

# Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors

Brandon H. Le<sup>a,1</sup>, Chen Cheng<sup>a,1</sup>, Anhtu Q. Bui<sup>a,1</sup>, Javier A. Wagmaister<sup>a,2</sup>, Kelli F. Henry<sup>a</sup>, Julie Pelletier<sup>b</sup>, Linda Kwong<sup>b</sup>, Mark Belmonte<sup>b</sup>, Ryan Kirkbride<sup>b</sup>, Steve Horvath<sup>c</sup>, Gary N. Drews<sup>d</sup>, Robert L. Fischer<sup>e</sup>, Jack K. Okamoto<sup>f</sup>, John J. Harada<sup>b</sup>, and Robert B. Goldberg<sup>a,3</sup>

<sup>a</sup>Department of Molecular, Cell, and Developmental Biology, and <sup>d</sup>Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095; <sup>b</sup>Section of Plant Biology, Division of Biological Sciences, University of California, Davis, CA 95616; <sup>c</sup>Department of Biology, University of Utah, Salt Lake City, UT 84112; <sup>e</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; and <sup>f</sup>United States Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2001.

Contributed by Robert B. Goldberg, March 19, 2010 (sent for review February 19, 2010)

**Most of the transcription factors (TFs) responsible for controlling seed development are not yet known. To identify TF genes expressed at specific stages of seed development, including those unique to seeds, we used Affymetrix GeneChips to profile *Arabidopsis* genes active in seeds from fertilization through maturation and at other times of the plant life cycle. Seed gene sets were compared with those expressed in prefertilization ovules, germinating seedlings, and leaves, roots, stems, and floral buds of the mature plant. Most genes active in seeds are shared by all stages of seed development, although significant quantitative changes in gene activity occur. Each stage of seed development has a small gene set that is either specific at the level of the GeneChip or up-regulated with respect to genes active at other stages, including those that encode TFs. We identified 289 seed-specific genes, including 48 that encode TFs. Seven of the seed-specific TF genes are known regulators of seed development and include the *LEAFY COTYLEDON (LEC)* genes *LEC1*, *LEC1-LIKE*, *LEC2*, and *FUS3*. The rest represent different classes of TFs with unknown roles in seed development. Promoter- $\beta$ -glucuronidase (*GUS*) fusion experiments and seed mRNA localization GeneChip datasets showed that the seed-specific TF genes are active in different compartments and tissues of the seed at unique times of development. Collectively, these seed-specific TF genes should facilitate the identification of regulatory networks that are important for programming seed development.**

embryo | transcriptome | mRNA localization

**S**eed development in higher plants begins with a double fertilization process that occurs within the ovule and ends with a dormant seed primed to become the next plant generation (1, 2). The major events that occur during seed development are shown schematically in Fig. 1 and are described elsewhere in detail (1–8). Genetic studies with *Arabidopsis* have uncovered several genes that play major roles in seed development (9–11), including those that govern endosperm formation (12, 13), embryo differentiation (14, 15), and seed coat development (8). In addition, molecular studies with *Arabidopsis* and other plants have identified the *cis*-control regions of several genes active during seed development, particularly those encoding storage proteins, and the transcription factors (TFs) that play a role in their regulation (16–22). Nevertheless, the identities of most regulators of seed development and their direct targets are largely unknown.

To date, many studies have been carried out by using microarrays to identify genes that are expressed at different times of the plant life cycle (23–26). Only a few of these studies, however, have focused exclusively on seeds and/or embryos to identify important regulators and processes required for seed development (27–31). In this paper, we present results of Affymetrix GeneChip experiments that profile genes that are active before, during, and after *Arabidopsis* seed for-

mation. Our experiments identified 48 TF genes that are active exclusively, or at elevated levels, in seeds. These seed-specific TF genes encode several classes of TFs, are active at different developmental times, and may be important for controlling stage-specific biological events during seed formation. Chimeric promoter- $\beta$ -glucuronidase (*GUS*) transgene experiments and laser capture microdissection (LCM) (32) GeneChip datasets demonstrated that the seed-specific TF genes are active in specific seed compartments and tissues, suggesting that they may play an important role in the differentiation and/or function of unique seed parts. Our data represent an important step toward identifying gene regulatory networks (33) in the *Arabidopsis* genome that are responsible for programming seed development. What these seed-specific TF genes do and how they are integrated into regulatory networks remain to be determined.

## Results

**Overview and GeneChip Analysis.** We carried out GeneChip hybridization experiments (*Materials and Methods*) by using mRNAs isolated from *Arabidopsis* unfertilized ovules (OV); seeds containing (i) zygotes (24H), (ii) globular-stage embryos (GLOB), (iii) cotyledon-stage embryos (COT), (iv) mature green embryos (MG), and (v) postmature green embryos (PMG); and post-germination seedlings (SDLG) to identify genes that are active before, during, and after seed development (Figs. 1 and 24 and *SI Materials and Methods*). These mRNAs represent gene sets that are active during periods when major events occur within the seed—including embryo differentiation, endosperm formation, seed coat development, storage reserve accumulation, and maturation (Fig. 1). We applied a stringent protocol to analyze our GeneChip data and restricted our analysis to mRNAs for which the detection call by the Microarray Analysis Suite (MAS) 5.0 software was P (Present) in both biological replicates to reduce the inclusion of false positives (*Materials and Methods*). Only probe sets with consensus detection calls of PP were considered to represent

Author contributions: B.H.L., A.Q.B., G.N.D., R.L.F., J.K.O., J.J.H., and R.B.G. designed research; B.H.L., C.C., A.Q.B., J.A.W., K.F.H., J.P., L.K., M.B., and R.K. performed research; S.H. contributed new reagents/analytic tools; B.H.L., C.C., and R.B.G. analyzed data; and B.H.L. and R.B.G. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

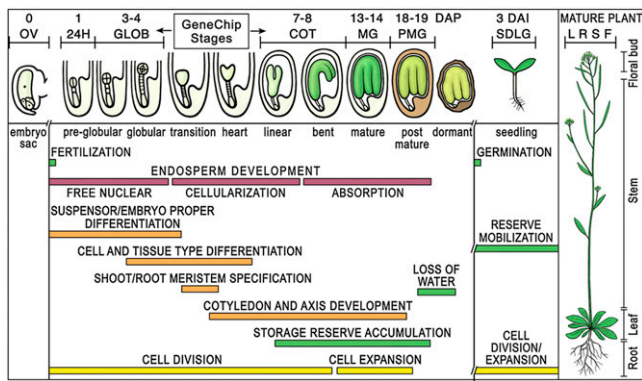
Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (accession no. GSE680).

<sup>1</sup>B.H.L., C.C., and A.Q.B. contributed equally to this work.

<sup>2</sup>Present address: Bayer CropSciences, Buenos Aires, Argentina, 1605.

<sup>3</sup>To whom correspondence should be addressed. E-mail: bobg@ucla.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/1003530107/DCSupplemental](http://www.pnas.org/cgi/content/full/1003530107/DCSupplemental).



