the diversity of protein functions required for cell viability and embryo development. Regarding the level of certainty that each gene identified is responsible for the mutant phenotype described, we recognize two distinct categories of EMB genes: those labeled as confirmed, either through molecular complementation or the analysis of additional mutant alleles disrupted in the same gene, and those labeled as not confirmed, where robust genetic data alone indicate close linkage between the gene and mutant phenotype (www.seedgenes.org). Approximately 100 of the EMB genes in Table S1 have identities that remain to be confirmed. Based on experience with other insertion lines, we estimate that >85% of these identities are correct. Genes of unknown function remain the most problematic because protein identity cannot be used to support the conclusion that an essential process has been disrupted.

To assess the terminal stage of development reached by mutant embryos prior to seed desiccation, we evaluated more than 700 mutants classified in SeedGenes as having an embryo-defective mutant phenotype. When more than one mutant allele was available, we chose what appeared to be the strongest allele based on phenotype and known location of the mutation. After removing 44 EMB loci associated with additional defects in gametophyte development, we divided the remaining 352 “true” EMB loci into terminal phenotype classes, ranging from preglobular to cotyledon. The results are shown in Figure 1A. Fifty-five percent of these loci are essential before the heart stage of development. Arrested embryos from these mutants are typically white, whereas many of those from mutants that reach the transition or cotyledon stages are pale green or green. More than 15% of “true” EMB genes are required at the earliest, preglobular stages of embryo development. Although such broad descriptions fail to capture important phenotype details, they demonstrate on a global scale that EMB genes differ in how far embryo development can proceed when their functions are disrupted. Furthermore, as discussed later, the existence of a substantial number of embryo-defective mutants with a Mendelian pattern of inheritance, characterized by 25% mutant seeds in siliques of selfed heterozygotes, along with a preglobular stage of embryo arrest, has implications for the timing and extent of the maternal-to-zygotic transition in gene expression during plant embryo development.

Historically, the distinction between embryo and gametophyte mutants in Arabidopsis has been vague and subjective. Some embryo mutants exhibit defects in gametophyte development, and some gametophyte mutants, when maintained as heterozygotes, produce a low frequency of viable mutant gametes. Embryo mutants with defects in male gametophyte development often exhibit a reduced percentage and nonrandom distribution of mutant seeds in siliques of selfed heterozygotes, along with a preglobular stage of embryo arrest, has implications for the timing and extent of the maternal-to-zygotic transition in gene expression during plant embryo development.
Because the SeedGenes database includes information, when available, on mutant seed ratios and distributions, we assembled a list of SeedGenes loci that appeared to have defects in both embryo and gametophyte development. These were assigned to an EMG (Embryo-Gametophyte) subclass of ES genes without a redundant paralog, and the remaining dataset of other (OT) “true” ES genes that may be redundant (BLASTP match e-30).

doi:10.1371/journal.pone.0028398.g001

**Figure 1. Terminal embryo phenotypes of the strongest known alleles of “true” ES genes without evidence of gametophyte defects.** (A) Entire dataset of 352 genes. (B) Comparison between unique “true” (UT) ES genes without a redundant paralog, and the remaining dataset of other (OT) “true” ES genes that may be redundant (BLASTP match e-30).

Protein functions of genes required for embryo and gametophyte development

Extensive overlap exists between the types of protein functions required for gametophyte development and those required for embryo development (Figure 2). Consequently, one cannot explain the difference between embryo and gametophyte mutants on the basis of protein function alone. Nevertheless, distinctive features of each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, whereas a complete loss of cytosolic translation in each dataset can be identified. For example, interfering with DNA replication and RNA modification typically disrupts embryo development, where
among mutants with severe defects in male gametophyte development. Most auxotrophic mutants exhibit embryo lethality rather than gametophyte lethality, which may be explained in part by the ability of maternal tissues to supply limited nutrients to mutant gametophytes. The disruption of male gametophyte development observed in several histidine auxotrophs placed in the EMG class of embryo lethals [50] may reflect a perturbation of the linked ATP recycling pathway rather than a loss of histidine. The abundance of EMB proteins with unknown functions included in the SeedGenes database, when compared with the gametophyte dataset, may reflect in part a reluctance of journals to include in part the SeedGenes database, when compared with the gametophyte dataset, may reflect in part a reluctance of journals to included in the SeedGenes database, when compared with the gametophyte dataset, may reflect in part a reluctance of journals to
generalize from those studies. We also distinguished insertions found in introns from those located in exons, although when a single flanking sequence was available and localized to an intron, we were uncertain whether an adjacent exon might also be disrupted. Overall, 89 of the 152 UT-EMB loci were associated with one or more mutant alleles that contained an exon insertion in the first or middle third of the gene. Another 23 had an intron insertion in the same (A/B) region, and 19 contained an insertion in the final third of the gene. When we compared the terminal phenotypes of 89 mutants carrying A/B exon disruptions with those of 42 mutants containing A/B intron or C region disruptions, we found no significant differences ($\chi^2 = 2.50; p = 0.65$) in the frequencies of early versus late embryo phenotypes (data not shown). We also found no significant differences ($\chi^2 = 7.69; p = 0.17$) in the phenotype distributions obtained when

