Early Disruption of Maternal–Zygotic Interaction and Activation of Defense-Like Responses in Arabidopsis Interspecific Crosses

Diana Burkart-Waco, a Kathie Ngo, a Brian Dilkes, b Caroline Josefsson, c and Luca Comai a

Department of Plant Biology and Genome Center, University of California, Davis, California 95616
Department of Horticulture and Landscape Architecture, Purdue University, Lafayette, Indiana 47907
Department of Biology, Vancouver Island University, Vancouver, Canada V9R 5S5

Seed death resulting from hybridization between Arabidopsis thaliana and Arabidopsis arenosa has complex genetic determination and involves deregulation 5 to 8 d after pollination (DAP) of AGAMOUS-LIKE genes and retroelements. To identify causal mechanisms, we compared transcriptomes of compatible and incompatible hybrids and parents at 3 DAP. Hybrids misexpressed endosperm and seed coat regulators and hyperactivated genes encoding ribosomal, photosynthetic, stress-related, and immune response proteins. Regulatory disruption was more severe in Columbia-0 hybrids than in C24 hybrids, consistent with the degree of incompatibility. Maternal loss-of-function alleles for endosperm growth factor and defense response regulators NON-EXPRESSOR OF PATHOGENESIS RELATED1 and HAiku1 increased hybrid seed survival. The activation of presumed POLYCOMB REPRESSIVE COMPLEX (PRC) targets, together with a 20-fold reduction in expression of FERTILIZATION INDEPENDENT SEED2, indicated a PRC role. Proximity to transposable elements affected natural variation for gene regulation, but transposon activation did not differ from controls. Collectively, this investigation provides candidates for multigenic orchestration of the incompatibility response through disruption of endosperm development, a novel role for communication between endosperm and maternal tissues and for pathways previously connected to immunity, but, surprisingly, does not identify a role for transposons.

INTRODUCTION

Parental combination as well as the genotype of each parent determines hybrid phenotype, suggesting that the outcome of hybridization depends on the balance of positive and negative interactions between parental genomes. In one outcome, heterosis, hybrids are more vigorous than either parent as exemplified by increased yield in maize intraspecific hybrids (reviewed in Springer and Stupar, 2007). Heterosis has been attributed to positive gene interactions (reviewed in Hochholdinger and Hoecker, 2007; Chen, 2010) ranging from genome-wide effects to single-locus effects (Lippman and Zamir, 2007; Krieger et al., 2010). Hybridization can also lead to reduced hybrid fitness, including seed failure, germination defects, hybrid necrosis, and sterility. Manifestation of these traits is thought to result from the interaction of diverged parental genes and result in hybrid incompatibly (Dobzhansky, 1936; Muller, 1942; Orr, 1995; reviewed in Bomblies et al., 2007; Rieseberg and Blackman, 2010). These interactions can affect different targets including pathogen responses (Bomblies et al., 2007; Jeuken et al., 2009; Yamamoto et al., 2010; Mizuno et al., 2011), suppression of transposable elements (TEs) (McClintock, 1984; Shaked et al., 2001; Kashkush et al., 2003; Madlung et al., 2005; Josefsson et al., 2006; Ungerer et al., 2006; Martienssen, 2010), small RNA pathways (Ha et al., 2009; Ng et al., 2012; Shivaprasad et al., 2012; Zhang et al., 2012), and developmental regulatory pathways, such as genomic imprinting in the seed (Josefsson et al., 2006). Importantly, although all these molecular mechanisms have been documented as affected by hybridization in at least one system, their relative contribution to hybrid incompatibility is unclear.

In Arabidopsis species, there is natural variation for interspecific hybrid incompatibility, and the molecular nature of seed death remains poorly understood (Bark-Waco et al., 2012). When Arabidopsis thaliana is crossed to its close relative Arabidopsis arenosa, abnormalities are observed in later stages of seed development (Josefsson et al., 2006; Walia et al., 2009; Burkart-Waco et al., 2012). This cross can only be performed using A. thaliana as the seed parent (Comai et al., 2000; Bushell et al., 2003). The genetic basis of the differential response between highly incompatible A. thaliana Columbia-0 (Col-0; 0% live seed) and moderately compatible C24 (17% live seed) was mapped to seven additive and several more epistatic quantitative trait loci.
(QTL) (Burkart-Waco et al., 2012). The identity of these QTL and the mode of action of the genes causing incompatibility are unknown. Studies of interploidy (tetraploid × diploid) interspecific Arabidopsis hybrids has revealed that ectopic expression of transcription factors involved in endosperm development contributes to incompatibility (Josefsson et al., 2006; Walia et al., 2009). These results have led to the hypotheses that (1) there are multiple factors governing the incompatibility response, and (2) aberrant gene regulation may underlie incompatibilities in Arabidopsis hybrids.

Here, we further characterize the molecular mechanism of hybrid incompatibility in Arabidopsis and assess TE activation in response to interspecific hybridization. We investigated seed development and the transcriptomes of developing hybrid seeds from incompatible Col-0 × A. arenosa and moderately compatible C24 × A. arenosa crosses. Early embryo development (2 to 6 d) was delayed in both incompatible and compatible hybrids, and differences in endosperm development were limited. Transcriptomes of seeds at 3 d after pollination (DAP) exhibit transcriptional disruption affecting genes at the maternal-filial interface, suggesting that the endosperm is the primary cause of incompatibility. We identified 1228 genes differentially expressed between the compatible and incompatible hybrid crosses, including 10 AGAMOUS-LIKE (AGL) transcripts and an overrepresentation of ribosomal, photosynthetic, and defense response genes. Some AGLs and the defense response genes also vary between Col-0 and C24 accessions in self-crosses, suggesting that maternal natural variation is conserved in hybrids at 3 DAP. Maternal transmission of nullimorphic alleles of the misregulated endosperm growth factor HAUKUT1 (KU1) and seed coat transcription factor TRANSPARENT TESTA GLABRA2 (TTG2) was sufficient to improve hybrid seed survival. Our data do not support a role for transposon activation in embryo lethality, but a loss-of-function mutation in an important regulator of pattern responses, NON-EXPRESSOR OF PR genes 1 (NPR1), improved hybrid seed survival, suggesting that a factor central to basal immune responses impacts postzygotic incompatibility in Arabidopsis. Because of the regulatory malfunction at many defense response genes in incompatible hybrids, we propose that hybridization can induce stress responses that result in hybrid necrosis (Bombilies et al., 2007).

