Insertional Mutagenesis of Genes Required for Seed Development in *Arabidopsis thaliana*

John McElver,*,[1] Iris Tzafrir,†,[1] George Aux,* Rebecca Rogers,† Carl Ashby,*,[2] Kelsey Smith,†,[3] Carla Thomas,* Amy Schetter,† Qing Zhou,* Mary Ann Cushman,†,[4] John Tossberg,* Todd Nickle,†,[5] Joshua Z. Levin,* Marcus Law,* David Meinke† and David Patton*

*Syngenta, Research Triangle Park, North Carolina 27709 and †Department of Botany,
Oklahoma State University, Stillwater, Oklahoma 74078

ABSTRACT

The purpose of this project was to identify large numbers of Arabidopsis genes with essential functions during seed development. More than 120,000 T-DNA insertion lines were generated following Agrobacterium-mediated transformation. Transgenic plants were screened for defective seeds and putative mutants were subjected to detailed analysis in subsequent generations. Plasmid rescue and TAIL-PCR were used to recover plant sequences flanking insertion sites in tagged mutants. More than 4200 mutants with a wide range of seed phenotypes were identified. Over 1700 of these mutants were analyzed in detail. The 350 tagged embryo-defective (*emb*) mutants identified to date represent a significant advance toward saturation mutagenesis of *EMB* genes in Arabidopsis. Plant sequences adjacent to T-DNA borders in mutants with confirmed insertion sites were used to map genome locations and establish tentative identities for 167 *EMB* genes with diverse biological functions. The frequency of duplicate mutant alleles recovered is consistent with a relatively small number of essential (*EMB*) genes with nonredundant functions during seed development. Other functions critical to seed development in Arabidopsis may be protected from deleterious mutations by extensive genome duplications.

SEED development has become a popular subject for genetic analysis because it plays a critical role in the life cycle of flowering plants (MEINKE 1995; LAUX and JURGENS 1997; MORDHORST *et al.* 1997). Many genes must be expressed as the zygote divides in a regulated manner, completes morphogenesis, and differentiates into a mature embryo capable of surviving desiccation and producing a viable plant. *Arabidopsis thaliana* has long been used as a model system for mutant analysis of seed development (MEINKE and SUSSEX 1979). Hundreds of embryo-defective (*emb*) mutants have been identified and used to address fundamental questions in plant biology. These recessive mutants have been recovered either by screening immature siliques for abnormal seeds (MEINKE 1994) or by screening at the seedling stage for defects indicative of a disruption of normal embryogenesis (JURGENS *et al.* 1994). Included among the many published mutants are examples of defects in growth, morphogenesis, cell division, cell differentiation, and pattern formation (MEINKE 1994; YADEGARI *et al.* 1994). Some of these embryo mutants are allelic to mutants known in other labs for their aberrations in vegetative development (CONKLIN *et al.* 1999; LUKOWITZ *et al.* 2001). Others are identical to pattern mutants isolated elsewhere at the seedling stage (JURGENS *et al.* 1994). A few mutants appear to be weak alleles of female gametophytic factors (SPRINGER *et al.* 1995). The common feature is that embryo development becomes abnormal after fertilization and before germination.

Genetic analysis of seed development has progressed in recent years from descriptions of mutant phenotypes to identification of mutant genes. Emphasis for gene isolation has been placed on collections of insertion mutants (FELDMANN 1991; BOUCHEZ *et al.* 1993). Several labs have participated in screens for T-DNA mutants defective in embryogenesis (CASTLE *et al.* 1993; YADEGARI *et al.* 1994; DEVIC *et al.* 1996; LINDSEY *et al.* 1996) while others have established systems to identify genes through transposon tagging (ALTMANN *et al.* 1995; OSBORNE *et al.* 1995; WISMAN *et al.* 1998; PARINOV *et al.* 1999; SPEULMAN *et al.* 1999; TISSIER *et al.* 1999) and gene trapping (TOPPING *et al.* 1994; SUNDARESAN *et al.* 1995; MARTIENSSEN 1998). Advances have also been

Corresponding author: David Meinke, Department of Botany, 104 Life Sciences East, Oklahoma State University, Stillwater, OK 74078. E-mail: meinke@okstate.edu

[1] These authors contributed equally to this work.

[2] *Present address:* Department of Biology, University of Virginia, Charlottesville, VA 22903.

[3] *Present address:* University of Oklahoma College of Medicine, Oklahoma City, OK 73190.

[4] *Present address:* Department of Biochemistry, University of Nevada, Reno, NV 89557.

[5] *Present address:* Mount Royal College, Calgary, Alberta T3E 6K6, Canada.

made in methods for map-based cloning (Lukowitz *et al.* 2000). The result has been significant advances in the number of *EMB* genes identified. Examples include mutants defective in transcription factors and associated proteins (Aida *et al.* 1997; Grossniklaus *et al.* 1998; Hardtke and Berleth 1998; Li and Thomas 1998; Lotan *et al.* 1998; Luerssen *et al.* 1998), metabolic enzymes (Patton *et al.* 1998; Lukowitz *et al.* 2001), transport and trafficking proteins (Lukowitz *et al.* 1996; Assaad *et al.* 2001), factors required for replication and translation (Springer *et al.* 1995; Tsugeki *et al.* 1996; Zhang and Somerville 1997; Uwer *et al.* 1998), and proteins with other essential functions (Albert *et al.* 1999; Jang *et al.* 2000; Schrick *et al.* 2000; Boisson *et al.* 2001).

We describe here the results of a large-scale project designed to saturate for T-DNA tagged mutants defective in seed development. A companion article (Budziszewski *et al.* 2001) presents a similar analysis of genes required for seedling development. Together these articles represent a coordinated effort to establish a large collection of insertion mutants defective in genes with essential functions during plant growth and development. Such knockouts of essential genes will be an important component of international efforts to build upon the Arabidopsis Genome Initiative (2000) and determine the biological function of every gene in Arabidopsis before the end of the decade (Somerville and Dangl 2000).

MATERIALS AND METHODS

Plant maintenance: T-DNA insertion lines were generated in the Columbia (COL) and Wassilewskija (WS) ecotypes of *A. thaliana*. Plants used for transformation experiments were grown in soil (Super Fine Germinating Mix; Conrad Fafard, Agawam, MA), watered from below once a week and maintained under fluorescent lights at $22° \pm 4°$ with 50% relative humidity and 16-hr light/8-hr dark cycles. Transgenic (T_1) plants screened for mutant phenotypes at Syngenta were grown in Aratrays (Lehle Seeds, Round Rock, TX) and covered with Aracons (Lehle) at bolting to minimize cross-pollination. Lines chosen for genetic analysis at Oklahoma State University were grown in pots containing a mixture of vermiculite, soil, and sand (Heath *et al.* 1986) and placed in a growth room at $24° \pm 2°$ under fluorescent lights on 16 hr light/8 hr dark cycles. Plants were watered daily from below with a fertilizer solution (Peters 15-16-17; Scotts, Chicago, IL). Heterozygous plants were identified by screening immature siliques for the presence of 25% defective seeds following self-pollination (Meinke 1994).

Plant transformation: A number of binary transformation vectors were used in this study. Initial populations were generated with activation tagging vectors pPCVICEn4HPT (Hayashi *et al.* 1992) and pSKI015 (Weigel *et al.* 2000) containing multiple enhancers in the right border and plasmid rescue elements for subsequent recovery of flanking sequences. Because this project did not emphasize the creation of gain-of-function mutations, most insertion lines were made with simple T-DNA vectors (pCSA104, pDAP101, pCSA110) with a BASTA-resistance gene, plasmid rescue element, and minimal borders designed for production of gene knockouts. The pCSA110 vector contained an additional fragment with a GUS reporter gene driven by the LAT52 pollen-specific promoter (Eyal *et al.* 1995). This vector was transformed into *qrt/qrt* plants (Preuss *et al.* 1994) to facilitate segregation analysis in pollen for related projects. Vectors were electroporated into disarmed *Agrobacterium tumefaciens* strain C58 cIrif containing the helper plasmid pMP90RK (Koncz and Schell 1986). Colonies with the inserted binary vector were selected on Luria plates supplemented with carbenicillin (100 μg/ml) and rifampicin (400 μg/ml). Single colonies were transferred to 5-ml liquid cultures of Luria broth with antibiotics (Sigma, St. Louis) and grown overnight at 28° with shaking. The entire culture was then transferred to 500 ml of fresh medium and grown to $OD_{600} = 2$.