Genetic redundancy and mutant allele strength

Two potential explanations for how mutants defective in embryo development might survive gametophyte development involve genetic redundancy and mutant allele strength. The first model proposes that duplicated genes with redundant functions compensate for the loss of essential EMB gene functions during gametophyte development. The second model suggests that most embryo-defective mutant alleles retain some function and predicts that true null alleles of EMB genes should be gametophyte lethals instead. Both models are insufficient to explain the results obtained. We evaluated the level of redundancy of 352 “true” EMB genes using BLASTP searches (www.ncbi.nlm.nih.gov) of the Arabidopsis proteome. EMB proteins lacking a match with at least an e-30 cutoff value to be relatively stringent in our selection of unique genes because we wanted to identify those EMB genes where survival of mutant gametophytes was least likely to result from genetic redundancy. The resulting dataset of 152 “unique true” (UT) EMB loci is presented in Table S5.

Distinguishing null alleles from those that retain a low level of protein function is particularly difficult with lethal mutants because homozygous mutant tissue cannot be readily analyzed. We therefore decided to record the insert locations for mutant alleles represented in the SeedGenes database, with the assumption that on average, disruptions in the beginning (A region) and middle (B region) of a gene may be more likely to eliminate protein function than those at the end of a gene (C region). We also distinguished insertions found in introns from those located in exons, although when a single flanking sequence was available and localized to an intron, we were uncertain whether an adjacent exon might also be disrupted. Overall, 89 of the 152 UT-EMB loci were associated with one or more mutant alleles that contained an exon insertion in the first or middle third of the gene. Another 23 had an intron insertion in the same (A/B) region, and 19 contained an insertion in the final third of the gene. When we compared the terminal phenotypes of 89 mutants carrying A/B exon disruptions with those of 42 mutants containing A/B intron or C region disruptions, we found no significant differences ($\chi^2 = 2.50; p = 0.65$) in the frequencies of early versus late embryo phenotypes (data not shown). We also found no significant differences ($\chi^2 = 7.69; p = 0.17$) in the phenotype distributions obtained when

### Table 1. Classification of genes required for embryo and gametophyte development.

<table>
<thead>
<tr>
<th>Class of essential genes</th>
<th>Subclass</th>
<th>Definition of class or subclass</th>
<th>Genes</th>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Embryo defective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True EMB</td>
<td></td>
<td>Defect in mutant embryo characterized</td>
<td>396</td>
<td>S1</td>
</tr>
<tr>
<td>UT-EMB</td>
<td></td>
<td>No documented gametophyte defect</td>
<td>352</td>
<td>S3</td>
</tr>
<tr>
<td><strong>Gametophyte defective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAM</td>
<td></td>
<td>Unique true EMB; No obvious paralog</td>
<td>152</td>
<td>S5</td>
</tr>
<tr>
<td>Viable</td>
<td></td>
<td>Defect in mutant gametophyte characterized</td>
<td>173</td>
<td>S2</td>
</tr>
<tr>
<td>Uncertain</td>
<td></td>
<td>Gametophyte defective; classification uncertain</td>
<td>13</td>
<td>S2</td>
</tr>
<tr>
<td><strong>Embryo / gametophyte defective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMG</td>
<td></td>
<td>&gt;10% mutant seeds predicted or observed</td>
<td>44</td>
<td>S2, S4</td>
</tr>
<tr>
<td>GEM</td>
<td></td>
<td>2–10% mutant seeds predicted or observed</td>
<td>25</td>
<td>S2, S4</td>
</tr>
<tr>
<td>GEM / GEM</td>
<td></td>
<td>Precise subclass uncertain or conflicted</td>
<td>4</td>
<td>S2</td>
</tr>
<tr>
<td>GEM / EMG</td>
<td></td>
<td>Precise subclass uncertain or conflicted</td>
<td>3</td>
<td>S2, S4</td>
</tr>
</tbody>
</table>

*Included in Table S2 but not discussed in the text.

doi:10.1371/journal.pone.0028398.t001

![Figure 2](image-url)
the unique and potentially redundant datasets of “true” EMB genes were compared (Figure 1B), although mutants with preglobular defects appeared to be somewhat more common among UT-EMB genes, consistent with the model that some mutants with late defects may be partially rescued by expression of a second, redundant gene.

Another point that argues against the idea that most embryo-defective mutants correspond to weak alleles of genes required for gametophyte development is that relatively few examples of gametophyte lethal alleles of known EMB loci have been documented in the literature. We have uncovered several cases, including EMB1989/NPB2, EMB2779/LCB1, RSW1/CESA1, EMB24/AaASP28, and EMB2776/PRP4/LIS [51–54], but the number is small when compared with the total number of essential genes identified to date. By contrast, more than 30 examples have been documented of weak mutant alleles of EMB genes that complete embryo development and exhibit other defects as seedlings or adult plants (Table S6). Based on these analyses, we conclude that in the vast majority of cases, the ability of emb mutant gametophytes to participate in fertilization is not likely to be explained by residual protein function provided by a weak mutant allele. Other factors besides genetic redundancy and weak mutant alleles must, therefore, be involved in supporting gametophyte development in heterozygous plants.