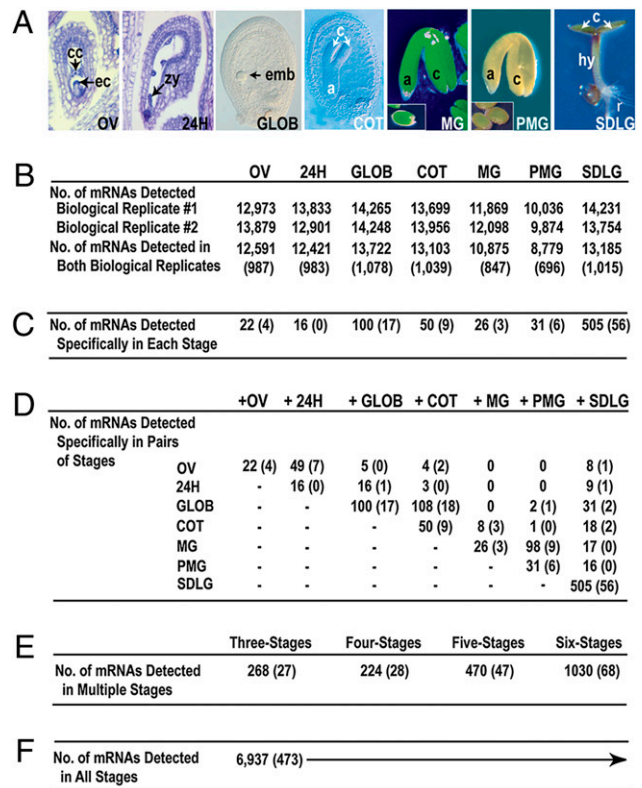
**Fig. 1.** Schematic representation of *Arabidopsis* seed development and stages of the life cycle used for GeneChip analysis. Seed cartoons were adapted from Bowman and Mansfield (57) and are not drawn to scale. Developmental events were modified from Goldberg et al. (1). Stages used for GeneChip analysis are described in *SI Materials and Methods*. Numbers correspond to days after pollination (DAP) or days after imbibition (DAI). Brackets mark the range of embryo stages included in each GeneChip seed sample. OV, unfertilized ovule; 24H, 24-h postpollination seed; GLOB, globular-stage seed; COT, cotyledon-stage seed; MG, mature-green-stage seed; PMG, postmature-green-stage seed; SDLG, seedling; L, leaf; R, root; S, stem; F, floral buds.

mRNAs present in any given developmental stage. Probe sets with discordant MAS 5.0 detection calls between biological replicates [e.g., P and A (Absent), or PA] were assigned a consensus detection call of Insufficient (INS) and removed from further analysis across all datasets used for comparative analysis (*Materials and Methods*). Thus, the seed stage-specific and seed-specific mRNA sets presented in this paper represent the minimum number of genes that are active at specific periods of seed development.

**mRNAs Present Before, During, and After Seed Development.** We detected  $\approx 9,000$ –14,000 unique mRNAs present at different developmental stages (Fig. 2B). Pearson correlation coefficients between biological replicates ranged from 0.96 to 0.99 (Fig. S1), indicating excellent concordance with each other. The number of mRNAs did not vary significantly in the periods before and after fertilization (OV and 24H), including the early stages of seed development (GLOB and COT) [ $P > 0.30$ , Analysis of Variance (ANOVA)]. By contrast, the number of mRNAs detected in MG and PMG seed stages decreased significantly ( $P < 0.001$ , ANOVA), and the values we obtained agreed with those in the late seed development AtGenExpress dataset (25). After germination, there was a significant increase in gene activity compared with seed MG and PMG stages ( $P < 0.001$ , ANOVA) (Fig. 2B). Collectively, we detected 15,563 diverse mRNAs throughout seed development (24H to PMG), and 16,701 mRNAs before, during, and after seed formation (OV to SDLG).

The average Pearson correlation coefficients between GeneChip experiments using mRNAs from different stages ranged from 0.97 for OV and 24H samples to 0.20 for PMG and SDLG samples. In general, Pearson correlation coefficients decreased as the seed stage pairs became more distant to each other developmentally. For example, the average correlation coefficients between GLOB and COT, GLOB and MG, and GLOB and PMG mRNAs were 0.87, 0.41, and 0.24, respectively. We grouped features on the *Arabidopsis* ATH1 GeneChip array into functional categories (34) (Fig. S24) and determined that specific functional groups were enriched or reduced in the mRNA populations at each developmental period (Table S1 in Dataset S1).

Taken together, these data show that the number of diverse seed mRNAs decreased significantly when the seed began preparing for dormancy, major changes in gene activity occurred



**Fig. 2.** Genes active before, during, and after *Arabidopsis* seed development. (A) Bright-field (OV, 24H), Nomarski (GLOB, COT), and whole-mount (MG, PMG, SDLG) photographs of pre-fertilization ovule, seed stages, and post-germination seedling used for GeneChip analysis, respectively. OV and 24H seed samples were visualized from 10  $\mu$ m stained paraffin sections (58). *Insets* show seeds used to dissect whole-mount MG and PMG embryos. Embryo in COT seed is at the linear cotyledon (LCOT) stage (Fig. 1). (B) Number of mRNAs detected at each stage of development. Numbers for biological replicates 1 and 2 indicate the number of probe sets with a MAS 5.0 detection call of P in each experiment (*Materials and Methods*). The number for both biological replicates indicates a consensus probe set detection call of PP and was used for subsequent analysis (*Materials and Methods*). Scatter plots and correlation coefficients comparing biological replicates are presented in Fig. S1. (C–F) Minimum number of specific and shared mRNAs at each developmental stage. The stringent filtering process used for these analyses is outlined in *Materials and Methods*. A total of 8,510 probe sets with INS (e.g., PA) or marginal (MM) consensus calls between biological replicates listed in B were removed across all developmental stages (*SI Materials and Methods*). The remaining probe sets were used to determine the number of stage-specific mRNAs (C) and mRNAs shared by two stages (D), three to six stages (E), or all stages (F). mRNAs (6,178 of the 6,937) shared by all stages (F) varied quantitatively across development ( $P < 0.05$ , ANOVA). Number in parentheses indicates TF mRNAs. The identities of mRNAs in each category (e.g., seed-stage-specific) are listed in Tables S3, S4, and S7 of Dataset S1. a, axis; c, cotyledon; cc, central cell; ec, egg cell; emb, embryo; hy, hypocotyl; r, roots; zy, zygote.

across seed development, and at least 16,000 genes are active throughout seed development.

**TF mRNAs Present Before, During, and After Seed Development.** We detected  $\approx 700$ –1,000 diverse TF mRNAs in each developmental stage, representing 36–55% of TF mRNAs represented on the *Arabidopsis* ATH1 GeneChip (Fig. 2B). Fewer TF mRNAs were detected in the MG and PMG stages compared with early seed stages (24H to COT) and postgermination SDLG, reflecting the decrease in mRNAs as a whole late in seed development (Fig. 2B). The proportion of TF transcripts relative to total mRNAs within a population, however, was the same for all stages (i.e.,  $\approx 8\%$ ). Collectively, we detected 1,327 diverse TF mRNAs throughout seed development



(24H to PMG) and 1,455 TF mRNAs before, during, and after seed formation (OV to SDLG).

We annotated features on the *Arabidopsis* ATH1 GeneChip array corresponding to major TF families to determine the spectrum of TF mRNAs present before, during, and after seed development (Fig. S2B). All major TF families were represented in the mRNA population at each developmental stage. However, significant differences were observed in the representation of specific TF families (Table S2 in Dataset S1). Taken together, these data suggest that at least 1,300 diverse TF mRNAs are required to program all of seed development, the number of TF mRNAs decreases before dormancy, and that the representation of specific TF mRNA families differs at specific developmental periods.

#### Each Seed Developmental Stage Has a Small Set of Specific mRNAs.

We identified a small number of mRNAs specific to each stage at the level of the GeneChip, including those encoding TFs from a variety of different families (Fig. 2C and Table S3 in Dataset S1). The stage-specific mRNAs included a range of functional categories, although the majority encoded proteins that were either unclassified or had no known function (Table S3 in Dataset S1). Approximately half of the seed-stage-specific mRNAs were also found to be seed-specific (see Fig. 4 below), suggesting that they play important roles in seed development. The largest numbers of seed-stage-specific mRNAs were observed in the GLOB and COT stages when major differentiation and morphological events occur during seed development (Fig. 1). For example, 100 GLOB-specific and 50 COT-specific mRNAs were identified, including 17 and 9 stage-specific TF mRNAs, respectively (Fig. 2C and Table S3 in Dataset S1). The GLOB-specific TF mRNAs included those encoding AUXIN RESPONSE FACTOR21 (ARF21, AT1G34410), LATERAL ORGAN BOUNDARIES35 (LBD35, AT5G35900), LBD15 (AT2G40470), and MINISEED3 (MINI3, AT1G55600)—the latter playing a major role in seed size (35). By contrast, we identified <75 mRNAs that were specific for the 24H, MG, and PMG stages, including 9 TF mRNAs (Fig. 2C). After germination, >500 SDLG-specific mRNAs were observed, including 56 SDLG-specific TF mRNAs (Fig. 2C and Table S3 in Dataset S1).