Transformed plants were generated with the vacuum infiltration method (Bechtold and Pelletier 1998) with modifications (Patton *et al.* 1996) and with the floral dip method (Clough and Bent 1998). Transformation efficiency ranged from 1 to 3% of progeny seeds. Optimal results were obtained with the dip method. Flowering plants were submerged up to the basal rosette for 10–30 sec in 500 ml of 10 mм $MgCl_2$ and 0.01% Silwet L-77 (Lehle) containing Agrobacterium cells from a 500-ml overnight culture. These plants were then subjected to the same treatment again 5 days later. Seeds from these plants were bulk harvested at maturity. Transformants carrying the pPCVICEn4HPT vector were selected on agar plates supplemented with Murashige and Skoog (1962) salts and 25 μg/ml hygromycin B (Sigma). All other transformants were selected in soil by spraying seedlings with 240 μg/ml glufosinate ammonium (BASTA; Hoechst, Frankfurt, Germany) in 0.005% Silwet L-77. Bulk seeds were mixed with filtered sand and spread over the soil surface with a salt shaker to a final density of 40 seeds/cm to allow for optimal selection. Seedlings were sprayed three times over a 2-week period with a Preval disposable paint applicator (Precision Valve, Yonkers, NY) and once after transplantation to Aratrays. Immature siliques of resistant transplants were examined under a dissecting microscope for defective seeds. Mature siliques were harvested from individual plants that appeared to be heterozygous for a seed mutation.

Genetic analysis: Ratios of resistant to sensitive (*R*:*S*) seedlings were obtained for each mutant by plating 50 mature seeds (T_2 or occasionally T_3 generation) on each of two 100 × 15-mm petri plates containing the inorganic salts of Murashige and Skoog (1962), 3% glucose, 0.8% (w/v) purified agar, and either 30 mg/liter hygromycin (Sigma) or 50 mg/liter glufosinate ammonium (Crescent Chemical, Hauppauge, NY). Seeds were surface sterilized as described by Errampalli *et al.* (1991) and transferred to specified locations on the agar surface using sterile forceps. Plates were stored at 4° for 2 days to improve germination rates and then placed beneath fluorescent lights and maintained under the same environmental conditions described above for plants grown in pots. Seedlings were scored for resistance 10–14 days after plating. Resistant seedlings selected for transplantation were removed from agar plates with fine forceps, transferred to soil, and covered with plastic for 2 days to maintain high humidity. Seedling mortality was <5% with this method. The number of resistant seedlings transplanted per transgenic line depended on the *R*:*S* ratio observed: 36 plants for a low ratio line (≤5:1) where tagging status could be resolved and 27 plants for moderate ratio lines (5–25:1) where 10 heterozygous transplants were identified and mature seeds harvested. Mutants from parental lines with low *R*:*S* ratios were classified as tagged if no wild-type plants were identified among surviving transplants, possibly tagged if one to three wild types were found, and not tagged if four or more wild types were encountered. Mature T_3 seeds were harvested as individual siliques from at least two heterozygotes per mutant and bulk harvested from the remaining heterozygotes to provide sufficient seed

TABLE 1

Oligonucleotide sequences for TAIL-PCR

Name	Sequence 5′ → 3′	Utility
LB1[a]	GCC TTT TCA GAA ATG GAT AAA TAG CCT TGC TTC C	1° TAIL
LB2[a]	GCT TCC TAT TAT ATC TTC CCA AAT TAC CAA TAC A	2° TAIL
LB3[a]	TAG CAT CTG AAT TTC ATA ACC AAT CTC GAT ACA C	3° TAIL, sequencing
RB1[b]	ATT AGG CAC CCC AGG CTT TAC ACT TTA TG	1° TAIL
RB2[b]	GTA TGT TGT GTG GAA TTG TGA GCG GAT AAC	2° TAIL
RB3[b]	TAA CAA TTT CAC ACA GGA AAC AGC TAT GAC	3° TAIL, sequencing
QRB1[c]	CAA ACT AGG ATA AAT TAT CGC GCG CGG TGT CA	1° TAIL
QRB2[c]	GGT GTC ATC TAT GTT ACT AGA TCG GGA ATT GA	2° TAIL
QRB3[c]	CGC CAT GGC ATA TGC TAG CAT GCA TAA TTC	3° TAIL, sequencing
SKI1[d]	AAT GGG TAA TTA CTC TTT CTT TTC CTC CAT ATT GA	1° TAIL
SKI2[d]	ATA TTG ACC ATC ATA CTC ATT GCT GAT CCA T	2° TAIL
SKI3[d]	TGA TCC ATG TAG ATT TCC CGG ACA TGA A	3° TAIL, sequencing
AD1a[e]	TGW GNA GSA NCA SAG A	TAIL-PCR, TAIL2K
AD1[f]	NTC GAS TWT SGW GTT	TAIL-PCR, TAIL2K
AD2[f]	NGT CGA SWG ANA WGA A	TAIL-PCR, TAIL2K
AD2[g]	AGW GNA GWA NCA WAG G	TAIL-PCR, TAIL2K
AD3[f]	WGT GNA GWA NCA NAG A	TAIL-PCR, TAIL2K
AD5[h]	STT GNT AST NCT NTG C	TAIL-PCR, TAIL2K

[a] For left border region of pCSA104, pCSA110, and pDAP101.

[b] For right border region of pCSA104 and pDAP101.

[c] For right border region of pCSA110.

[d] For left border region of pPCVICEn4HPT and pSKI015.

[e] A modified AD1 as compared to LIU and WHITTIER (1995).

[f] LIU *et al.* (1995).

[g] LIU and WHITTIER (1995).

[h] TSUGEKI *et al.* (1996).

stocks for subsequent analysis. Plant tissue from tagged mutants was frozen in liquid nitrogen, stored at −80°, and sent to Syngenta on dry ice for molecular analysis.

Phenotypic analysis: Mutants were placed into broad phenotypic classes on the basis of terminal seed phenotypes observed under a dissecting microscope (MEINKE 1994). This was first done with primary transformants and later confirmed by analysis of progeny plants. Mutant seeds that lacked a visible embryo were classified as zygotic (very small seed) or preglobular (intermediate seed size) even when a globular embryo was later found with Nomarski optics (MEINKE 1994). Mutants classified as globular contained a spherical arrested embryo that was visible upon dissection; transition mutants produced a range of globular, triangular, and heart-shaped embryos; and cotyledon mutants contained embryos with more advanced cotyledons. The remaining *emb* mutants were classified as variable (wide range of seed sizes and embryo stages) or other (unusual features not found in standard classes). Seeds with a glassy appearance characteristic of *titan* mutants (LIU and MEINKE 1998) were examined further with Nomarski optics. Pigment mutants without defects in seed morphogenesis were classified as albino (extensive loss of seed pigmentation), pale mature (moderate loss of seed pigmentation), or fusca (inappropriate accumulation of anthocyanin). All genetic and phenotypic data were assembled into an Access (Microsoft, Redmond, WA) database designed to facilitate data tracking.

Plasmid rescue: During the initial phase of this project, plasmid rescue was used to recover border sequences flanking T-DNA inserts (FELDMANN 1992). Genomic plant DNA (∼1 μg) isolated according to REITER *et al.* (1992) was digested with restriction enzyme (New England Biolabs, Beverly, MA) and resulting fragments were ligated and electroporated into *Escherichia coli* cells (XL-1 Blue; Stratagene) using standard protocols. Resistant colonies were selected on 100 μg/ml am-

picillin, plasmid DNA was isolated using QIAGEN (Valencia, CA) or Promega (Madison, WI) kits, and selected clones were analyzed by restriction digestion and sequencing.

Thermal asymmetric interlaced PCR: Most border sequences were recovered in a high-throughput manner using a modified thermal asymmetric interlaced (TAIL)-PCR protocol (LIU *et al.* 1995). DNA from lyophilized plant tissue was obtained using the Puregene DNA isolation kit (Gentra Systems, Minneapolis) with the following modifications. After resuspension of the DNA pellet in hydration buffer, samples were treated with RNase I, incubated at 37° for 30 min, extracted with phenol:chloroform:isoamyl alcohol (24:24:1), and extracted again with chloroform:isoamyl alcohol. DNA was precipitated with 0.1 volume 3 M sodium acetate, pH 5.2, and an equal volume of isopropyl alcohol. Samples were then centrifuged at $20,000 \times g$ for 5 min, and the pellet was rinsed with 70% ethanol, air dried, resuspended in TE buffer at pH 8.0, and analyzed by agarose gel electrophoresis. Several modifications were made in the TAIL-PCR method of LIU *et al.* (1995). The amount of genomic DNA was reduced to 1 ng per primary reaction; T-DNA-specific primers with a higher T_m were designed (Table 1); and the tertiary reaction volume was reduced to 50 μl. With PE 9600 thermocyclers (Applied Biosystems, Foster City, CA), reaction mixes were held on ice until the block warmed to 30°. When MJ Research (Waltham, MA) Tetrad DNA Engines were used, a 2-min hold at 4° was added to the beginning of the program and reactions were placed on the thermocycler block during this refrigerated dwell. The slow ramp program step used with these machines was 0.2°/sec to 72°.