Significance of terminal embryo phenotypes

Disruptions of some UT-EMB genes permit growth and development to continue beyond the globular stage of embryo development even though the mutant alleles in question appear to be nulls. How can this apparent contradiction be explained? Because diploid sporocytes are small in comparison to multicellular embryos late in development, they are unlikely to provide gametophytes with sufficient quantities of stored gene products to explain the late embryo phenotypes found in these mutants. Other factors must therefore be involved. In order to evaluate such loci, we focused on 36 UT-EMB genes with A/B exon disruptions that exhibited a terminal embryo phenotype at the transition or cotyledon stage. One striking feature of the resulting dataset, shown in Table S7, is the diversity of possible causes for the late embryo phenotypes observed. These include a diverse array of functional proteins required for specialized pathways in metabolism (tag1, hyd1, emb1879), and peripheral roles of EMB proteins in essential cellular processes (emb10, emb2184, emb2754). Other notable features of this dataset include the abundance of genes encoding chloroplast-localized proteins (15 loci) and proteins with unknown cellular functions (12 loci). We conclude that in some cases, the late embryo phenotypes observed in these knockouts are consistent with known protein functions that become important late in embryo development. In other words, elimination of these specialized functions would not be expected to result in lethality immediately after fertilization. In other cases, the protein product likely performs a peripheral role in an essential process required throughout growth and development. The survival of mutant gametophytes in emb/EMB heterozygotes is best addressed, therefore, using EMB genes with knockout phenotypes at early stages of embryo development, because in those cases, survival of the mutant gametophyte contrasts with the documented, early lethality found in the mutant embryo.

We then focused on a dataset of 25 UT-EMB genes with a preglobular or zygotic terminal phenotype, which is more consistent with an essential gene product required throughout development (Table 2). Once again, a wide range of protein functions is represented, from purine, proline, and pantothenate biosynthesis (fac1, pts, emb2772) and RNA modification in the cytosol (emb2191, emb2820) to chromosome dynamics (emb2773, emb2782) and microtubule assembly (tnhs, fgh, par, emb2084, ggt2). Surprisingly, chloroplast-localized proteins are missing from this dataset and genes encoding proteins with unknown functions are reduced in occurrence. We propose that in some of these knockouts, residual gene products from the wild-type allele transcribed in diploid sporocytes enable mutant gametophytes to participate in fertilization. Development of homozygous mutant seeds arrests a short time later, when these reserves become depleted. Diffusion of nutrients from surrounding maternal tissues cannot be the primary source of metabolites for all stages of male gametophyte development because pollen-tube growth in Arabidopsis can proceed in vitro on a simple, defined medium [44,55].

Nature of stored gene products

In principle, the gene products contributed by diploid sporocytes to mutant gametophytes could be either RNA or protein, or a combination of both. Male gametophytes have long been known to be active in transcription. Post-meiotic gene expression was first demonstrated in maize by analyzing isozyme patterns for dimeric enzymes in pollen from heterozygous plants [56]. Female gametophytes have been more difficult to study, due in part to the presence of surrounding maternal tissues. Translation of cytosolic mRNAs is an essential feature of gametophyte development in Arabidopsis. Mutant alleles unable to support this essential process are likely not transmitted through male or female gametases [48]. Furthermore, pollen-tube growth in vitro is blocked in the presence of translation inhibitors [42]. This means that diploid sporocytes do not contribute to haploid gametophytes sufficient levels of functional proteins derived from all essential genes required to complete growth and development. Nevertheless, the possibility remains that stored products of some essential genes might be proteins instead of RNAs.

Knockout mutants disrupted in essential subunits of RNA polymerase are unable to transmit the mutant allele through female gametes and exhibit reduced transmission of the mutant allele through male gametes [51]. On the female side, this indicates that transcription within the gametophyte is required for normal development and that maternal sources of RNA polymerase are insufficient to meet transcriptional demands. On the male side, microsporocytes appear to contribute RNA polymerase activity to male gametophytes, either as stored transcript or functional protein, but this activity is not sufficient to meet the full transcriptional demand in mutant pollen, resulting in reduced rates of transmission. RNA polymerase knockouts, however, fail to address the broader question of which other products of essential genes are contributed from sporocytes to gametophytes, and whether some of those products are proteins rather than transcripts. A different pattern has been documented for the SHORT SUSPENSOR (SSP) locus, where a transcript produced in male gametophytes is translated in the zygote and modulates an important signaling pathway early in embryo development [57].