We compared the mRNA sets represented in pairs of developmental stages to determine whether there were seed-period-specific mRNAs in addition to those unique to individual stages (Fig. 2D). We observed that pairs of seed stages that were close to each other developmentally (e.g., OV-24H, GLOB-COT) had small sets of mRNAs that were not detected at the level of the GeneChip at other stages of development (Fig. 2D and Table S4 in Dataset S1). For example, OV and 24H seeds had 49 specific mRNAs that were absent in other stages as well as in post-germination SDLG. Similarly, GLOB and COT seeds had 108 mRNAs, including 18 encoding TFs, that were not detected at any other stage investigated, whereas MG and PMG seeds had 98 mRNAs, including 9 TFs, that were absent at other developmental periods. Neither 24H and MG nor 24H and PMG seeds, by contrast, had any detectable pair-specific mRNAs (Fig. 2D). Analysis of Gene Ontology (GO) terms enriched in both the seed-stage-specific and seed-period-specific mRNA sets (SI Materials and Methods) indicated that the GLOB and COT mRNA sets were enriched in sequences encoding gene regulatory functions (i.e., TFs) ( $P < 0.01$ ), whereas the MG and PMG mRNA sets were enriched in sequences encoding seed dormancy and embryo developmental functions [e.g., late embryo abundant (lea) proteins] ( $P < 0.01$ ) (Table S5 in Dataset S1).

We used the Harada-Goldberg LCM GeneChip dataset (Materials and Methods) to determine where the seed-stage-specific and seed-period-specific mRNAs were localized within the seed (Table S6 in Dataset S1). Most of the GLOB- and GLOB-COT-specific mRNAs were localized either within the endosperm or the seed coat. Very few were present exclusively within the embryo, although many of the endosperm mRNAs were also colocalized within the suspensor. Most

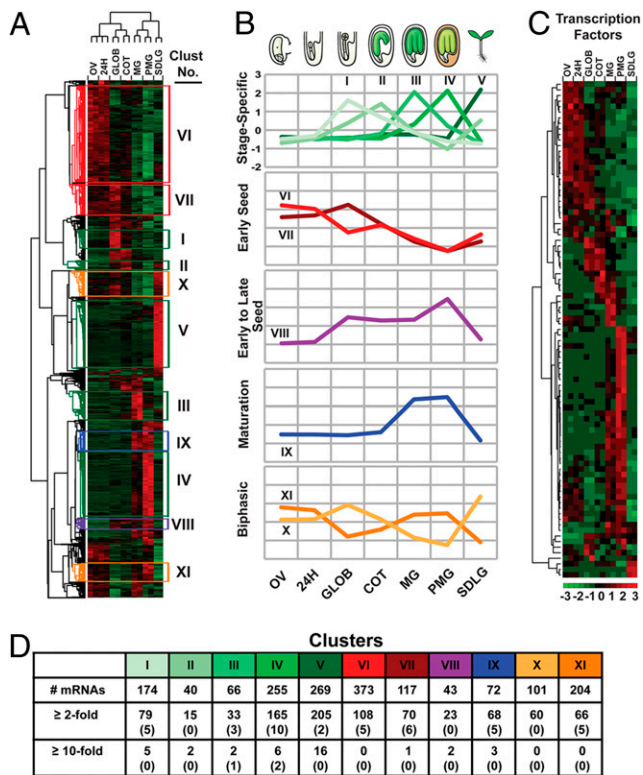
embryo mRNAs were probably below the detection limit of our GeneChip experiments (see below) because of their small contribution to the entire seed mRNA population at these developmental stages (Figs. 1 and 2A). By contrast, the MG- and MG-PMG-specific mRNAs were localized primarily within the embryo or seed coat at these stages of seed development when the embryo is fully differentiated and occupies the majority of space within the seed (Figs. 1 and 2A).

We carried out quantitative reverse transcription PCR (qRT-PCR) experiments with 19 seed-stage-specific mRNAs (Fig. 2C) and seed-period-specific mRNAs (Fig. 2D), including 13 TF mRNAs, to determine whether they were present at other stages, but below the detection limit of the GeneChip (Table S10 in Dataset S1). In our experiments, the detection level was  $\approx 2 \times 10^{-5}$ , or one transcript per 200,000 (Fig. S3A), which is similar to that determined by others using Affymetrix GeneChips (36, 37). All tested mRNAs were validated by qRT-PCR in their target stages at levels similar to those observed with the GeneChip (Table S10 in Dataset S1 and Fig. S3B; Pearson correlation coefficient = 0.76). Most mRNAs tested ( $\approx 70\%$ ), however, were also detected in one or more other stages at greatly reduced levels ( $\approx 10$ – $10,000$  fold) (Table S10 in Dataset S1). The small number of stage-specific mRNAs determined by qRT-PCR to be present in other stages at levels close to that of the target stage, or with slightly reduced levels (2- to 3-fold reduction), had GeneChip signal intensities bordering on the limits of detection (Table S10 in Dataset S1 and Fig. S3C). Taken together, these results indicate that each stage and period of seed development has a small set of mRNAs, including those encoding TFs, that is either absent from other stages or present at highly reduced levels, and that many of these mRNAs are localized in specific parts of the seed.

#### Most Diverse Seed mRNAs Are Present Before, During, and After Seed Development.

By contrast with the small number of seed-stage- and seed-period-specific mRNAs (Fig. 2C and D), most seed mRNAs ( $\approx 7,000$ ), including those encoding TFs ( $\approx 500$ ), were detected before (OV), during (24H to PMG), and after seed formation (SDLG) (Fig. 2F and Table S7 in Dataset S1), indicating that most diverse seed mRNA sequences are present from fertilization through dormancy. A large number of mRNAs were also observed in mosaic combinations of three to six stages (e.g., OV, 24H, and PMG; COT, PMG, and SDLG), although the total number ( $\approx 2,000$ ) was significantly less than those mRNAs shared across development (Fig. 2E and Table S4 in Dataset S1). Most of the shared mRNAs (4,237 of 6,937 or  $\approx 61\%$ ), including those encoding TFs (301 of 473 or  $\approx 64\%$ ), changed significantly in prevalence by at least 2-fold during at least one period of development (e.g., OV-24H) ( $P < 0.01$ ,  $t$  test).