To improve the efficiency of TAIL-PCR, an abbreviated protocol (TAIL2K) was developed where the secondary and tertiary reactions were replaced with a single 40-μl reaction subjected to modified cycling conditions. Initial linear amplification

TABLE 2

Summary of insertional mutagenesis project

Project objective	Results to date
T-DNA insertion lines screened	>120,000
Putative seed mutants identified	>4,200
Putative seed mutants analyzed	2,131
Confirmed *emb* mutants	1,501
Confirmed pigment mutants	204
Tagged *emb* mutants	354
Tagged pigment mutants	48

with the secondary T-DNA primer was carried out for 5 cycles (94°, 10 sec; 64°, 1 min; 72°, 2.5 min), followed by 15 "super" cycles (94°, 10 sec; 64°, 1 min; 72°, 2.5 min; 94°, 10 sec; 64°, 1 min; 72°, 2.5 min; 94°, 10 sec; 44°, 1 min; 72°, 2.5 min). Reactions were then subjected to a low annealing temperature PCR for 5 cycles (94°, 10 sec; 44°, 1 min; 72°, 3 min). A final extension at 72° for 5 min was followed by a hold at 4°. The MJ Research thermocycler was utilized in the calculated temperature control method to execute cycling to emulate the PE9600.

Twelve parallel TAIL-PCR reaction sets were performed on DNA recovered from each tagged mutant. Six degenerate (AD) primers (Table 1) were used in combination with two sets of specific primers, one set for each T-DNA border. A 5-μl sample of each completed TAIL-PCR reaction was subjected to electrophoresis on a 2% agarose gel. Successful TAIL-PCR reactions (10 μl) were incubated at 37° for 1 hr with 2 units of exonuclease I (United States Biochemicals, Cleveland) and 0.5 units of shrimp alkaline phosphatase (United States Biochemicals) added in 2 μl of 1× PCR buffer, followed by a 10-min 70° heat inactivation step. The treated reaction mixture was used as template for direct DNA sequencing (ABI Prism 377) with Big Dye terminator chemistry (Applied Biosystems), using the tertiary border primer.

Border confirmation and second border rescue: Borders recovered through plasmid rescue were confirmed using Southern blots probed with PCR products generated from Arabidopsis sequences near the insertion point. The presence of a polymorphic hybridizing band with DNA from heterozygous plants was used as evidence that a legitimate border had been identified. With borders isolated through TAIL-PCR, insertion points identified by BLAST (Altschul *et al.* 1997) comparison to Arabidopsis genomic sequence were confirmed by direct PCR, using a genome-specific primer (GSP) located 100–500 bases from the insertion site (oriented toward the insertion) in combination with the secondary or tertiary T-DNA primer (Table 1). All GSPs were designed as pairs flanking an insertion point. Each GSP was then paired with the appropriate border primer in confirmation reactions, using DNA from heterozygous plants. A wild-type reaction with GSPs alone was run as a control. When TAIL-PCR failed to recover both sides of an insertion, two additional GSPs more distant (~2 and 5 kb) from the insertion site and pointing toward the expected second border were designed. Long PCR conditions (Advantage 2; Clontech, Palo Alto, CA) were then employed following the manufacturer's directions. Seven PCR reactions were performed using appropriate pairs of genomic and T-DNA border primers. When this method failed, distant GSPs were paired with internal T-DNA primers designed at 1-kb intervals in both orientations across the entire T-DNA. Any borders recovered with this approach were classified as abnormal because they lacked the expected 24-bp T-DNA

Strategy for Genetic Analysis of Seed Mutants

Screen Parental (T_1) Transplants for Defective Seeds

Harvest Mature Progeny (T_2) Seeds from Putative Mutants

Grow Progeny (T_2) Plants in Soil; Analyze Seed Phenotype

Plate T_2 Seeds on Selection Medium; Calculate R:S Ratio

Low R:S   Transplant Seedlings; Resolve Tagging Status

Mod R:S   Transplant Seedlings; Harvest Progeny Seeds

Plate Seeds from 10 Different Subfamilies

Identify Low R:S Subfamily; Resolve Status

High R:S   Stop Analysis Unless Phenotype Exceptional

FIGURE 1.—Identification of tagged mutants defective in seed development.

imperfect repeat characteristic of right and left borders. Mutants that failed this abnormal second border rescue were set aside.

Border sequence analysis: Sequence trace files for each tagged mutant were imported into a mutant-specific Sequencher project (Gene Codes, Ann Arbor, MI). Trace files were typically examined for overall quality and manually trimmed. Consensus sequences assembled from multiple PCR products were then compared with the Arabidopsis genome using BLAST (Altschul *et al.* 1997). All sequence information was entered into a work prioritization database where border recovery status was tracked.

RESULTS

Identification of mutants: The results of our efforts to saturate the genome with seed mutations are summarized in Table 2. More than 120,000 independent insertion lines were generated and siliques of resistant transplants were screened for defective seeds. This strategy made more efficient use of growth space than screening sibling plants in the next generation. Of the 4200 putative seed mutants identified through our initial screen, ~70% were defective in embryo morphology and the remainder were altered in seed pigmentation. The frequency of embryo defectives identified (2.5% of lines screened) was similar to that reported with smaller populations (Castle *et al.* 1993; Devic *et al.* 1996). To estimate the degree of saturation achieved, we first examined segregation of the selectable marker in several hundred lines chosen at random. The results indicated that T_1 plants averaged 1.5 insertion sites and that the entire population contained at least 180,000 independent insertions.

TABLE 3

Expected response of seedlings on selection medium

Total inserts[a]	Mutant tagged	Unlinked inserts[b]	Ratio $R:S$ seedlings[c]	% EMB/EMB transplants[d]
1	Yes	0	2	0
2	Yes	1	11	27
3	Yes	2	47	32
1	No	1	3	33
2	No	2	15	33
3	No	3	63	33

Parental (T_1) lines are assumed to be heterozygous for an *emb* mutation that prevents survival of homozygotes.

[a] Number of functional insertion sites in parental line.

[b] Unlinked relative to *EMB* locus.

[c] Expected ratio of resistant to sensitive seedlings on selection medium.

[d] Percentage of resistant (T_2) transplants that should appear wild type.

FIGURE 2.—Distribution of parental $R:S$ ratios for 229 tagged mutants (black bars) and 487 untagged mutants (gray bars) from the low ratio class indicative of a single site of T-DNA insertion. For each mutant, ratios were calculated for 100 seedlings on selection medium and rounded to the nearest 0.5 for this graph. Results include both embryo-defective and seed pigment mutants.

Genetic characterization of mutants: The strategy developed for high-throughput genetic analysis of mutants recovered from this screen is outlined in Figure 1. During the first phase of the project, T_2 seeds from each putative mutant were planted in soil and grown to maturity to confirm the presence of a seed phenotype and then plated on selection medium to estimate the number of unlinked insertion sites and resolve tagging status. To maximize throughput in later phases of the project, we eliminated screening of T_2 plants in soil except when seed was limited. Tagging status was resolved by transplanting resistant seedlings from putative mutants with a low $R:S$ ratio indicative of a single insertion site and screening the resulting plants for defective seeds. Expected responses of tagged and untagged mutants are summarized in Table 3. Mutants with T-DNA inserts linked to the mutation were expected to give intermediate results. Tagged mutants with a single functional insert should produce only heterozygous (*emb/EMB*) transplants because wild-type (*EMB/EMB*) seedlings lack the resistance (T-DNA) marker associated with the mutation and are eliminated during selection.

The difference between $R:S$ ratios expected for tagged and untagged mutants could not be used to resolve tagging status given the small number of seedlings scored. This is illustrated in Figure 2 with a histogram of ratios obtained for a large collection of mutants recovered from lines with a low $R:S$ ratio. Although the average ratio for tagged mutants (2.2 ± 1.2) was lower than that for untagged mutants (3.0 ± 1.0), consistent with results predicted from Table 3, there was sufficient overlap to require analysis of transplants to resolve tagging status. Resistant seedlings from mutants with a moderate ratio indicative of two unlinked insertion sites were also transplanted to soil but in this case progeny seeds from 10 heterozygotes were plated again to identify subfamilies
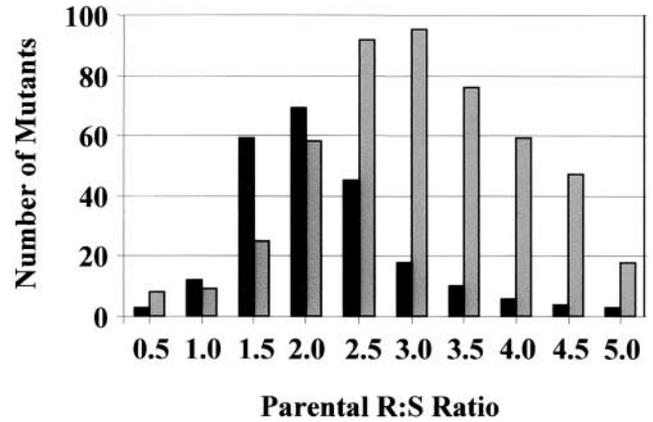
without the second insert. Analysis was usually terminated if the desired subfamily with a low $R:S$ ratio was not identified. Mutants with high ratios indicative of multiple insertion sites required more extensive work to resolve tagging status and were not usually examined further.