Transcriptional profiles of essential genes in male gametophytes

We searched for evidence of stored, pre-meiotic transcripts of EMB genes in haploid gametophytes by comparing the published transcript profiles [42–43] for several different classes of essential genes in Arabidopsis: (a) 59 GAM or GEM loci with severe defects in male transmission (Table S8); (b) 48 GEM or EMG loci with moderate defects in male transmission (Table S9); and (c) 75 UT-
defects in male transmission. This suggests that survival of absence. Unfortunately, the published proteome datasets for be present in mature pollen grains when their transcripts are pollen maturation. This model predicts that EMB proteins should pollen development and not on stored mRNAs translated after mutant pollen tubes often depends on proteins synthesized during development. We expected to find higher levels of transcript in young microspores than in pollen tubes or mature pollen grains when isolation. By contrast, we anticipated that transcript levels would increase or pre-meiotic transcripts supported male gametophyte development. Microspores have a documented preglobular embryo phenotype. EMB loci with early defects in seed development Arabidopsis [58–62] are too limited in scope to support or refute this prediction. Only a handful of proteins derived from UT-EMB genes with early stages of embryo arrest are included in these datasets, and even the GAM, GEM, and EMG classes are infrequently represented (Table S11).

Transcripts from UT-EMB genes with early embryo defects are frequently detected in young microspores, and their levels do not increase substantially during pollen development. Both the transcript profiles and inheritance patterns of these EMB genes are therefore consistent with the utilization of stored transcript during pollen development and inconsistent with gametophyte expression. By contrast, genes required for male gametophyte development are often up-regulated during pollen development and their transcripts are frequently detected both at pollen maturity and in pollen tubes. As expected, EMG and GEM loci with moderate defects in male gametophyte development were expected to have intermediate or variable profiles.

Our analysis of male transcriptome datasets is summarized in Table 3. Overall, the results from different laboratories are remarkably consistent and highlight informative trends. Transcripts from UT-EMB loci with early defects in seed development are more likely to be absent from mature pollen ($\chi^2 = 17.1; p<0.001$) than transcripts from gametophyte loci with severe defects in male transmission. This suggests that survival of emb mutant pollen tubes often depends on proteins synthesized during pollen development and not on stored mRNAs translated after pollen maturation. This model predicts that EMB proteins should be present in mature pollen grains when their transcripts are absent. Unfortunately, the published proteome datasets for EMB loci with functional male gametophytes and a preglobular (25 loci; Table 2) or globular (50 loci; Table S10) stage of embryo arrest. We expected to find higher levels of transcript in young microspores than in pollen tubes or mature pollen grains when pre-meiotic transcripts supported male gametophyte development. By contrast, we anticipated that transcript levels would increase or remain steady throughout gametophyte development for GAM loci with severe defects in male transmission, consistent with post-meiotic gene expression. EMG and GEM loci with a combination of embryo defects and moderate defects in male gametophyte development were expected to have intermediate or variable profiles.