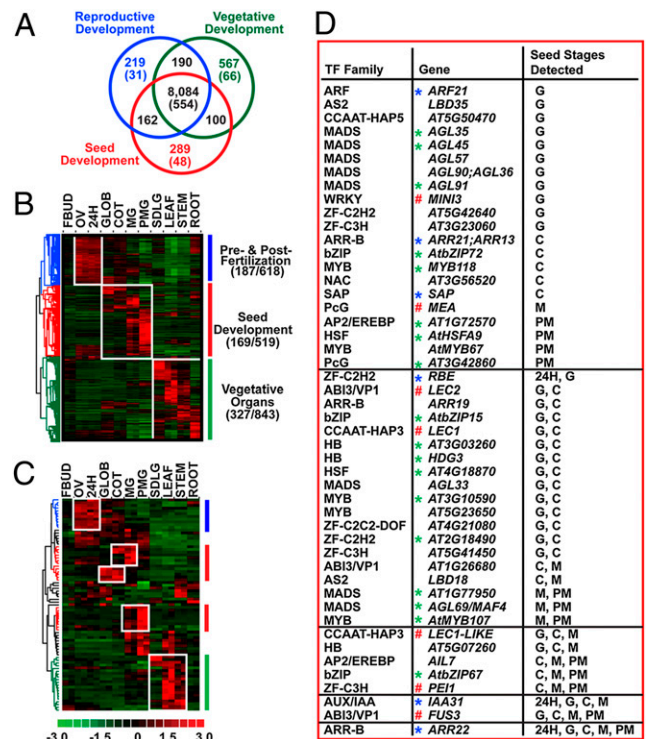
We carried out unsupervised hierarchical clustering analysis on a subset of 2,000 shared mRNAs that had the largest standard deviation in signal levels across all developmental stages (i.e., varied the most quantitatively) to identify up-regulated sets of mRNAs (Fig. 3A and Table S7 in Dataset S1). The different mRNA samples (e.g., OV to SDLG) clustered according to their temporal relationships during seed formation, reflecting the correlation coefficients between developmentally similar and dissimilar mRNA populations (Fig. 3A, top brackets). For example, the OV and 24H shared mRNAs were more similar in their quantitative profiles than the OV and SDLG shared mRNA sets. We identified 11 prominent mRNA clusters in the shared mRNA population (clusters I to XI), including those encoding TFs, demonstrating that complex patterns of quantitative mRNA changes occurred during seed development coinciding with unique developmental events (Fig. 3A–C and Table S7 in Dataset S1). Shared mRNAs grouped into seed-stage-up-regulated clusters (GLOB, COT, MG, and PMG) (I–IV) and seed-period-up-regulated clusters for (i) early seed development (OV to GLOB) (VI and VII), (ii) early to late seed development (GLOB to PMG) (VIII), (iii) late seed development (MG and PMG) (IX), and (iv) postgermination SDLG (V) (Fig. 3A and B). In addition, two biphasic mRNA clusters (X and XI) were identified that were up-regulated in temporally



**Fig. 3.** Quantitative regulation of mRNAs shared by all stages of seed development. (A) Unsupervised hierarchical clustering of GeneChip samples (Fig. 1) and probe sets with a consensus detection call of PP in all stages of development (Fig. 2F) was carried out by using dChip 1.3 (56) as described in *Materials and Methods*. Only the top 2,000 probe sets with the most varying signals across all stages were included in the clustering analysis (*SI Materials and Methods*). Numbered boxes highlight individual clusters of coregulated mRNAs. The identities of the top 2,000 mRNAs used for this clustering analysis and the mRNAs in each cluster are listed in [Table S7 of Dataset S1](#). GO terms that are enriched significantly in each cluster ( $P < 0.01$ ) are presented in [Tables S8 and S9 of Dataset S1](#). (B) Graphical representation of cluster mRNA accumulation patterns. Lines represent the average mRNA accumulation pattern for all mRNAs in each cluster. (C) Unsupervised hierarchical clustering of 89 TF mRNAs included in the top 2,000 most varying probe sets shared by all stages of seed development (Fig. 2F) presented in A. The identities of TF mRNAs in each cluster are listed in [Tables S7 and S9 of Dataset S1](#). (D) The number of mRNAs in each cluster shown in B and the number of mRNAs per cluster that increased significantly in prevalence  $\geq 2$ -fold and  $\geq 10$ -fold relative to the mean signal intensity of each cluster ( $P < 0.05$ ). Scale from  $-3$  (green) to  $+3$  (red) represents the relative number of standard deviations from the mean signal intensity for each probe set across all developmental stages.

distinct developmental periods. Cluster X mRNAs were up-regulated in GLOB seeds and postgermination SDLG, whereas cluster XI contained mRNAs that were highly prevalent in OV-24H and MG-PMG seeds (Fig. 3A and B). Approximately 30% (cluster XI) to 94% (cluster IX) of the mRNAs in each cluster increased in prevalence by at least 2-fold, with a smaller proportion of mRNAs (1–5%) in most clusters increasing  $>10$ -fold ( $P < 0.05$ ,  $t$  test) (Fig. 3D). The highest mRNA abundance change for the shared mRNAs was  $\approx 35$ -fold and occurred during maturation (MG and PMG, cluster IX) and after seed germination (SDLG, cluster V) ([Table S9 in Dataset S1](#)).

GO term analysis of cluster mRNAs showed enrichment for sequences programming distinct processes at specific stages of seed development ([Tables S8 and S9 in Dataset S1](#)). For example, the mRNAs present in cluster I (GLOB), cluster II (COT), and cluster III (MG) were enriched for sequences involved in carbohydrate metabolic processes ( $P < 7.7 \times 10^{-6}$ ), rhamnose biosynthesis ( $P <$



**Fig. 4.** Identification of *Arabidopsis* seed-specific mRNAs. (A) GeneChip data obtained for all stages of the life cycle (Fig. 2 and Fig. S4) were partitioned into three groups: (i) reproductive development [OV and floral buds (FBUD) (blue circle)], (ii) seed development [(24H, GLOB, COT, MG, and PMG) (red circle)], and (iii) vegetative development [SDLG, leaf, stem, and root (green circle)]. Processing and filtering of the data are outlined in *SI Materials and Methods*. Number in parentheses indicates number of TF mRNAs. The identities of seed-specific, seed-specific TF, reproductive-organ-specific, and vegetative-organ-specific mRNAs are listed in [Table S13 of Dataset S2](#). GO terms that are enriched significantly in the seed-, reproductive-, and vegetative-specific mRNA sets ( $P < 0.01$ ) are presented in [Table S15 of Dataset S2](#). Seed development accumulation patterns and representation of functional groups for the 289 seed-specific mRNAs are shown in Fig. S5. (B and C) Unsupervised hierarchical clustering of mRNAs (B) and TF mRNAs (C) shared by all periods of the life cycle [i.e., intersection of mRNA sets in A] was carried out by using dChip 1.3 (56) as described in *SI Materials and Methods* and Fig. 3A legend. Only the top 2,000 probe sets with the most varying signals across all periods of the life cycle were included in the clustering analysis shown in B (*SI Materials and Methods*). All 77 TF mRNAs included in the top 2,000 most varying probe sets shared by all life cycle periods (B) were used for the clustering analysis shown in C. Blue, red, and green bars highlight mRNA clusters that are up-regulated in (i) OV and 24H seeds, (ii) GLOB, COT, MG, and PMG seeds, and (iii) SDLG and vegetative organs, respectively. The number of mRNAs in each cluster that increased significantly in prevalence  $\geq 2$ -fold relative to the mean signal intensity of each cluster ( $P < 0.05$ ) is listed next to the bars in B. The identities of mRNAs shared throughout the life cycle (A) and those that are present in up-regulated clusters (B and C) are listed in [Table S17 of Dataset S2](#). GO terms that are enriched significantly in each cluster ( $P < 0.01$ ) are presented in [Table S18 of Dataset S2](#). (D) TF families and stage specificity of seed-specific TF mRNAs identified in A. Seed-specific TF mRNAs were classified into families as shown in Fig. S5. Homozygous T-DNA insertion lines for TF genes marked with a \* did not produce a detectable seed phenotype (green \* by us and blue \* by others) ([Table S20 in Dataset S2](#)). Mutations in seed-specific TF genes marked with a # were shown previously by us (e.g., *lec1*, *lec1-like*, *lec2*, *mea*) and by others (e.g., *pei1*, *fus3*), to produce a seed-defective phenotype ([Table S20 in Dataset S2](#)). Scale from  $-3$  (green) to  $+3$  (red) represents the relative number of standard deviations from the mean signal intensity for each probe set across all developmental stages.