More than 2100 putative seed mutants identified through the initial screen of parental lines were subjected to genetic analysis using this approach. Included in this collection were 1501 *emb* and 204 pigment mutants with confirmed seed phenotypes, ~20 mutants with defects seen only after germination, and ~220 putative mutants from high-ratio lines where seed phenotypes were not confirmed because resistant seedlings were not transplanted. Another 220 lines with no apparent phenotype were found, most of them early in the project before screening of parental (T_1) transplants had been perfected. Forty lines segregating for two different seed mutations were also encountered. Half of these lines carried two different embryo defectives. Most of the others had one pigment mutation and one embryo defective. As shown in Table 4, approximately one-half of the confirmed mutants contained a single insertion site on the basis of genetic data, another 20–30% appeared to contain two unlinked insertions, and the remainder gave high ratios indicative of multiple insertion sites. The proportion of mutants assigned to each ratio class was consistent between embryo defectives and pigment mutants.

The tagging status of >1000 confirmed seed mutants has been resolved to date (Table 5). This includes >90% of the mutants recovered from parental lines with one or two insertion sites. A total of 354 tagged mutants defective in embryo development and another 48 tagged mutants altered in seed pigmentation have been identified

TABLE 4

Parental *R:S* ratios of seed mutants analyzed

Class[a]	R:S	Inserts[b]	% embryo defectives[c,e]	% pigment mutants[d,e]
S	0	0	1	1
L	0.1–5	1	52	51
M	5.1–15	2	24	21
MH	15.1–25	2–3	11	8
H	25.1–99	3	8	9
R	100	3+	4	10

[a] Based on the ratio of resistant to sensitive T_2 seedlings on selection medium: all sensitive (S), low (L), moderate (M), moderately high (MH), high (H), and all resistant (R).

[b] Estimated number of functional T-DNA insertion sites based on genetic data.

[c] Calculated for 1219 *emb* mutants.

[d] Calculated for 143 pigment mutants.

[e] Some mutants were not assigned to a parental ratio class because T_2 seed was limited and *R:S* ratios were determined in subsequent generations.

TABLE 5

Tagging status of seed mutants with different parental *R:S* ratios

Class[a]	Tagged	Not tagged	Possibly tagged[b]	Unresolved	Total
S	0	11[c]	0	0	11
L	227	487	16	13	743
M	100	155	12	60	327
MH	28	54	4	48	134
H	3	1	0	115	119
R	0	2	0	68	70
Total	358	710	32	304	1404

Numbers represent confirmed mutants assigned to each ratio class. Some mutants were not assigned because T_2 seed was limited and *R:S* ratios were determined in subsequent generations.

[a] Ratio of resistant to sensitive plants as defined in Table 4.

[b] Genetic evidence inconclusive; rare wild-type transplants found.

[c] Some of these mutants may be tagged with a nonfunctional insert.

(Table 2). This represents a significant advance when compared to past studies (CASTLE *et al.* 1993; YADEGARI *et al.* 1994; DEVIC *et al.* 1996). Ratios were calculated for ~450,000 seedlings on selection plates and >30,000 resistant transplants were screened for defective seeds. Most of the unresolved mutants contained multiple insertion sites. The number of untagged mutants identified was about twice the number of tagged mutants. The frequency of tagged mutants (~35% of total mutants resolved) was independent of the initial number of insertion sites. In other words, tagged mutants were identified from parental lines with one and two insertion sites at about the same frequency. The tagging status of some mutants classified as possibly tagged was resolved by comparing border sequences obtained from heterozygous and sibling wild-type transplants.

Phenotypic characterization of mutants: Terminal phenotypes of mutant embryos were first noted for parental (T_1) plants and then later confirmed in subsequent generations. A simple classification system based on seed size and embryo morphology as viewed through a dissecting microscope was used to assign mutants to phenotypic classes. Detailed examination of cleared seeds with Nomarski optics was reserved for mutants with exceptional defects and mutants with glassy seeds indicative of a *titan* phenotype (LIU and MEINKE 1998; MCELVER *et al.* 2000). The distribution of terminal phenotypes observed (Table 6) was similar to that reported for other mutant collections (CASTLE *et al.* 1993). The large number of tagged mutants identified here made it possible to compare terminal phenotypes of tagged mutants to those of the entire collection. If most mutants with phenotypes extending beyond the globular stage were weak alleles of genes required at early stages of development, then tagged mutants should on average arrest at some-

what earlier stages because many T-DNA insertions are likely to be null alleles (KRYSAN *et al.* 1999). The results as illustrated in Table 6 do not support this conclusion. The distributions of phenotypic classes for tagged mutants were not significantly different from those observed with the entire collection.

Recovery of border sequences: A classification system was devised to summarize the status of border sequence identification for tagged mutants. The number of mutants assigned to each class is shown in Table 7. Class A mutants had sequence information from both sides of the insert and were therefore assigned the highest confidence level for gene identification. Most of the class B mutants, with sequence data from only one side of the insert, were assumed to have a truncated or rearranged insert that interfered with recovery of the second border. Although preliminary gene assignments for many of these mutants are likely to be correct, the presence of an extended deletion adjacent to an aberrant border could in some cases remove a linked gene responsible for the mutant phenotype. Mutants assigned to this class have therefore been excluded from the functional analyses described later. Class Cs mutants had sequence information from both sides of the insert but appeared to contain a deletion that spanned several adjacent genes. These mutants could nevertheless be included with Class A/B mutants on the physical map because the phenotype was associated with a defined chromosomal locus. Class C mutants had sequence data from two or more linked borders, indicative of chromosomal rearrangements or multiple linked inserts. Mutants with even more complex patterns of unlinked insertion sites were assigned to Class D. Some of these mutants could with further work be resolved and upgraded.

TABLE 6

Terminal phenotypes of seed mutants analyzed

Phenotypic class	% total mutants	% tagged mutants
Embryo defectives (1501 total; 354 tagged)		
Zygotic	2	2
Zygotic/preglobular	3	2
Preglobular	20	19
Preglobular/globular	6	6
Globular	30	35
Transition	13	16
Cotyledon	16	12
Variable	8	7
Other	2	1
Total	100	100
Pigment mutants (204 total; 48 tagged)		
Fusca	5	8
Albino	52	56
Pale mature	43	36
Total	100	100

TABLE 7

Summary of border sequence identification

Mutant class	No. of mutants	Sequenced borders	Border sequence identity
A	129	2	Single gene region identified
B	51	1	Single gene region identified
Cs	11	2	Deletion of adjacent genes
C	27	≥ 2	Linked sites, not adjacent
D	56	≥ 2	Unresolved, conflicting data
P	71	≤ 1	Molecular analysis pending
F	8	0	Border rescue failed

Border sequence assignments: The process of assigning border sequences to a specific genomic locus is illustrated in Table 8. Results are presented for 12 class A mutants representative of the collection. In each case, consensus flanking sequences from both sides of the insert matched the same locus. The next most homologous locus differed enough in sequence to allow confirmation that the correct gene had been identified. Rare mismatches between consensus border sequences and the corresponding genomic locus were not unexpected given the high throughput nature of the project. Mutants where such a definitive genome assignment could not be made were set aside for further analysis.

Distribution of target genes: One hundred and ninety-one tagged mutants with sequenced insertions at a single chromosomal location were chosen for detailed analysis. These mutants were from classes A, B, and Cs as defined above. The 167 *EMB* genes disrupted in these 191 mutants are distributed throughout the genome. The number of target genes identified on each chromosome (Table 9) is proportional to the amount of DNA present and the total number of predicted genes (Arabidopsis Genome Initiative 2000). A sequence-based physical map of *EMB* genes identified to date is shown in Figure 3. Results are shown for genes represented by a single mutant allele (left line of each chromosome set) and more than one mutant allele (right line). The distribution of target genes along the length of each chromosome appears to be random with the exception of centromeric regions, which are known to contain a lower density of genes. The combined results of genetic (Franzmann *et al.* 1995) and physical maps support the conclusion that *EMB* genes are dispersed throughout the genome. The relationship between these gene loca-

tions and the presence or absence of localized genome duplications remains to be determined.

Identification of target genes: Predicted functions of the 12 *EMB* genes noted in Table 8 are listed in Table 10. Included are a variety of metabolic enzymes; chromosome scaffold proteins; enzymes required for protein synthesis, folding, and turnover; and proteins of unknown function. The remaining gene identities will be disclosed through a National Science Foundation-funded 2010 project designed to facilitate the identification and analysis of essential genes in Arabidopsis. Project deliverables include synthesized information on 300 *EMB* and 75 seed pigment genes; duplicate mutant alleles for many of these genes identified through a combination of forward and reverse genetics; distribution quantities of mutant seeds available to the community; information on gene expression patterns, mutant phenotypes, and related sequences; and a database accessed via the internet. Target gene identities for most of the class A, B, and Cs mutants described here will be made public according to the schedule presented on the project web page (http://www.seedgenes.org).