Table 2. Unique “true” EMB genes with a documented preglobular embryo phenotype.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene symbol</th>
<th>Identity status</th>
<th>Function class</th>
<th>Predicted protein function</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g03360</td>
<td>RRP4</td>
<td>C</td>
<td>2.3</td>
<td>tRNA Processing; Exonuclease</td>
</tr>
<tr>
<td>At1g09770</td>
<td>AICDC5</td>
<td>NC</td>
<td>7</td>
<td>MYB Domain Transcription Factor; Cell Cycle Control</td>
</tr>
<tr>
<td>At1g48175</td>
<td>EMB2191</td>
<td>C</td>
<td>2.3</td>
<td>tRNA Adenosine Deaminase</td>
</tr>
<tr>
<td>At1g5350</td>
<td>EMB1275</td>
<td>C</td>
<td>12</td>
<td>Calpain-Type Cysteine Protease</td>
</tr>
<tr>
<td>At1g55900</td>
<td>EMB1860</td>
<td>C</td>
<td>4.2</td>
<td>Inner Mitochondrial Membrane Import Protein</td>
</tr>
<tr>
<td>At1g71440</td>
<td>PFI</td>
<td>C</td>
<td>11.1</td>
<td>Tubulin Folding Cofactor E; Microtubule Dynamics</td>
</tr>
<tr>
<td>At2g30920</td>
<td>EMB3002</td>
<td>NC</td>
<td>9</td>
<td>Methyltransferase; Ubiquinone Biosynthesis</td>
</tr>
<tr>
<td>At2g32590</td>
<td>EMB2795</td>
<td>C</td>
<td>6.2</td>
<td>Condensin Complex Subunit; Sister Chromatid Segregation</td>
</tr>
<tr>
<td>At2g34780</td>
<td>EMB1611</td>
<td>C</td>
<td>13</td>
<td>Uncertain</td>
</tr>
<tr>
<td>At2g38280</td>
<td>FAC1</td>
<td>C</td>
<td>10.1</td>
<td>AMP Deaminase; Purine Biosynthesis</td>
</tr>
<tr>
<td>At2g38670</td>
<td>PECT1</td>
<td>C</td>
<td>10.5</td>
<td>Phosphatidylethanolamine Biosynthesis</td>
</tr>
<tr>
<td>At3g10220</td>
<td>EMB2804</td>
<td>C</td>
<td>11.1</td>
<td>Tubulin Folding Cofactor B; Microtubule Dynamics</td>
</tr>
<tr>
<td>At3g20070</td>
<td>TTN9</td>
<td>C</td>
<td>13</td>
<td>Unknown</td>
</tr>
<tr>
<td>At3g60740</td>
<td>TTN1</td>
<td>C</td>
<td>11.1</td>
<td>Tubulin Folding Cofactor D; Microtubule Dynamics</td>
</tr>
<tr>
<td>At4g03240</td>
<td>AIFH</td>
<td>C</td>
<td>12</td>
<td>Frataxin; Mitochondrial Fe-S Protein Biogenesis</td>
</tr>
<tr>
<td>At4g21800</td>
<td>QQ72</td>
<td>C</td>
<td>11.1</td>
<td>ATP/GTP Binding Protein; Microtubule Localization</td>
</tr>
<tr>
<td>At4g39920</td>
<td>POR</td>
<td>C</td>
<td>11.1</td>
<td>Tubulin Folding Cofactor C; Microtubule Dynamics</td>
</tr>
<tr>
<td>At5g13480</td>
<td>FY</td>
<td>C</td>
<td>2.3</td>
<td>Nuclear RNA Binding / Polyadenylation Protein</td>
</tr>
<tr>
<td>At5g13690</td>
<td>CYL1</td>
<td>C</td>
<td>10.4</td>
<td>Alpha-N-Acetyl-Glucosaminidase</td>
</tr>
<tr>
<td>At5g14800</td>
<td>EMB2772</td>
<td>C</td>
<td>10.1</td>
<td>Proline Biosynthesis</td>
</tr>
<tr>
<td>At5g15540</td>
<td>EMB2773</td>
<td>C</td>
<td>6.2</td>
<td>Adherin; Sister Chromatid Cohesion</td>
</tr>
<tr>
<td>At5g15920</td>
<td>EMB2782</td>
<td>C</td>
<td>6.2</td>
<td>SMC Family Protein; Chromosome Dynamics</td>
</tr>
<tr>
<td>At5g24670</td>
<td>EMB2820</td>
<td>C</td>
<td>2.3</td>
<td>tRNA Adenosine Deaminase</td>
</tr>
<tr>
<td>At5g48840</td>
<td>PTS</td>
<td>C</td>
<td>10.1</td>
<td>Pantothenate Synthetase</td>
</tr>
<tr>
<td>At5g59440</td>
<td>ZEU51</td>
<td>C</td>
<td>10.1</td>
<td>Thymidylate Kinase; Nucleotide Biosynthesis</td>
</tr>
</tbody>
</table>

aGene responsible for mutant phenotype confirmed (C) or not confirmed (NC) through allelism tests or molecular complementation.
bProtein function classes are defined in Table S1, last spreadsheet tab.
doi:10.1371/journal.pone.0028398.t002
are more abundant in pollen and are retained in sufficient levels after fertilization to support further embryo development remains an open question.

Transcriptional profiles of essential genes in microsporocytes

To test our prediction that transcripts of EMB genes are often contributed from diploid microsporocytes to haploid gametophytes, we evaluated three recent reports of transcript profiles for isolated male meiocytes. Two of these studies utilized direct RNA sequencing [63,64]; the other relied on microarray analysis [65]. One report [64] characterized the full transcriptional landscape of isolated microsporocytes; the other two [63,65] emphasized the identification of meiosis-specific gene products. One report [64] characterized the full transcriptional landscape of isolated microsporocytes; the other two [63,65] emphasized the identification of meiosis-specific gene products. Of the 183 GAM, GEM, and EMB genes listed in Table 2 and Tables S8, S9, S10 combined, just one is included among the 844 meiosis-specific loci described by Yang et al. [64]. This is not surprising because most of the genes in our dataset encode proteins thought to be required at multiple stages of development. Similarly, only 5% of the 183 essential genes in our dataset are included among 752 transcripts that Libeau et al. [65] conclude are enriched in microsporocytes (χ² = 15.0 RPKM) is not significantly higher for UT-EMB genes with early embryo defects than for male gametophyte essentials (χ² = 0.61; p = 0.44), or for all genes with transcripts detected in microsporocytes (χ² = 0.22; p = 0.64). We therefore propose that in most cases where transcription of an EMB gene in microsporocytes supports gametophyte development, survival of mutant (mb) gametophytes results from relatively modest levels of residual transcript contributed to young microsperocytes and subsequently translated during pollen development. By contrast, male gametophyte defects in gametophyte lethals arise when post-meiotic expression of an essential gene is required to supplement whatever functional transcript remains from pre-meiotic expression in microsporocytes. Confirmation of this model could be achieved in the future by determining whether down-regulation of an EMB gene in microsporocytes changes an embryo lethal into a gametophyte lethal.