$9.6 \times 10^{-4}$ ), and fatty acid metabolism ( $P < 5 \times 10^{-4}$ ), respectively. These GO terms most likely reflect key biological events that occur in seeds during these time periods; for example, starch accumulation



in the outer integument seed coat layer (GLOB), mucilage formation in the seed coat (COT), and fatty acid synthesis in the embryos during maturation (MG) (5, 38). Each cluster contained TF mRNAs (Fig. 3 C and D) that were up-regulated >2-fold in prevalence and may be important for regulating the GO-term biological processes that occur during the corresponding developmental period (Fig. 3C and Tables S8 and S9 in Dataset S1). For example, AtMyb61 mRNA (AT1G09540), ATAF1 mRNA (AT1G01720), ATbZIP53 mRNA (AT3G62420), and SPLAYED (SYD) mRNA (AT2G28290) were prevalent in cluster I (GLOB), cluster IV (PMG), cluster IX (MG and PMG), and cluster XI (OV-24H and MG-PMG), respectively. These TF mRNAs have been shown to regulate seed coat mucilage extrusion (AtMyb61) (39), ABA response (ATAF1) (40), maturation gene expression (ATbZIP53) (41), and cotyledon boundary-shoot meristem formation (SYD) (42), respectively (Table S9 in Dataset S1).

Taken together, these data show that (i) most diverse seed mRNAs, including TF mRNAs, are present before, during, and after seed formation, (ii) shared seed mRNAs undergo significant prevalence changes and are grouped into stage- and period-specific clusters of coup-regulated mRNA sets, and (iii) coup-regulated mRNAs within each cluster encode proteins involved in important seed processes.

**Most Seed mRNA Sequences Are Present Throughout the Plant Life Cycle.** We carried out GeneChip hybridization experiments by using mRNAs isolated from floral buds (FBUD), leaves (L), stems (S), and roots (R) (Fig. S4) and compared the combined reproductive (FBUD, OV) and vegetative (L, R, S, SDLG) mRNA populations with those present throughout seed development (24H, GLOB, COT, MG, and PMG) to determine the representation of seed mRNAs at other times of the life cycle (Fig. 4A and Table S13 in Dataset S2). Most diverse seed mRNA sequences were represented in the reproductive and vegetative organs of the mature plant (Fig. 4A), in addition to being present before (OV) and after seed germination (SDLG) (Fig. 2). We identified a minimum of 8,084 diverse mRNAs, including 562 TF mRNAs, that were shared by the seed, floral, and vegetative mRNA populations (Fig. 4A) by using our stringent filtering criterion (*Materials and Methods*). A smaller number of life cycle mosaic mRNAs (100–200) were shared by different combinations of two of the three developmental periods (e.g., seed and vegetative development; Fig. 4A and Table S16 in Dataset S2). Unsupervised hierarchical clustering analysis of the top 2,000 most varying life cycle-shared mRNAs (*Materials and Methods*) indicated there were clusters of up-regulated mRNA sets, including those encoding TFs, specific for each period of the life cycle, including seed development (Fig. 4B and C and Table S17 in Dataset S2). Clusters of up-regulated GLOB, COT, MG, and PMG seed mRNAs were identified in the life cycle-shared mRNA population, including those encoding TFs (Fig. 4B and C), that contained sequences also present in up-regulated shared seed mRNA clusters at the same developmental stages (Fig. 3). For example, ~80% of the mRNAs that were present within the PMG cluster of the shared seed mRNA population (cluster IV, Fig. 3A and B) were also found within the PMG seed cluster of the life cycle shared mRNA population (Fig. 4B). GO term analysis of the up-regulated seed mRNAs (Fig. 4B) showed enrichment for sequences involved in processes that occur during seed development and maturation (e.g., water deprivation, fatty acid biosynthesis; Table S18 in Dataset S2). Thus, many stage- and period-specific seed mRNAs were up-regulated in the context of both seed development and the entire plant life cycle.

Collectively, we detected a total of 18,504 diverse mRNAs, including 1,675 TF mRNAs, during seed, reproductive, and vegetative periods of the life cycle at the level of the GeneChip, a value only 20% higher than the 15,563 diverse mRNAs present during seed development. Taken together, these data indicate that (i) there is a large overlap in the mRNA populations of developing seeds, from fertilization through dormancy, with those present in floral and veg-

etative organs of the mature plant, (ii) there is a small set of life cycle-shared mRNAs that is up-regulated during specific periods of seed development, and (iii) at least 18,504 diverse mRNAs are required to program the sporophytic phase of the plant life cycle, a value consistent with the 20,000 diverse mRNAs present in the AtGenExpress GeneChip database (25).

**Seeds Contain a Small Set of Specific mRNAs Enriched for Sequences Encoding Seed-Specific TFs.** By contrast with the large overlap between the mRNA populations present throughout the life cycle, we identified a small set of mRNAs that was specific at the level of the GeneChip for each developmental period, including seeds from fertilization through dormancy (Fig. 4A and Table S13 in Dataset S2). For example, we identified 289 mRNAs that were seed-specific and not detected at other times of development, including 48 seed-specific TF mRNAs (Fig. 4A and D and Table S13 in Dataset S2). qRT-PCR experiments with 36 of the seed-specific TF mRNAs indicated that they were either absent from mature plant organs (L, R, S, FBUD) or represented in one or more organs at highly reduced levels. In general, the seed-specific TF mRNAs were reduced in prevalence by 100- to 60,000-fold when present in floral and vegetative mRNA populations (Table S11 in Dataset S2). For example, five seed-specific TF mRNAs [AtbZIP72 (AT5G07160), heat shock protein AT-HSFA9 (AT5G54070), AGL33 (AT2G26320), myb-related protein (AT5G23650), and Homeobox (HB) protein (AT5G07260)] were undetectable at both the GeneChip and qRT-PCR levels in mature plant organs. AT-HSFA9 mRNA had been shown previously to be seed-specific (43). By contrast, LEC1 and LEC2 mRNAs were absent from L, R, and S at the qRT-PCR level but present at a 600-fold (LEC2 mRNA) to 3,000-fold (LEC1 mRNA) reduced levels in FBUD (Table S11 in Dataset S2). Comparison of our seed-specific TF mRNA set against relevant *Arabidopsis* gene expression datasets (e.g., AtGenExpress) validated that the seed-specific TF mRNAs uncovered here were either not detectable at other periods of the life cycle or present in one or more mature plant organ systems at reduced levels.

The seed-specific mRNAs were distributed into all major functional groups, although 25% encoded proteins that were either unclassified or had no known function (Fig. S5A and Table S14 in Dataset S2). Remarkably, the largest known functional group represented in the seed-specific mRNA population was transcription (18%), reflecting a significant enrichment (48/289) in TF mRNAs ( $P < 1 \times 10^{-4}$ , Fisher Exact Test) (Table S14 in Dataset S2), including 17 of the 23 major TF families represented on the GeneChip (compare Figs. S2B and S5B). Three TF families, ARR-B, MADS, and CCAAT, were overrepresented significantly in the seed-specific TF mRNA population ( $P < 0.02$ , Fisher Exact test) (Table S14 in Dataset S2), whereas others (e.g., Trihelix, GRAS, bHLH) had no representatives (Fig. S5B). Analysis of GO term enrichment categories encoded by the seed-specific mRNA set also showed an enrichment for sequences involved in transcriptional regulation ( $P < 0.01$ ), seed development ( $P < 0.001$ ), somatic embryogenesis ( $P < 0.01$ ), and oil storage ( $P < 0.001$ ) (Table S15 in Dataset S2). As expected, mRNAs encoding known seed-specific protein markers, such as storage proteins, lea proteins, and oleosin, were also enriched in the seed-specific mRNA population (Table S13 in Dataset S2), reflecting the overrepresentation of sequences in the protein destination and storage category ( $P < 4 \times 10^{-5}$ , Fisher Exact Test) (Fig. S5A and Table S14 in Dataset S2).

Approximately 80% of the seed-specific mRNAs, including those encoding TFs, were either stage specific (e.g., 24H, GLOB, COT, MG, or PMG) or period specific for two contiguous stages (e.g., 24H-GLOB, GLOB-COT, MG-PMG) during seed development (Fig. 4D and Fig. S5C). Most were either GLOB- or COT-stage-specific, or were present specifically during the GLOB-COT period of development (Fig. 4D and Fig. S5C), mirroring what was observed with the seed-stage-specific mRNA set that included half of the seed-specific mRNAs (Fig. 2B and C). Together, these data

indicate that there is a small set of seed-specific mRNAs, enriched for sequences encoding TFs, that is either absent from mature plant organs or present at highly reduced levels, and that most of these mRNAs accumulate at specific times of seed development when major events required for seed formation occur (Fig. 1).