Duplicate mutant alleles: Included in the collection of 191 class A, B, and Cs mutants presented here are 146 genes disrupted once, 20 genes represented by two mutant alleles, and 1 gene with five separate alleles. This ~8-kb gene has been identified before (A. Ray, personal communication) and is known as *SUS1* (Schwartz *et al.* 1994), *SIN1* (Ray *et al.* 1996), and *CAF1* (Jacobsen *et al.* 1999). Other published genes found here include *FUS6* (Castle and Meinke 1994), *EDD* (Uwer *et al.* 1998), *RSW1* (Arioli *et al.* 1998), *SLP* (Apuya *et al.* 2001), *KEULE* (Assaad *et al.* 2001), and *RML1/CAD2* (Vernoux *et al.* 2000). The *BIO1* gene inferred from biochemical analysis (Shellhammer and Meinke 1990) and molecular complementation (Patton *et al.* 1996) of mutant embryos was also recovered. The frequency of duplicate mutant alleles identified is similar to that described for another collection of mutants examined in detail (Franzmann *et al.* 1995). Seed mutations in that project were mapped relative to visible markers and complementation tests were performed between mutants that mapped to similar regions of the

TABLE 8

Examples of border sequence assignments to genome locations

Line no.[a]	Accession no.[b]	Insert position[c]	Consensus sequence length[d]	Genome sequence identity	
				Best match[e]	Next match[f]
14986	AP000410	51816	631	538/549	20/20
		54505	690	572/572	22/23
55132	AL138650	15928	761	645/645	19/19
		15953	773	643/643	22/22
29862	AL162295	54121	496	421/423	20/20
		54134	765	683/687	20/20
01917	AL161565	9010	1479	1299/1344	56/61
		9011	1771	743/746	261/282
16713	AB023039	46037	1031	894/903	86/104
		46227	673	624/626	27/28
36172	AB011482	51071	739	641/644	19/19
		51981	738	645/648	26/28
29863	AC025417	103077	660	528/550	21/22
		103096	388	309/314	18/18
00978	AC016972	58221	634	589/589	18/18
		58301	425	394/394	63/68
18048	AC006439	66380	624	486/488	25/26
		66394	606	416/419	18/18
18247	AC007290	6171	666	517/519	18/18
		6194	238	153/156	19/19
21157	AC006929	60173	338	231/232	21/21
		60189	601	521/523	29/32
07724	AC002387	65413	483	391/391	128/132
		65432	726	640/642	110/113

[a] Reference number for mutant in Syngenta collection.

[b] Contains best border sequence match and provides framework for insertion site coordinates.

[c] Coordinate numbers decrease from insertion site for first border and increase from insertion site for second border.

[d] Edited sequence from TAIL-PCR products, plasmid rescue inserts, adaptor or direct PCR products, or some combination thereof. Differs from number of bases in best match, identified from BLAST output, due to presence of vector sequences at product ends.

[e] Number of sequence identities with the assigned locus/number of bases identified from BLAST output.

[f] Number of sequence identities with a separate locus/number of bases identified as next best sequence match.

genome. In both of these studies, the occurrence of duplicate mutant alleles is consistent with the presence of ~500–750 target genes essential for seed development. Table 11 compares the results obtained here with those expected for a genome containing 500–1000 target *EMB* genes with equal rates of mutation. Although this method has obvious limitations (FRANZMANN *et al.* 1995) related to differences in gene size and susceptibility to T-DNA disruption, the results obtained support the conclusion that a rather small number of Arabidopsis genes are readily disrupted by mutation to give a seed phenotype.

One issue that could be addressed for the first time with this large collection of tagged mutants was whether *EMB* genes disrupted by T-DNA insertion were larger than average. Our analysis revealed that the average size (4.4 ± 2.6 kb) of the predicted genomic coding region for 21 *EMB* genes represented by duplicate mutant al-

leles was indeed larger than the average (2.0 ± 1.5 kb) reported for the entire genome (ARABIDOPSIS GENOME INITIATIVE 2000). This is consistent with the accepted conclusion that large genes are more frequent targets for T-DNA insertion than small genes (KRYSAN *et al.* 1999). We were more surprised to find that *EMB* genes (class A) represented by a single mutant allele were also somewhat larger than average (2.9 ± 2.1 kb). This suggests that forward genetic screens for seed mutants have to date favored the identification of somewhat larger than average genes and may have missed knockouts in many small genes with essential functions during seed development. Estimates of target *EMB* gene numbers in Arabidopsis may therefore need to be increased somewhat as additional data are obtained from saturation mutagenesis and reverse genetic experiments. Whether essential genes are larger or smaller on average than other Arabidopsis genes also remains to be resolved.

TABLE 9

Chromosomal distribution of tagged *EMB* genes

Chromosome	Predicted genes[a]	% total	*EMB* genes identified	% total *EMB*	Chi square[b]
1	6,543	25.7	48	28.7	0.8
2	4,036	15.8	22	13.2	0.9
3	5,220	20.5	29	17.4	1.0
4	3,825	15.0	21	12.6	0.8
5	5,874	23.0	47	28.1	2.5
Total	25,498	100.0	167	100.0	—

Genes represented by mutants with single insertion site are as defined in text.

[a] Number of genes predicted by the ARABIDOPSIS GENOME INITIATIVE (2000).

[b] Distribution of *EMB* genes is not significantly different from that expected based on the total number of predicted genes on each chromosome.

Functional diversity of genes identified: All *EMB* sequences were searched against the GenBank nonredundant peptide, nucleotide, EST, and SwissProt databases using the default BLAST 2.0 algorithm. A cutoff of $\leq 1e^{-25}$ expected value was used for a positive match. With small proteins that could not reach this level with a perfect match, a 70% amino acid identity threshold was used. Detailed analyses were performed on 110 *EMB* genes (class A) with confirmed insertion sites. Forty-six of these (42%) could not be assigned to a functional class, consistent with percentages observed for the entire genome (ARABIDOPSIS GENOME INITIATIVE 2000). The remaining 64 proteins spanned a wide range of functional classes, from metabolism (15), cell growth (8), transcription (17), translation (9), and protein fate (11) to protein transport and traffic (2) and plant defense (2). On the basis of TargetP analysis (EMANUELSSON *et al.* 2000), >25% of the 110 *EMB* proteins included in this study appear to be targeted to chloroplasts. This is about twice the percentage expected on the basis of
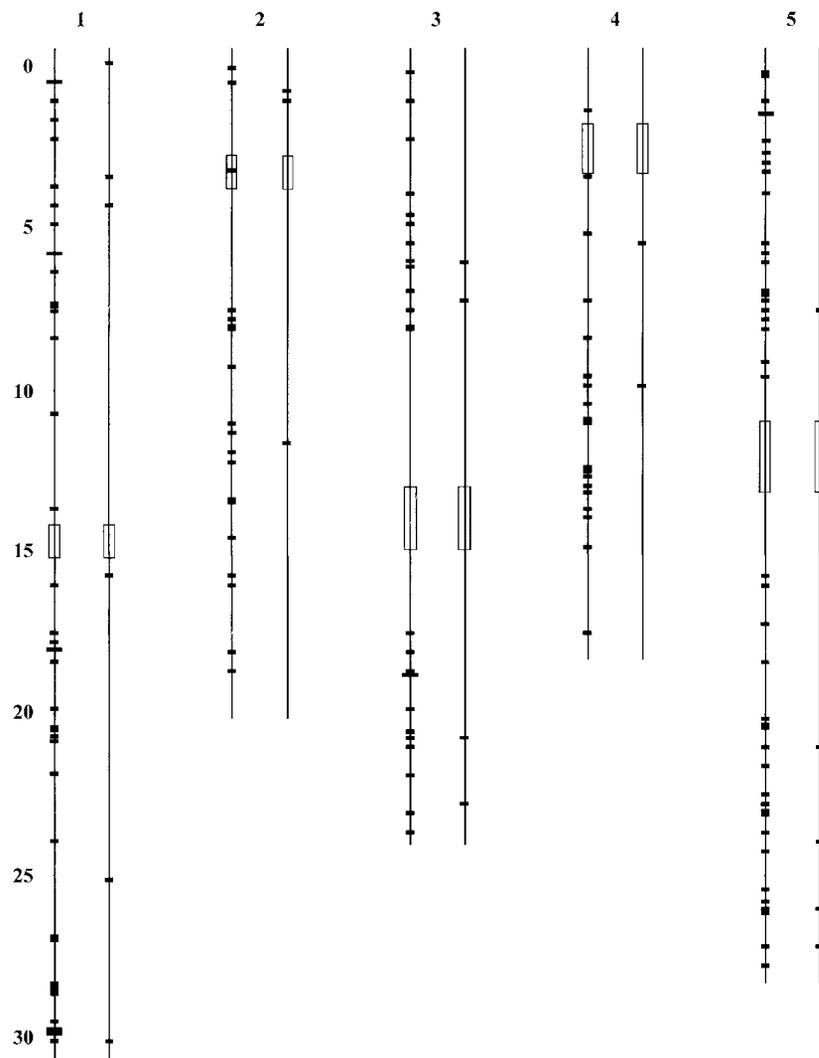