Transcriptional profiles of essential genes in megasporecytes

We focused our analysis on genes required for male gametophyte development because data for female meiocytes are more difficult to obtain and evaluate. One could argue that because the number of confirmed, female gametophyte mutants identified to date is fewer than the number of male gametophyte mutants, megaspores may be more effective than microsporocytes at

| Table 3. Summary of pollen development transcriptome data for selected essential genes. |
|-----------------------------------------------|---------------------------------|-------------------|-------------------|-------------------|-------------------|
| Transcript accumulation pattern found in male gametophytes | Unique true EMB genes | Male gametophyte defectives |
|---------------------------------------------------------------|----------------|-------------------|-------------------|-------------------|-------------------|
| Transcript detected early in pollen development but not in mature pollen | Preglobular | Globo | Moderate | Severe | Reference |
| Transcript detected throughout pollen development | 70% | 64% | 55% | 27% | [42] |
| Other patterns of transcript accumulation | 17% | 29% | 43% | 61% | [42] |
| Transcript detected throughout pollen development; ≥2× higher at late stages | 13% | 7% | 2% | 12% | [42] |
| Transcript detected throughout pollen development; ≥2× higher at early stages | 4% | 10% | 12% | 13% | [42] |
| Comparison of results obtained from different laboratories | Transcript detected in mature pollen | Substantial transcript detected in mature pollen | Transcript detected in mature pollen | Transcript detected in mature pollen | Transcript detected in pollen tube |
| Transcript detected in mature pollen | 26% | 36% | 43% | 69% | [42] |
| Substantial transcript detected in mature pollen | 9% | 12% | 21% | 47% | [42] |
| Transcript detected in mature pollen | 13% | 14% | 29% | 42% | [43] |
| Transcript detected in mature pollen | 13% | 19% | 21% | 51% | [44] |
| Transcript detected in pollen tube | 13% | 14% | 21% | 45% | [44] |

*Based on published, large-scale microarray datasets for wild-type plants [42–44].
*Non-redundant genes with embryo arrest stage shown; no gametophyte defects observed.
Moderate (EMG and GEM) and severe (GAM) classes are described in the text.
*Percentages of essential genes with transcript detected are noted. Total number of genes analyzed: preglobular (23/25); globular (42/50); moderate (42/48); severe (51/59). Excluded genes were not part of the microarray dataset or had no transcript detected during pollen development.
*Level detected was among top 50% of all transcript levels recorded.
*Level detected was among top 25% of all transcript levels recorded. This cutoff value for detection [42] gave results more consistent with those obtained elsewhere [43,44].
*Most of the genes identified with a positive signal using data from one laboratory were the same as those identified using data from the other laboratories.

doi:10.1371/journal.pone.0028398.t003
contributing functional gene products to progeny haploid cells. This cannot, however, be true for all essential genes, because RNA polymerase knockouts are not transmitted through the female side, which indicates that some essential genes must be transcribed in female gametophytes. While it seems logical that knockouts of genes required for rapid tip growth should exhibit male rather than female gametophyte lethality, the mechanism responsible for severe female defects in knockouts with minimal disruption of male gametophytes remains obscure. In some cases, gene duplication may be involved, with a redundant gene supplying the missing function on the male side when the paralog required for female development is disrupted.

The recent publication of transcriptional profiles for Arabidopsis megasporocytes [66] and female gametophytes [67] makes it possible to analyze maternal contributions to haploid gametophytes and the role of pre-meiotic transcription in supporting embryo development. Transcripts of many essential genes are detected in megasporocytes. Of the 75 UT-EMB genes with early embryo defects, 62% produced transcripts detected in at least three of four megasporocyte samples analyzed [66]. This is consistent with the model that some transcripts produced in megasporocytes are used to support female gametophyte development. We then asked whether genes associated with severe defects in female transmission often fail to produce transcripts detected in megasporocytes. Unexpectedly, when 19 such genes (Table S2) were examined, the likelihood of transcripts being present in megasporocytes was not significantly different \( \chi^2 = 0.24; p = 0.62 \) from that observed for the 75 UT-EMB genes with early phenotypes, whose knockouts survive gametophyte development. This result mirrors the trend noted above for microsporocyte transcripts involving male gametophyte essential genes. Furthermore, we found no significant difference \( \chi^2 = 1.81; p = 0.40 \) in the likelihood of transcripts being present in megasporocytes, and available in the egg cell for contribution to the zygote, when we compared our datasets of UT-EMB genes with preglobular, globular, and cotyledon mutant phenotypes. We have therefore been unable to establish a role for pre-meiotic transcripts in supporting the continued growth of mutant embryos to a globular stage and beyond. By contrast, observed differences in the severity of phenotypes between male and female gametophyte mutants disrupted in essential genes, and between embryo and gametophyte lethals overall, appear in some cases to reflect differences in the amount of residual, essential gene products donated from sporocytes to developing gametophytes. The practical applications of these different patterns of reproductive lethality in Arabidopsis remain to be explored.