### Seed-Specific TF mRNAs Are Localized in Different Seed Compartments.

We used chimeric seed-specific promoter-*GUS* transgenes to localize seed-specific TF gene transcriptional activity within seed compartments (i.e., embryo, endosperm, seed coat) from fertilization through dormancy (Fig. 5*A* and Fig. S64). Several unique transcriptional patterns were observed, including transcription in the (i) entire embryo, (ii) embryonic organs including cotyledons and axis, (iii) endosperm, and (iv) chalazal endosperm (Fig. 5*A*). The latter transcriptional pattern was observed for nine different seed-specific promoter-*GUS* reporter genes (Fig. S64). In general, the embryo transcriptional patterns correlated with the accumulation of corresponding seed-specific TF mRNAs late in development (COT-PMG), whereas the endosperm/seed coat transcriptional patterns correlated with seed-specific TF mRNAs that accumulated early in development (GLOB-COT) (Fig. 5*A* and Fig. S7).

We used the Harada-Goldberg LCM GeneChip dataset to localize the seed-specific TF mRNAs within all major seed compartments and tissues during development (*Materials and Methods*), including the

(i) embryo (embryo proper and suspensor), (ii) endosperm (peripheral, micropylar, and chalazal), and (iii) seed coat (general and chalazal) (Fig. 5*B* and *C*, Figs. S6B and S7, and Table S19 in Dataset S2). Each seed-specific TF mRNA had a unique temporal- and compartment-specific accumulation pattern (Fig. 5*B* and *C*, and Fig. S7). For example, *AtbZIP67* mRNA (AT3G44460) accumulated in the (i) embryo proper, (ii) peripheral, micropylar, and chalazal endosperm regions, and (iii) general and chalazal seed coat during the heart (HRT) to MG stages (Fig. 5*B* and *C*). By contrast, *myb* TF mRNA AT3G10590 was detected primarily in the chalazal endosperm and seed coat regions during the preglobular (PG) to MG period of development (Fig. 5*B* and *C*). In general, both the whole seed mRNA accumulation patterns and the *GUS* transgene localization profiles were congruent with both the temporal and spatial TF mRNA accumulation patterns identified by using the Harada-Goldberg LCM GeneChip dataset (Fig. 5*A* and *C* and Fig. S6). Together, these data indicate that the seed-specific TF mRNAs are localized within unique seed compartments at precise times during seed development and that transcriptional processes are primarily responsible for generating the seed-specific TF mRNA accumulation patterns.

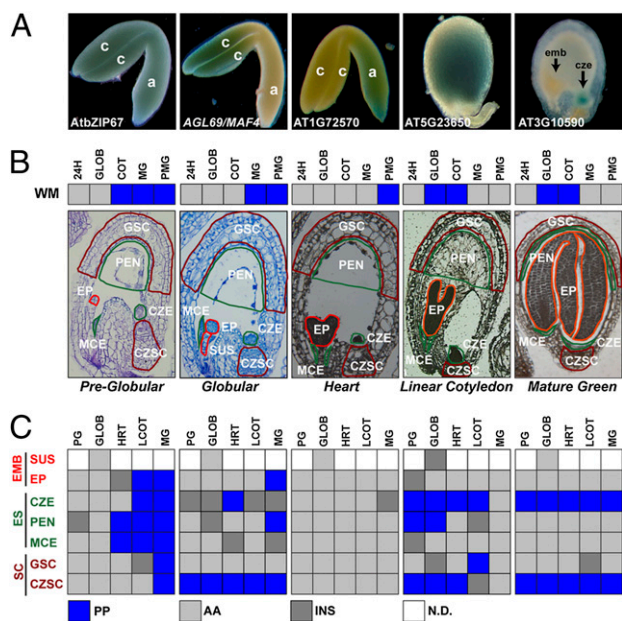
### Mutations in Genes Encoding Most Seed-Specific TFs Do Not Result in a Detectable Phenotype.

Seven of the 48 TF mRNAs uncovered in our seed-specific TF mRNA set encoded important regulatory proteins that resulted in seed-lethal phenotypes when their corresponding genes were mutated, including *LEC1*, *LEC2*, *FUS3*, *PEI1*, *MINISEED3*, and *MEDEA* (Fig. 4*D* and Table S20 in Dataset S2) (35, 44–46). Mutations (47) in an additional 24 seed-specific TF genes did not result in a seed-lethal phenotype or detectably alter seed development (Fig. 4*D* and Table S20 in Dataset S2). Taken together, these data indicate that the seed-specific TF gene set is enriched for important known regulators of seed development, but that mutant alleles of most seed-specific TF genes do not yield a seed phenotype and, as such, the functions of most seed-specific TF genes investigated here are not yet known.

### Discussion

We profiled *Arabidopsis* mRNA sets before, during, and after seed formation, and compared these mRNA sets to those from mature plant organ systems to uncover key transcriptional regulators of seed development. Our experiments showed that at least 16,000 mRNAs are required to program *Arabidopsis* seed development—from fertilization through dormancy. This is undoubtedly a lower limit due to (i) our use of whole seed mRNAs, (ii) the inability of the GeneChip to detect rare sequences in complex mRNA populations such as whole seeds, and (iii) the incompleteness of the Affymetrix ATH1 GeneChip, which is missing  $\approx 20\%$  of known *Arabidopsis* genes (36). If we assume that sequences not included on the GeneChip represent a random collection of genes, then at least 19,000 diverse mRNAs are required to program seed development and make an *Arabidopsis* seed.

Most diverse seed mRNA sequences are present before fertilization, persist from zygote formation through dormancy, and are represented after seed formation in germinating seedlings and mature plant organ systems. Thus, most seed mRNAs are used in different developmental contexts throughout the plant life cycle, although significant quantitative changes occur in individual mRNA prevalences that correspond with specific seed developmental stages and/or periods of the life cycle. The reduction in the number of seed mRNAs detected during late development (MG-PMG) is probably due to mRNA turnover resulting from the general shutdown in transcriptional processes as the seed enters dormancy (48). More than a generation ago, we (49) and others (50), used RNA/cDNA hybridization experiments to investigate mRNA populations during seed development. Our conclusions from that era, using primitive technology that provided the foundation for questions being addressed currently with sophisticated genomics approaches, are in remarkable agreement with those reported here. That is, most diverse seed mRNA sequences



**Fig. 5.** Seed-specific TF gene activity in different *Arabidopsis* seed compartments, regions, and tissues. (A) Localization of GUS enzyme activity in seeds and embryos of transgenic lines carrying different seed-specific TF gene upstream regions (Fig. 4*D*) fused with the *GUS* reporter gene (SI *Materials and Methods*). Squares in the horizontal bars below each GUS-stained embryo or seed show the GeneChip MAS 5.0 consensus call for the seed-specific TF gene in whole-mount seeds at different developmental stages (Fig. 2). Blue and gray squares represent consensus detection calls of PP and AA, respectively (see *Materials and Methods*). (B) Bright-field photographs of *Arabidopsis* 5–7 μm paraffin seed sections at different developmental stages. Highlighted areas represent compartments, regions, and tissues captured by LCM (32). (C) Seed-specific TF mRNA localization within seeds at different stages of development. Blue, light gray, and dark gray squares indicate GeneChip MAS 5.0 consensus detection calls of PP, AA, and INS, respectively (see *Materials and Methods*). White squares indicate not determined (N.D.). a, axis; c, cotyledon; cze, chalazal endosperm; czsc, chalazal seed coat; emb, embryo; ep, embryo proper; es, endosperm; gsc, general seed coat; mce, micropylar endosperm; pen, peripheral endosperm; sc, seed coat; sus, suspensor; PG, preglobular stage seed; HRT, heart-stage seed; LCOT, linear cotyledon-stage seed.



persist throughout development, are represented in mature plant organs, and stage-specific quantitative changes occur in specific seed mRNA sets (49, 50).