FIGURE 3.—Localization of 167 *EMB* genes on a sequence-based physical map of Arabidopsis. Each bar represents a tagged gene identified from a collection of 191 mutants (classes A, B, and Cs) with a single locus of insertion. Genes disrupted once in this collection are shown on the left line of each chromosome set. Those represented by multiple mutant alleles are shown at right. Long bars designate sequenced Arabidopsis Genome Initiative (AGI) clones containing 2 different *EMB* genes. Thick bars represent *EMB* genes on linked clones. Open rectangles correspond to centromeric regions as defined by genetic analysis (ARABIDOPSIS GENOME INITIATIVE 2000). Numbers indicate the estimated length of each chromosome in megabases. The line for each chromosome has been extended 0.5 Mb beyond the telomere to facilitate visualization of gene locations. Nucleolar organizer regions located on chromosomes 2 and 4 are not shown.

TABLE 10

Examples of essential genes identified

Line no.	Locus symbol	BAC location	AGI name	Predicted protein product
14986	*TTN6*	K10D20.17	At3g20625	Iso-T deubiquitinating enzyme
55132	*TTN8*	T5N23.30	At3g54670	SMC1 cohesin
29862	*TTN1*	T4C21.150	At3g60740	Tubulin folding cofactor D
01917	*EMB1027*	T25K17.110	At4g26300	Arginyl-tRNA synthetase
16713	*EMB1276*	MIF21.12	At5g48230	Acetyl-CoA C-acetyltransferase
36172	*BIO1*	MUA2.17	At5g57590	7-KAP: DAP aminotransferase
29863	*EMB1586*	T12C24.30	At1g12770	DEAD/DEAH box helicase
00978	*TTN4*	F23N20.18	At1g71190	Unknown protein
18048	*TTN5*	T30D6.10	At2g18390	ARL2 GTPase
18247	*TTN7*	F20F1.4	At2g27170	SMC3 cohesin
21157	*SLP*	T1E2.8	At2g28000	Chaperonin-60α
07724	*EMB1067*	F4L23.16	At2g45330	Unknown protein

genome-wide predictions. In contrast, the proportion of *EMB* proteins predicted to be associated with secretory pathways is lower than expected.

Most *EMB* proteins with assigned functions have apparent homologs in other organisms. In contrast, most *EMB* proteins with uncertain functions are plant-specific based on the BLAST criteria described. Six of these proteins have to date been identified only in Arabidopsis. Another two unassigned proteins are found only in plants and bacteria; one appears to be localized to chloroplasts and the other to mitochondria. About one-half of the unassigned proteins contain at least one PROSITE (Hofmann *et al.* 1999) or Pfam (Bateman *et al.* 2000) motif that could with further analysis provide some insight into protein function. The other one-half may require more detailed experimentation before predictions of protein function can be made.

TABLE 11

Number of *EMB* genes represented by $\geq$1 mutant alleles

Mutant alleles	Syngenta population[b]	Predicted[a] 500	1000	Franzmann *et al.* population[c]	Predicted 500	1000
1	146	132	158	95	100	115
2	20	25	15	11	13	7
3	0	3	1	4	1	0
4	0	0	0	0	0	0
5	1	0	0	0	0	0

[a] Observed and expected number of genes disrupted once, twice, and three or more times in collections of 191 mutants (Syngenta) and 129 mutants (Franzmann *et al.* 1995). Predictions assume that all genes mutate at equal frequency (Franzmann *et al.* 1995) and that either 500 or 1000 target *EMB* genes are present in the genome.

[b] Limited to class A, B, and Cs mutants as defined in the text.

[c] Mutants were mapped relative to visible markers and duplicate alleles were identified through complementation tests (Franzmann *et al.* 1995).

DISCUSSION

Assigning functions to all 26,000 Arabidopsis genes is a monumental task that will require complementary research strategies and extensive multinational coordination. Part of this effort must be devoted to characterizing gene expression patterns and validating biochemical functions of protein products within the plant cell. A second critical component will be to determine the biological significance of each gene product and its relevance to plant growth and development. This will require forward genetics and saturation mutagenesis on a genomics level. When gene functions are not redundant, this objective can be met by determining which gene disruptions result in a mutant phenotype.

We describe here an efficient system for identifying large numbers of genes with essential and nonredundant functions during seed development. More than 120,000 T-DNA insertion lines were screened for defects in seed development, 4200 seed mutants with a wide range of defects were identified, 1705 confirmed mutants were analyzed in detail, 354 tagged embryo-defective mutants and 48 tagged pigment mutants were recovered, 24 examples of duplicate alleles were discovered, and 167 *EMB* genes with confirmed insertion sites were placed on the physical map. Related efforts to identify genes required for seedling development are described in a companion article (Budziszewski *et al.* 2001). When combined with results from other laboratories working at different stages of the life cycle, these projects should help to identify the full complement of essential Arabidopsis genes and define the minimal gene set required to produce a flowering plant.

One question that needs to be addressed is whether mutants defective in essential genes should by definition (a) fail to complete the life cycle, (b) exhibit an obvious phenotype, or (c) display any phenotype from dramatic to subtle. We believe the first definition is too narrow and dependent on growth conditions whereas the third

definition is too broad and requires that a variety of methods be used to search for minor defects. We have therefore chosen the second definition of an essential gene for this project—one that is needed for normal growth and development and gives rise to a visible phenotype when disrupted. We believe that this definition represents an effective standard for dealing with a diverse collection of plant genes. A distinction must also be made between essential genes, which are amenable to analysis with knockouts because their functions and expression patterns are not duplicated elsewhere in the genome, and essential functions, some of which are encoded by redundant gene families. The essential genes described in this report therefore represent a subset of the minimal gene set required to fulfill every function required for seed development. Alternative strategies will be needed in the future to determine the consequences of disrupting groups of genes with related and/or overlapping functions.

Considerable progress has already been made in the identification of essential genes in Saccharomyces (WINZELER *et al.* 1999; TOHE and OGUCHI 2000) and Caenorhabditis (FRASER *et al.* 2000; GONCZY *et al.* 2000). The systematic approach used to analyze the yeast genome cannot be duplicated with Arabidopsis at present because efficient systems of homologous recombination and gene replacement are not available. The RNA interference methods used to dissect gene function in Caenorhabditis (FRASER *et al.* 2000) show promise with plant systems (CHUANG and MEYEROWITZ 2000; METTE *et al.* 2000) but may be more difficult to exploit given the presence of extensive gene duplications. The best approach available at present is therefore to screen large collections of insertion lines for mutant phenotypes and accumulate information on essential genes associated with a disruption of normal growth and development.

Several different classes of seed mutants must be included to meet this objective: gametophytic mutants where gene expression occurs before fertilization but also affects seed development (SPRINGER *et al.* 1995); embryo defectives characterized by abnormal morphology (MEINKE 1994); seed pigment mutants with normal morphology but altered pigmentation (NORRIS *et al.* 1995); high chlorophyll fluorescence mutants (MEURER *et al.* 1996); and other types of mutants with biochemical defects that exhibit a cellular phenotype without altering seed morphology or pigmentation. We chose to focus on embryo defectives and a limited number of pigment mutants for this study. Additional examples of mutants with pigment defects that become apparent before and after germination are described by BUDZISZEWSKI *et al.* (2001).

Saturating the Arabidopsis genome with mutations affecting seed development presents a significant challenge because of the large number of target genes involved. Confirmation of gene identities through molecular complementation also requires considerable effort

and multiple crosses when working with lethals. The identification of duplicate mutant alleles will therefore play an important role both in assessing the degree of saturation and confirming gene identities inferred from the analysis of a single allele. These additional alleles should come from two major sources: forward genetic screens where the same gene has been disrupted more than once and reverse genetic analysis where an attempt has been made to find another T-DNA insertion within the same gene. The public availability of research facilities designed to expedite reverse genetics in Arabidopsis (KRYSAN *et al.* 1999) should reduce the effort required to confirm most of the target genes found in this project.