The transition from maternal to zygotic gene expression

In addition to evaluating sporocyte transcription and gametophyte viability in embryo-defective mutants of Arabidopsis, we utilized the dataset of EMB genes presented here to address another central issue in plant development related to stored, essential gene products. Two fundamental questions have figured prominently in recent discussions of gene expression during plant embryo development. One involves the extent of paternal allele contributions to the developing embryo and endosperm tissue. The other concerns when during development the transition from maternal to zygotic gene expression takes place. Both of these questions have been widely studied in maize [68–70] and Arabidopsis [71–77]. However, significant conflicts remain over the interpretation of results obtained, reliability of methods employed, and validity of conclusions drawn.

Our interest in this field began almost 20 years ago, when we isolated and characterized a mutant (emb175) with an unusual pattern of inheritance: heterozygous plants produced 50% mutant seeds regardless of pollen genotype [36]. Additional alleles of this gene, later named MEA, were isolated and characterized elsewhere, the gene was shown to encode a polycomb group protein that regulates cell proliferation, and the paternal allele was found to be imprinted by DNA methylation, particularly in the endosperm [78–80]. Several other genes with a similar pattern of inheritance [www.seedgenes.org] but differing seed phenotypes and mechanisms of maternal influence have also been described [81,82]. Evaluation of these genes and mutant phenotypes has frequently centered on maternal genome contributions and paternal genomic imprinting in relation to parental conflict over resource allocation from mother to offspring. Despite considerable interest in the origin and significance of these loci, their numbers are quite limited when compared to EMB genes with a normal, zygotic pattern of inheritance.

We were therefore surprised by a report 12 years ago that claimed widespread, delayed activation of the paternal genome during seed development in Arabidopsis and concluded that early embryo and endosperm development are primarily, if not exclusively, under maternal control [71]. This conclusion, which appeared inconsistent with the inheritance pattern of early embryo-lethal mutants, was based on the analysis of β-glucuronidase (GUS) enhancer trap lines that marked the expression of 19 different genes known to be active during ovule and early seed development. By comparing the presence or absence of GUS staining in embryos produced through reciprocal crosses to wild-type plants, the authors concluded that the paternal gene in each case was not expressed until 3 to 4 days after pollination, when the

---

**Table 4.** Summary of microsporocyte transcriptome data for selected essential genes.

<table>
<thead>
<tr>
<th>Transcript detected in microsporocytes</th>
<th>Unique true <em>EMB</em> genes</th>
<th>Male gametophyte defectives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preglobular</td>
<td>Globular</td>
</tr>
<tr>
<td>&gt;=1.0 RPKM</td>
<td>88%</td>
<td>82%</td>
</tr>
<tr>
<td>&gt;5.0 RPKM*</td>
<td>41%</td>
<td>41%</td>
</tr>
<tr>
<td>&gt;15.0 RPKM*</td>
<td>14%</td>
<td>17%</td>
</tr>
</tbody>
</table>

*Based on published transcriptome dataset of Yang et al. [64]. Percentages of essential genes with transcripts detected at different abundance levels are noted.

†Non-redundant genes with early embryo arrest and no gametophyte defects. Genes analyzed: preglobular [25]; globular [50].

‡Moderate (EMG and GEM) and severe (GAM) classes are described in the text. Genes analyzed: moderate [48]; severe [59].

§Reads per kilobase per million reads (RPKM); high numbers indicate abundant transcripts.

‖Transcripts with RPKM<1.0 considered present.

Genes with no transcript detected and those with RPKM<1.0 are excluded from the percentages shown.

doi:10.1371/journal.pone.0028398.t004
embryo had reached the early globular stage. Additional studies with EMB30/GNOM and tests for allele-specific single nucleotide polymorphisms (SNPs) in transcripts produced following reciprocal crosses between accessions seemed to support these observations. This model was soon challenged through further work on Arabidopsis [72]. With maize, published reports were both consistent [68] and inconsistent [69] with the model. Down-regulating RNA polymerase II activity after fertilization was later found to interfere with endosperm but not early embryo development in Arabidopsis [73]. Once again, this suggested that the embryo is transcriptionally quiescent before the globular stage and is dependent instead on stored maternal transcripts.

Recently, a large-scale RNA sequencing effort [77] designed to detect accession-specific SNPs from Arabidopsis embryos at the 2–4 cell stage indicated that genome-wide, transcripts from the maternal allele predominate over those from the paternal allele through at least the globular stage of development. Following a cross between Columbia and Landsberg erecta accessions, almost 90% of the 135,000 informative sequence reads derived from 4,000 loci genome-wide appeared to be of maternal origin at the 2–4 cell embryo stage. Overall, the paternal contribution increased somewhat by the globular stage, and some loci produced equal amounts of transcripts from both maternal and paternal alleles, but early in development, the vast majority of loci seemed to be under maternal control. The authors stated that their work reconciled observations made by different laboratories of both maternal and zygotic effects during embryogenesis in Arabidopsis. But their data still conflict with the existence of large numbers of embryo-defective mutants with a Mendelian pattern of inheritance and cellular defects observed well before the globular stage.