We identified a small set of mRNAs, significantly enriched for sequences encoding TFs, which is either specific for seeds at the qRT-PCR level or present in one or more mature plant organs at levels significantly below those of seed mRNAs that are shared with other periods of the life cycle. These seed-specific mRNAs represent <2% of the total mRNAs present during seed development (289/15,500). At least half of the seed-specific mRNAs, including the seed-specific TF mRNAs, accumulate at specific stages of seed development when major events required for seed formation occur (e.g., GLOB, COT, MG). The remaining seed-specific mRNAs accumulate within temporally contiguous periods that correspond with key seed developmental events as well (e.g., GLOB-COT). Most of the seed-stage- and seed-specific mRNAs identified here accumulate within the GLOB-COT period of seed development, correlating with the period when the majority of *Arabidopsis* embryo-defective mutants arrest in seed development (51)—a time when critical morphogenetic and biochemical events occur that are required for embryo and seed formation (2) (Fig. 1). Other mRNAs are specific for the MG-PMG period of seed development when maturation occurs. Almost all of the GLOB-COT-specific mRNAs are localized within the endosperm, whereas the MG-PMG mRNAs are represented within the embryo, although these mRNAs can also be present in other regions of the seed. The temporal and spatial mRNA accumulation patterns for the seed-specific mRNAs correlate with biological processes that occur uniquely within seeds during the plant life cycle; that is, the formation of a triploid endosperm and a maturation period when seeds accumulate high levels of food reserves, prepare for dehydration, and enter dormancy (2, 4).

We identified 48 seed-specific TF mRNAs that most likely play important roles in regulating seed development. This is a lower limit of the number of seed-specific regulators because of the stringent filtering process we used in comparing GeneChip datasets generated in this study (*Materials and Methods*). If we lower our stringency to include discordant MAS 5.0 consensus calls containing one P and consider TF genes not present on the GeneChip, then the number of seed-specific TF mRNAs could approach 100 or more, but is still a small proportion of the 1,400 diverse TF mRNAs that we detected throughout seed development. Seed TF mRNAs that are shared with other periods of the life cycle clearly play important roles in seed development; however, the seed-specific TF mRNAs uncovered here probably guide processes unique to seeds.

The functions of most seed-specific TF mRNAs uncovered here are not known; however, the seed-specific TF mRNA set is enriched for known regulators of seed development (e.g., LEC1, LEC2, L1L, FUS3, MEDEA) that were uncovered in genetic screens for embryo defective mutants, strongly suggesting that the other seed-specific TF mRNAs will play critical regulatory roles as well. These regulators have been shown to be critical for controlling events unique to seeds; that is endosperm formation and maturation (15, 52). The critical question is, of course, what role do the remaining seed-specific TF mRNAs in our dataset play in seed development? The localization patterns of most of these seed-specific TF mRNAs also suggest that they are involved in regulating either the differentiation and/or function of the endosperm early in seed development or events required for maturation in either specific embryo regions and/or the seed coat late in development. One clue as to the function of several seed-specific TF mRNAs is the observation that nine are localized in the chalazal endosperm layer during the GLOB-COT phase and correlate with the transcription of their corresponding genes (Fig. S6). This seed layer plays an important role in embryo development transferring critical nutrients from maternal to embryonic tissues (2), and it is possible that the nine chalazal-specific TF genes form a regulatory

network required for the differentiation and/or function of this seed region. Analysis of the chalazal endosperm-specific TF gene promoters shows an enrichment for a CARGCW8AT motif ( $P < 1 \times 10^{-4}$ ), which is an AGL15 (AT5G13790) TF binding site (Fig. S64) (53). AGL15 mRNA is present in the Harada-Goldberg LCM chalazal endosperm dataset, suggesting that AGL15 might act upstream of the chalazal endosperm-specific TF genes and play a role in activating at least one chalazal endosperm gene regulatory network.

The roles of most regulatory genes in controlling seed development and how seed gene sets are organized into regulatory networks are not well understood. The seed-specific TF genes uncovered in our study should provide an important starting point for understanding how gene activity is coordinated during seed development to make a seed. Clearly, how specific compartments of the seed are differentiated and the roles that seed-specific TF genes play in this process remain to be determined.

## Materials and Methods

**Plant Material.** Detailed information on (i) growth of *Arabidopsis* plants, (ii) stages of seed development, and (iii) characteristics of plant material are presented in *SI Materials and Methods*.

**RNA Isolation and Affymetrix GeneChip Hybridization.** Details of RNA isolation, biotinylated cRNA synthesis, and hybridization with Affymetrix *Arabidopsis* ATH1 22K GeneChips (36, 54) are presented in *SI Materials and Methods*. Two biological replicates were analyzed for each sample and processed at the same time to reduce variability. Signal intensities and detection calls [(P), (A), or (M)] were determined by using Affymetrix MAS 5.0 software default parameters (36, 54). All of the ATH1 22K GeneChip data were deposited in the Gene Expression Omnibus (GEO) as Series GSE680. Experiments were also carried out by using the first generation Affymetrix *Arabidopsis* AtGenome1 8K GeneChip (54, 55) with OV, 24H, COT, and MG seed RNAs. These data are not discussed here in detail but are also deposited in GEO as part of Series GSE680.

**Analysis of GeneChip Hybridization Data.** The consensus call for each probe set was assigned as PP, AA, or MM by combining the detection calls of both biological replicates (36, 54). Probe sets with different detection calls between biological replicates (e.g., P in replicate 1 and A in replicate 2) were assigned a consensus detection call of INS, or insufficient. We applied a stringent filter to our data by using only probe sets with a consensus call of PP (i.e., P in both biological replicates), and removing probe sets that had consensus calls of either INS or MM from all datasets used for comparative analysis. A detailed description of the process we used to analyze, filter, and compare the results of our GeneChip hybridization experiments, including hierarchical clustering using dChip 1.3 software (56) and GO term enrichment analysis, is presented in *SI Materials and Methods*.

**Real-Time Quantitative RT-PCR Validation of GeneChip Data.** qRT-PCR reactions were carried out by using the procedures described in *SI Materials and Methods*. Primer sequences used for the qRT-PCR reactions are listed in Table S12 of Dataset S2. qRT-PCR data were evaluated using three criteria: (i) observed vs. expected qRT-PCR product Tm, (ii) technical replicate Ct value reproducibility, and (iii) observed vs. expected qRT-PCR product size. Only qRT-PCR results that satisfied all three criteria were considered reliable and used in this paper.

**Localization of TF Gene Activity in Specific Seed Compartments.** TF promoter-*GUS* transgene experiments were carried out according to procedures outlined in *SI Materials and Methods*. Localization of mRNAs to specific seed compartments were determined by using the Harada-Goldberg LCM GeneChip datasets that are deposited in GEO as Series GSE12402, GSE11262, GSE15160, GSE12403, and GSE15165 for PG, GLOB, HRT, LCOT, and MG stages of seed development, respectively.

**ACKNOWLEDGMENTS.** We thank Weimin Deng for maintaining our *Arabidopsis* plants, Zixing Fang for assistance with the dChip analysis, and Zugen Chen for outstanding advice and help with the GeneChip hybridizations. This work was supported by grants from the National Science Foundation Plant Genome Program and Ceres (to R.B.G. and J.J.H.), US Public Health Service National Research Service Award GM07104 (K.F.H.), and National Institutes of Health Training Grant in Genomic Analysis and Interpretation T32HG002536 (B.H.L.).