Biological activities of genes reported here extend across a broad spectrum of protein functions and cellular compartments. This is not surprising in light of the many biochemical functions required to make a differentiated plant embryo from a heterotrophic zygote. Many genes known to be essential for growth in other organisms nevertheless remain to be identified. The probable explanations are that saturation has not been achieved and that redundancy (genomic and biochemical) has compensated for the loss of many gene products. Detailed comparison of *EMB* gene locations relative to known genome duplications may provide further insights into why the loss of certain gene products is tolerated whereas disruption of others interferes with normal seed development. Some interesting trends are already beginning to emerge. The large number of *EMB* proteins predicted to be targeted to chloroplasts, for example, is consistent with the conclusion that loss of chloroplast function results in embryonic lethality whereas disruption of photosynthetic function alters seed pigmentation but not morphogenesis. Proteins required for transcription, translation, and protein destination also appear to be more common in the present collection than across the whole genome, whereas those involved with intracellular traffic and signaling are underrepresented. The abundance of plant-specific proteins with unclassified functions suggests that plants have diverged from other major groups of organisms with respect to some essential gene functions. Making sense of these functions within a cellular, organismal, and evolutionary context, as stated in Arabidopsis community objectives for the coming decade, will require much more detailed analysis and experimental confirmation in the years ahead.

LITERATURE CITED

AIDA, M., T. ISHIDA, H. FUKAKI, H. FUJISAWA and M. TASAKA, 1997 Genes involved in organ separation in Arabidopsis: an analysis of the *cup-shaped cotyledon* mutant. Plant Cell **9:** 841–857.

ALBERT, S., B. DESPRES, J. GUILLEMINOT, N. BECHTOLD, G. PELLETIER et al., 1999 The *EMB* 506 gene encodes a novel ankyrin repeat containing protein that is essential for the normal development of Arabidopsis embryos. Plant J. **17:** 169–179.

ALTMANN, T., G. FELIX, A. JESSOP, A. KAUSCHMANN, U. UWER et al., 1995 Ac/Ds transposon mutagenesis in *Arabidopsis thaliana*: mutant spectrum and frequency of Ds insertion mutants. Mol. Gen. Genet. **247:** 646–652.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG et al., 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

APUYA, N. R., R. YADEGARI, R. L. FISCHER, J. J. HARADA, J. L. ZIMMERMAN et al., 2001 The Arabidopsis embryo mutant *schlepperless* has a defect in the *chaperonin-60α* gene. Plant Physiol. **126: 717–730.**

ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

ARIOLI, T., L. PENG, A. S. BETZNER, J. BURN, W. WITTKE et al., 1998 Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science **279:** 717–720.

ASSAAD, F., Y. HUET, U. MAYER and G. JURGENS, 2001 The cytokinesis gene *KEULE* encodes a Sec1 protein that binds the syntaxin KNOLLE. J. Cell Biol. **152:** 531–544.

BATEMAN, A., E. BIRNEY, R. DURBIN, S. R. EDDY, K. L. HOWE et al., 2000 The Pfam protein families database. Nucleic Acids Res. **28:** 263–266.

BECHTOLD, N., and G. PELLETIER, 1998 In planta Agrobacterium-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. Methods Mol. Biol. **82:** 259–266.

BOISSON, M., V. GOMORD, C. AUDRAN, N. BERGER, B. DUBREUCQ et al., 2001 *Arabidopsis glucosidase I* mutants reveal a critical role of N-glycan trimming in seed development. EMBO J. **20:** 1010–1019.

BOUCHEZ, D., C. CAMILLERI and M. CABOCHE, 1993 A binary vector based on Basta resistance in plant transformation of *Arabidopsis thaliana*. C. R. Acad. Sci. Ser. III Sci. Vie **316:** 1188–1193.

BUDZISZEWSKI, G. J., S. P. LEWIS, L. W. GLOVER, J. REINEKE, G. JONES et al., 2001 *Arabidopsis* genes essential for seedling viability: isolation of insertional mutants and molecular cloning. Genetics **159:** 1765–1778.

CASTLE, L. A., and D. W. MEINKE, 1994 A *FUSCA* gene of Arabidopsis encodes a novel protein essential for plant development. Plant Cell **6:** 25–41.

CASTLE, L. A., D. ERRAMPALLI, T. L. ATHERTON, L. H. FRANZMANN, E. S. YOON et al., 1993 Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. Mol. Gen. Genet. **241:** 504–514.

CHUANG, C. F., and E. M. MEYEROWITZ, 2000 Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **97:** 4985–4990.

CLOUGH, S. J., and A. F. BENT, 1998 Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735–743.

CONKLIN, P. L., S. R. NORRIS, G. L. WHEELER, E. H. WILLIAMS, N. SMIRNOFF et al., 1999 Genetic evidence for the role of GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis. Proc. Natl. Acad. Sci. USA **96:** 4198–4203.

DEVIC, M., S. ALBERT and M. DELSENY, 1996 Induction and expression of seed-specific promoters in Arabidopsis embryo-defective mutants. Plant J. **9:** 205–215.

EMANUELSSON, O., H. NIELSEN, S. BRUNAK and G. VON HEIJNE, 2000 Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:** 1005–1016.

ERRAMPALLI, D., D. PATTON, L. CASTLE, L. MICKELSON, K. HANSEN et al., 1991 Embryonic lethals and T-DNA insertional mutagenesis in *Arabidopsis*. Plant Cell **3:** 149–157.

EYAL, Y., C. CURIE and S. MCCORMICK, 1995 Pollen specificity elements reside in 30 bp of the proximal promoters of two pollen-expressed genes. Plant Cell **7:** 373–384.

FELDMANN, K. A., 1991 T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. Plant J. **1:** 71–82.

FELDMANN, K. A., 1992 T-DNA insertion mutagenesis in *Arabidopsis*: seed infection/transformation, pp. 274–289 in *Methods in Arabidopsis Research*, edited by C. KONCZ, N.-H. CHUA and J. SCHELL. World Scientific Publishing, Singapore.

FRANZMANN, L. H., E. S. YOON and D. W. MEINKE, 1995 Saturating the genetic map of *Arabidopsis thaliana* with embryonic mutations. Plant J. **7:** 341–350.

FRASER, A. G., R. S. KAMATH, P. ZIPPERLEN, M. MARTINEZ-CAMPOS, M. SOHRMANN et al., 2000 Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature **408:** 325–330.

GONCZY, P., G. ECHEVERRI, K. OEGEMA, A. COULSON, S. J. JONES et al., 2000 Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. Nature **408:** 331–336.

GROSSNIKLAUS, U., J. P. VIELLE-CALZADA, M. A. HOEPPNER and W. B. GAGLIANO, 1998 Maternal control of embryogenesis by *MEDEA*, a *polycomb* group gene in *Arabidopsis*. Science **280:** 446–450.

HARDTKE, C. S., and T. BERLETH, 1998 The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J. **17:** 1405–1411.

HAYASHI, H., I. CZAJA, H. LUBENOW, J. SCHELL and R. WALDEN, 1992 Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. Science **258:** 1350–1353.

HEATH, J. D., R. WELDON, C. MONNOT and D. W. MEINKE, 1986 Analysis of storage proteins in normal and aborted seeds from embryo-lethal mutants of *Arabidopsis thaliana*. Planta **169:** 304–312.

HOFMANN, K., P. BUCHER, L. FALQUET and A. BAIROCH, 1999 The PROSITE database, its status in 1999. Nucleic Acids Res. **27:** 215–219.

JACOBSEN, S. E., M. P. RUNNING and E. M. MEYEROWITZ, 1999 Disruption of an RNA helicase/RNAse III gene in *Arabidopsis* causes unregulated cell division in floral meristems. Development **126:** 5231–5243.

JANG, J. C., S. FUJIOKA, M. TASAKA, H. SETO, S. TAKATSUTO et al., 2000 A critical role of sterols in embryonic patterning and meristem programming revealed by the fackel mutants of *Arabidopsis thaliana*. Genes Dev. **14:** 1485–1497.

JURGENS, G., R. A. TORRES RUIZ and T. BERLETH, 1994 Embryonic pattern formation in flowering plants. Annu. Rev. Genet. **28:** 351–371.

KONCZ, C., and J. SCHELL, 1986 The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol. Gen. Genet. **204:** 383–396.

KRYSAN, P. J., J. C. YOUNG and M. R. SUSSMAN, 1999 T-DNA as an insertional mutagen in Arabidopsis. Plant Cell **11:** 2283–2290.

LAUX, T., and G. JURGENS, 1997 Embryogenesis—a new start in life. Plant Cell **9:** 989–1000.

LI, Z., and T. L. THOMAS, 1998 *PEI1*, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. Plant Cell **10:** 383–398.

LINDSEY, K., J. F. TOPPING, P. S. C. F. DA ROCHA, K. L. HORNE, P. R. MUSKETT et al., 1996 Insertional mutagenesis to dissect embryonic development in Arabidopsis, pp. 51–76 in *Embryogenesis: The Generation of a Plant*, edited by T. L. WANG and A. CUMING. BIOS Scientific Publishers, Oxford.

LIU, C.-M., and D. W. MEINKE, 1998 The *titan* mutants of *Arabidopsis* are disrupted in mitosis and cell cycle control during seed development. Plant J. **16:** 21–31.