Because all of the publications on the maternal-to-zygotic transition in Arabidopsis have cited at most a handful of EMB genes with known defects early in development, we decided to establish a more robust dataset of known EMB genes with arrested embryos that fail to reach the globular stage. The resulting collection of 70 genes, corresponding to about 18% of all 396 EMB genes identified to date, is presented in Table S12. This dataset does not include genes such as EMB30/GNOM, where the initial defects are observed long before embryo development becomes arrested, and other mutants with late terminal phenotypes where early defects remain to be uncovered. Nevertheless, with this dataset in hand, we were intrigued to learn of another study on the maternal-to-zygotic transition in Arabidopsis that followed an approach similar to Autran et al. [77] but expanded the number of genes and embryo stages examined, included the analysis of reciprocal crosses, and increased the washing of isolated embryos to remove contaminating RNA derived from the maternal seed coat (Michael Nodine and David Bartel, personal communication). Their results for more than 7,000 loci present a different picture, with maternal and paternal transcripts for most genes found in equal amounts, even at the 1–2 cell stage. Transcripts from the 70 preglobular EMB genes listed in Table S12 exhibit a similar, bi-parental pattern of accumulation, consistent with their known mode of inheritance.

When these results are evaluated in the context of several decades of research on embryo-defective mutants, we believe there is compelling evidence for early activation of the zygotic genome in Arabidopsis, and that unlike the situation observed in most animal systems, early expression of essential genes is required to support a wide range of cellular functions beginning shortly after fertilization. From a historical perspective, elucidating the role of paternal gene expression in plant embryo development provides a compelling illustration of the value of large datasets of genes with essential functions during growth and development. A more detailed picture of the molecular basis of reproductive lethality in Arabidopsis should emerge as additional genes with essential functions are identified in the future.

Materials and Methods
Establishing a dataset of genes required for gametophyte development
Several approaches were taken to identify and evaluate genes reported to be required for gametophyte development in Arabidopsis. We began with a list of all known genes with a mutant phenotype published 8 years ago [84], examined candidate genes from The Arabidopsis Information Resource (www.arabidopsis.org) that were thought to be associated with phenotype information, and performed a variety of PubMed searches (www.ncbi.nlm.nih.gov) with relevant keywords (gamete, gametophyte, mutant, mutation, knockout, lethal, essential, male, female, and pollen). Publications describing the large-scale identification of genes required for male and female gametophyte development [31,35] were evaluated to identify genes most likely to be responsible for the gametophyte defects observed. Emphasis was placed on genes with multiple mutant alleles and flanking sequence derived from both sides of the insert. We also utilized information presented at SeedGenes and in publications on embryo-defective mutants to determine which mutants exhibited additional defects in gametophyte development. Transmission efficiency (TE) of the mutant allele, determined in crosses between heterozygous and wild-type plants, was defined as: number of
heterozygous F1 plants) / (number of wild-type F1 plants). Subclasses of genes with different mixtures of embryo and gametophyte defects (GAM, GEM, EMG) and levels of transmission reduction (e.g. 0%, MMM, MM, M, +) are defined in the second and third tabbed spreadsheets in Table S2. The percentage of mutant seeds expected in sired silicles of heterozygotes with reduced transmission of the mutant allele through one or both gametes was calculated as: [(q TE)/(1+q TE)] ×[(q TE)/(1+q TE)]×100, where “1” represents normal transmission efficiency of the wild-type allele. This reduces to [1/(1+1)] ×[1/(1+1)]×100, or 25% mutant seeds in the absence of defects in gamete transmission.

Supporting Information

Table S1  Updated SeedGenes dataset of 396 EMB genes of Arabidopsis.

Table S2 173 Arabidopsis genes with a gametophyte-defective phenotype.

Table S3 352 true EMB genes without evidence of gametophyte defects.

Table S4 72 GAM and EMG genes with both embryo and gametophyte defects.

Table S5 152 unique true EMB genes of Arabidopsis.

Table S6 Examples of EMB genes with other phenotypes in weak alleles or RNAI lines.

References


Table S7 36 unique true EMB genes with late phenotypes in putative knockout alleles.

Table S8 59 GAM or GEM genes with severe defects in male gamete transmission.

Table S9 48 GEM or EMG genes with moderate defects in male gamete transmission.

Table S10 50 unique true EMB genes with early globular to globular phenotypes.

Table S11 Detection of proteins required for male gametophyte and embryo development in pollen.

Table S12 70 EMB genes where mutant embryos arrest at the preglobular stage of development.

Acknowledgments

We thank former members of the Meinke laboratory at Oklahoma State University, Allan Dickerman at the Virginia Bioinformatics Institute, and David Patton at Syngenta for contributing to the establishment of the SeedGenes database. Michael Nodine in the laboratory of David Bartel at the Massachusetts Institute of Technology kindly shared unpublished data on transcript patterns for preglobular EMB loci.

Author Contributions

Conceived and designed the experiments: DM RM. Wrote the paper: DM. Established, analyzed, and curated the embryo datasets: DM RM JL. Established, analyzed, and curated the gametophyte datasets: JL DM. Analyzed the transcriptome and proteome datasets: RM DM.