1. Goldberg RB, de Paiva G, Yadegari R (1994) Plant embryogenesis: Zygote to seed. *Science* 266:605–614.
2. Raghavan V (2006) *Double Fertilization: Embryo and Endosperm Development in Flowering Plants* (Springer, Berlin).
3. Goldberg RB, Barker SJ, Perez-Grau L (1989) Regulation of gene expression during plant embryogenesis. *Cell* 56:149–160.
4. Harada JJ (1997) *Seed maturation and the control of germination. Cellular and Molecular Biology of Seed Development, Advances in Cellular and Molecular Biology of Plants*, eds Larkins BA, Vasil IK (Kluwer Academic Publishers, Dordrecht, the Netherlands), Vol 4, pp 545–592.
5. Baud S, Boutin J-P, Miquel M, Lepiniec L, Rochat C (2002) Anintegrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiol Biochem* 40: 151–160.
6. Olsen OA (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* 16(Suppl):S214–S227.
7. Laux T, Würschum T, Breuninger H (2004) Genetic regulation of embryonic pattern formation. *Plant Cell* 16(Suppl):S190–S202.
8. Haughn G, Chaudhury A (2005) Genetic analysis of seed coat development in *Arabidopsis*. *Trends Plant Sci* 10:472–477.
9. Jenik PD, Gillmor CS, Lukowitz W (2007) Embryonic patterning in *Arabidopsis thaliana*. *Annu Rev Cell Dev Biol* 23:207–236.
10. Devic M (2008) The importance of being essential: EMBRYO-DEFECTIVE genes in *Arabidopsis*. *C R Biol* 331:726–736.
11. Meinke D, Muralla R, Sweeney C, Dickerman A (2008) Identifying essential genes in *Arabidopsis thaliana*. *Trends Plant Sci* 13:483–491.
12. Gehring M, Choi Y, Fischer RL (2004) Imprinting and seed development. *Plant Cell* 16 (Suppl):S203–S213.
13. Huh JH, Bauer MJ, Hsieh TF, Fischer RL (2008) Cellular programming of plant gene imprinting. *Cell* 132:735–744.
14. Breuninger H, Rikirsch E, Hermann M, Ueda M, Laux T (2008) Differential expression of WOX genes mediates apical-basal axis formation in the *Arabidopsis* embryo. *Dev Cell* 14:867–876.
15. Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. *Trends Plant Sci* 13:624–630.
16. Braybrook SA, et al. (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci USA* 103:3468–3473.
17. Chandrasekharan MB, Bishop KJ, Hall TC (2003) Module-specific regulation of the beta-phaseolin promoter during embryogenesis. *Plant J* 33:853–866.
18. Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in *Arabidopsis*. *Development* 130:6065–6073.
19. Mönke G, et al. (2004) Seed-specific transcription factors ABI3 and FUS3: Molecular interaction with DNA. *Planta* 219:158–166.
20. Yamamoto A, et al. (2009) *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* 58: 843–856.
21. Kawashima T, et al. (2009) Identification of cis-regulatory sequences that activate transcription in the suspensor of plant embryos. *Proc Natl Acad Sci USA* 106:3627–3632.
22. Vasil V, et al. (1995) Overlap of Viviparous1 (VP1) and abscisic acid response elements in the Em promoter: G-box elements are sufficient but not necessary for VP1 transactivation. *Plant Cell* 7:1511–1518.
23. Benedito VA, et al. (2008) A gene expression atlas of the model legume *Medicago truncatula*. *Plant J* 55:504–513.
24. Ma L, et al. (2005) Organ-specific expression of *Arabidopsis* genome during development. *Plant Physiol* 138:80–91.
25. Schmid M, et al. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37:501–506.
26. Jiao Y, et al. (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat Genet* 41:258–263.
27. Girke T, et al. (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol* 124:1570–1581.
28. Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell* 14:1191–1206.
29. Spencer MW, Casson SA, Lindsey K (2007) Transcriptional profiling of the *Arabidopsis* embryo. *Plant Physiol* 143:924–940.
30. Verdier J, et al. (2008) Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol Biol* 67:567–580.
31. Day RC, Herridge RP, Ambrose BA, Macknight RC (2008) Transcriptome analysis of proliferating *Arabidopsis* endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant Physiol* 148:1964–1984.
32. Kerk NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM (2003) Laser capture microdissection of cells from plant tissues. *Plant Physiol* 132:27–35.
33. Davidson EH, Levine MS (2008) Properties of developmental gene regulatory networks. *Proc Natl Acad Sci USA* 105:20063–20066.
34. Mayer K, et al. (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402:769–777.
35. Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *Proc Natl Acad Sci USA* 102:17531–17536.
36. Redman JC, Haas BJ, Tanimoto G, Town CD (2004) Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *Plant J* 38:545–561.
37. Lockhart DJ, et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675–1680.
38. Windsor JB, Symonds VV, Mendenhall J, Lloyd AM (2000) *Arabidopsis* seed coat development: morphological differentiation of the outer integument. *Plant J* 22: 483–493.
39. Penfield S, Meissner RC, Shoue DA, Carpita NC, Bevan MW (2001) MYB61 is required for mucilage deposition and extrusion in the *Arabidopsis* seed coat. *Plant Cell* 13: 2777–2791.
40. Jensen MK, et al. (2008) Transcriptional regulation by an NAC (NAM-ATAF1,2-CUC2) transcription factor attenuates ABA signalling for efficient basal defence towards *Blumeria graminis* f. sp. *hordei* in *Arabidopsis*. *Plant J* 56:867–880.
41. Alonso R, et al. (2009) A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of *Arabidopsis* seed maturation gene expression based on heterodimerization and protein complex formation. *Plant Cell* 21:1747–1761.
42. Kwon CS, Chen C, Wagner D (2005) WUSCHEL is a primary target for transcriptional regulation by SPLAYED in dynamic control of stem cell fate in *Arabidopsis*. *Genes Dev* 19:992–1003.
43. Kotak S, Vierling E, Bäumlein H, von Koskull-Döring P (2007) A novel transcriptional cascade regulating expression of heat stress proteins during seed development of *Arabidopsis*. *Plant Cell* 19:182–195.
44. Harada JJ (2001) Role of *Arabidopsis* LEAFY COTYLEDON genes in seed development. *J Plant Physiol* 158:405–409.
45. Li Z, Thomas TL (1998) PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in *Arabidopsis*. *Plant Cell* 10:383–398.
46. Sorensen MB, Chaudhury AM, Robert H, Bancharrel E, Berger F (2001) Polycomb group genes control pattern formation in plant seed. *Curr Biol* 11:277–281.
47. Alonso JM, et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301:653–657.
48. Walling L, Drews GN, Goldberg RB (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc Natl Acad Sci USA* 83:2123–2127.
49. Goldberg RB, Hoshchek G, Tam SH, Ditta GS, Breidenbach RW (1981) Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis. *Dev Biol* 83:201–217.
50. Galau GA, Dure L, 3rd (1981) Developmental biochemistry of cottonseed embryogenesis and germination: Changing messenger ribonucleic acid populations as shown by reciprocal heterologous complementary deoxyribonucleic acid–messenger ribonucleic acid hybridization. *Biochemistry* 20:4169–4178.
51. McElver J, et al. (2001) Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* 159:1751–1763.
52. Kiyosue T, et al. (1999) Control of fertilization-independent endosperm development by the MEDEA polycomb gene in *Arabidopsis*. *Proc Natl Acad Sci USA* 96:4186–4191.
53. Tang W, Perry SE (2003) Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. *J Biol Chem* 278:28154–28159.
54. Hennig L, Menges M, Murray JA, Gruissem W (2003) *Arabidopsis* transcript profiling on Affymetrix GeneChip arrays. *Plant Mol Biol* 53:457–465.
55. Zhu T, Wang X (2000) Large-scale profiling of the *Arabidopsis* transcriptome. *Plant Physiol* 124:1472–1476.
56. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. *Genome Biol*, 2: research0032.0031–0032.0011.
57. Bowman JL, Mansfield SG (1993) Embryogenesis: Introduction. *Arabidopsis: An Atlas of Morphology and Development*, ed Bowman JL (Springer, New York), pp 351–361.
58. Lotan T, et al. (1998) *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93:1195–1205.