LIU, Y. G., N. MITSUKAWA, T. OOSUMI and R. F. WHITTIER, 1995 Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J. **8:** 457–463.

LIU, Y.-G., and R. F. WHITTIER, 1995 Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25: 674–681.

LOTAN, T., M. OHTO, K. M. YEE, M. A. WEST, R. LO et al., 1998 Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell 93: 1195–1205.

LUERSSEN, H., V. KIRIK, P. HERRMANN and S. MISERA, 1998 FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. Plant J. 15: 755–764.

LUKOWITZ, W., U. MAYER and G. JURGENS, 1996 Cytokinesis in the Arabidopsis embryo involves the syntaxin-related KNOLLE gene product. Cell 84: 61–71.

LUKOWITZ, W., C. S. GILLMOR and W. R. SCHEIBLE, 2000 Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiol. 123: 795–805.

LUKOWITZ, W., T. C. NICKLE, D. W. MEINKE, R. L. LAST, P. L. CONKLIN et al., 2001 Arabidopsis cyt1 mutants are deficient in a mannose-1-phosphate guanylyltransferase and point to a requirement of N-linked glycosylation for cellulose biosynthesis. Proc. Natl. Acad. Sci. USA 98: 2262–2267.

MARTIENSSEN, R. A., 1998 Functional genomics: probing plant gene function and expression with transposons. Proc. Natl. Acad. Sci. USA 95: 2021–2026.

McELVER, J., D. PATTON, M. RUMBAUGH, C.-M. LIU, L. J. YANG et al., 2000 The TITAN5 gene of Arabidopsis encodes a protein related to the ADP ribosylation factor family of GTP binding proteins. Plant Cell 12: 1379–1392.

MEINKE, D. W., 1994 Seed development in Arabidopsis, pp. 253–295 in Arabidopsis thaliana, edited by E. M. MEYEROWITZ and C. R. SOMERVILLE. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

MEINKE, D. W., 1995 Molecular genetics of plant embryogenesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46: 369–394.

MEINKE, D. W., and I. M. SUSSEX, 1979 Embryo-lethal mutants of Arabidopsis thaliana: a model system for genetic analysis of plant embryo development. Dev. Biol. 72: 50–61.

METTE, M. F., W. AUFSATZ, J. VAN DER WINDEN, M. A. MATZKE and A. J. MATZKE, 2000 Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. 19: 5194–5201.

MEURER, J., K. MEIERHOFF and P. WESTHOFF, 1996 Isolation of high-chlorophyll-fluorescence mutants of Arabidopsis thaliana and their characterisation by spectroscopy, immunoblotting and northern hybridisation. Planta 198: 385–396.

MORDHORST, A. P., M. A. J. TOONEN and S. C. DEVRIES, 1997 Plant embryogenesis. Crit. Rev. Plant Sci. 16: 535–576.

MURASHIGE, T., and F. SKOOG, 1962 A revised medium for rapid growth and bioassay with tobacco tissue cultures. Physiol. Plant. 15: 473–497.

NORRIS, S. R., T. R. BARRETTE and D. DELLAPENNA, 1995 Genetic dissection of carotenoid synthesis in arabidopsis defines plastoquinone as an essential component of phytoene desaturation. Plant Cell 7: 2139–2149.

OSBORNE, B. I., U. WIRTZ and B. BAKER, 1995 A system for insertional mutagenesis and chromosomal rearrangement using the Ds transposon and Cre-lox. Plant J. 7: 687–701.

PARINOV, S., M. SEVUGAN, Y. DE, W. C. YANG, M. KUMARAN et al., 1999 Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in Arabidopsis. Plant Cell 11: 2263–2270.

PATTON, D. A., S. VOLRATH and E. R. WARD, 1996 Complementation of an Arabidopsis thaliana biotin auxotroph with an Escherichia coli biotin biosynthetic gene. Mol. Gen. Genet. 251: 261–266.

PATTON, D. A., A. L. SCHETTER, L. H. FRANZMANN, K. NELSON, E. R. WARD et al., 1998 An embryo-defective mutant of Arabidopsis disrupted in the final step of biotin synthesis. Plant Physiol. 116: 935–946.

PREUSS, D., S. Y. RHEE and R. W. DAVIS, 1994 Tetrad analysis possible in Arabidopsis with mutation of the QUARTET (QRT) genes. Science 264: 1458–1460.

RAY, S., T. GOLDEN and A. RAY, 1996 Maternal effects of the short integument mutation on embryo development in Arabidopsis. Dev. Biol. 180: 365–369.

REITER, R. S., R. M. YOUNG and P. A. SCOLNIK, 1992 Genetic linkage of the Arabidopsis genome: methods for mapping with recombinant inbred and random amplified polymorphic DNAs (RAPDs), pp. 170–190 in Methods in Arabidopsis Research, edited by C. KONCZ, N.-H. CHUA and J. SCHELL. World Scientific Publishing, Singapore.

SCHRICK, K., U. MAYER, A. HORRICHS, C. KUHNT, C. BELLINI et al., 2000 FACKEL is a sterol C-14 reductase required for organized cell division and expansion in Arabidopsis embryogenesis. Genes Dev. 14: 1471–1484.

SCHWARTZ, B. W., E. C. YEUNG and D. W. MEINKE, 1994 Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of Arabidopsis. Development 120: 3235–3245.

SHELLHAMMER, J., and D. W. MEINKE, 1990 Arrested embryos from the bio1 auxotroph of Arabidopsis thaliana contain reduced levels of biotin. Plant Physiol. 93: 1162–1167.

SOMERVILLE, C., and J. DANGL, 2000 Genomics. Plant biology in 2010. Science 290: 2077–2078.

SPEULMAN, E., P. L. METZ, G. VAN ARKEL, B. TE LINTEL HEKKERT, W. J. STIEKEMA et al., 1999 A two-component enhancer-inhibitor transposon mutagenesis system for functional analysis of the Arabidopsis genome. Plant Cell 11: 1853–1866.

SPRINGER, P. S., W. R. McCOMBIE, V. SUNDARESAN and R. A. MARTIENSSEN, 1995 Gene trap tagging of PROLIFERA, an essential MCM2–3-5-like gene in Arabidopsis. Science 268: 877–880.

SUNDARESAN, V., P. SPRINGER, T. VOLPE, S. HAWARD, J. D. G. JONES et al., 1995 Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev. 9: 1797–1810.

TISSIER, A. F., S. MARILLONNET, V. KLIMYUK, K. PATEL, M. A. TORRES et al., 1999 Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell 11: 1841–1852.

TOHE, A., and T. OGUCHI, 2000 An improved integration replacement/disruption method for mutagenesis of yeast essential genes. Genes Genet. Syst. 75: 33–39.

TOPPING, J. F., F. AGYEMAN, B. HENRICOT and K. LINDSEY, 1994 Identification of molecular markers of embryogenesis in Arabidopsis thaliana by promoter trapping. Plant J. 5: 895–903.

TSUGEKI, R., E. Z. KOCHIEVA and N. V. FEDOROFF, 1996 A transposon insertion in the Arabidopsis SSR16 gene causes an embryo-defective lethal mutation. Plant J. 10: 479–489.

UWER, U., L. WILLMITZER and T. ALTMANN, 1998 Inactivation of a glycyl-tRNA synthetase leads to an arrest in plant embryo development. Plant Cell 10: 1277–1294.

VERNOUX, T., R. C. WILSON, K. A. SEELEY, J. P. REICHHELD, S. MUROY et al., 2000 The ROOT MERISTEMLESS1/CADMIUM SENSITIVE2 gene defines a glutathione-dependent pathway involved in initiation and maintenance of cell division during postembryonic root development. Plant Cell 12: 97–110.

WEIGEL, D., J. H. AHN, M. A. BLAZQUEZ, J. O. BOREVITZ, S. K. CHRISTENSEN et al., 2000 Activation tagging in Arabidopsis. Plant Physiol. 122: 1003–1014.

WINZELER, E. A., D. D. SHOEMAKER, A. ASTROMOFF, H. LIANG, K. ANDERSON et al., 1999 Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285: 901–906.

WISMAN, E., G. H. CARDON, P. FRANSZ and H. SAEDLER, 1998 The behaviour of the autonomous maize transposable element En/Spm in Arabidopsis thaliana allows efficient mutagenesis. Plant Mol. Biol. 37: 989–999.

YADEGARI, R., G. R. DE PAIVA, T. LAUX, A. M. KOLTUNOW, N. APUYA et al., 1994 Cell differentiation and morphogenesis are uncoupled in Arabidopsis raspberry embryos. Plant Cell 6: 1713–1729.

ZHANG, J. Z., and C. R. SOMERVILLE, 1997 Suspensor-derived polyembryony caused by altered expression of valyl-tRNA synthetase in the twn2 mutant of Arabidopsis. Proc. Natl. Acad. Sci. USA 94: 7349–7355.

Communicating editor: C. S. GASSER