RESULTS

Developmental Delays in seed of A. thaliana × A. arenosa Hybrids

We compared seed development of A. thaliana ecotypes Col-0 and C24, which respectively produce 0 to 1% and ~17% live seed after interspecific hybridization to A. arenosa (Burkart-Waco et al., 2012). Photomicrographs of developing seed from A. thaliana controls and interspecific F1 hybrid seed are shown in Figure 1. A. arenosa embryos grow much slower than A. thaliana (see Supplemental Figure 1 online, scale bar = 60 µm). For example, in A. arenosa, embryos are only one to four celled at 5 DAP, but endosperm proliferates rapidly in comparison to that of A. thaliana and hybrids (Bushell et al., 2003). Staining with vital dyes failed to detect dead cells in any cross at 2 to 6 DAP, although dead seed were clearly visible as early as 9 to 11 DAP, bars = 60 µm (see Supplemental Figure 2 online). Compared with A. thaliana seeds, interspecies hybrid embryos from both ecotypes were similar at 2 DAP, but a few C24 hybrids advanced to higher stages (Figure 1). At all stages beyond this, the hybrid embryos were developmentally delayed and similar between Col-0 and C24, with the exception of a few C24 × A. arenosa embryos displaying developmental progression to heart stage by 6 DAP, whereas no Col-0 embryos made this transition. Endosperm of Col-0 × A. arenosa hybrid seed was delayed at 3 DAP, as shown in Supplemental Figures 3A and 3B online (74.0 nuclei ± 3.8; t test, P = 0.001 versus 95.6 nuclei ± 4.1 in selfed Col-0), but C24 interspecific hybrid embryos appeared nearly normal at 3 DAP (82.0 ± 9.2; t test, P = 0.12 versus 94.5 ± 4.2 in selfed C24). Interestingly, the maternally derived inner integument was thinner in Col-0 hybrids than in selfed seed (see Supplemental Figure 3E online). In summary, incompatible Col-0 hybrids at 3 DAP displayed subtle reduction of endosperm proliferation and maternal tissue symptoms when compared with C24 hybrids (see Supplemental Figure 3 online). The stronger delay in the hybrid embryos could result from incompatibilities that either affected the embryo and not the endosperm, or the endosperm and, indirectly, the embryo. In addition, embryonic delay may result from the genetic influence of the A. arenosa seed father (see Supplemental Figure 1 online).

RNA Expression Profiling by Sequencing

To identify any molecular abnormalities during early seed development, we performed transcriptome analysis of 3 DAP seed by RNA sequencing. This stage was chosen because hybrid endosperms are relatively normal in C24 and only slightly delayed in Col-0. We chose to compare seeds with similar numbers of endosperm nuclei, rather than selecting seeds by embryo stage. Key epigenetic regulators of seed development are transcribed primarily in the endosperm (Baroux et al., 2002; Gehring et al., 2004; Berger and Chaudhury, 2009), and the abnormal regulation of gene expression in the endosperm has been suggested to underlie interspecific incompatibility (Haig and Westoby, 1991; Gutierrez-Marcos et al., 2003; Josefsson et al., 2006). The high degree of phenotypic variability observed later in seed development, even within a seed class (e.g., seeds at 11 DAP in Supplemental Figure 2B online), suggest that this focus on early development when organ growth was more consistent (see Supplemental Figure 3B online) is less sensitive to spurious secondary effects on growth.

To examine the differences in gene expression between hybrid and control seed, we used high-throughput RNA sequencing, as illustrated in Supplemental Figure 4 online, and our cDNA sequencing statistics are presented in Supplemental Table 1 online. We used 3 DAP seed from Col-0 × A. arenosa and C24 × A. arenosa hybrids and used Col-0, C24, and A. arenosa seed from selfing or intersib matings for controls (see Supplemental Figure 4A online). We compared mapping of sequence reads to a reference cDNA database from A. thaliana alone or from both parents and opted for the first approach (see Methods for technical details). A gene was considered expressed if a mean of four sequence reads (averaged across both replicates) was mapped unequivocally to it. We found 20,768 genes expressed in both Col-0 and C24 (see Supplemental Figure 5 online). In the hybrid endosperm data, 18,483 genes were expressed in both Col-0 × A. arenosa and C24 × A. arenosa (see Supplemental Figure 5 online). Control and hybrid expression profiles were compared...
Figure 1. Early Seed Development of *A. thaliana* and *A. thaliana × A. arenosa* Hybrids.
with sets of genes whose expression was characterized in different seed compartments (http://seedgenenetowork.net; Le et al., 2010). We refer to these annotated sets of transcripts as “Harada-Goldberg” transcripts, after the senior authors of the Le et al. (2010) study. Greater than 92% of the genes identified as expressed at the globular stage in Le et al. were detected in controls. Similarly, >91% of the genes expressed at the globular stage and genes expressed were detected in our hybrid seed transcriptomes. The majority of molecular signatures observed in the hybrid transcriptome analysis matched the seed phenotypes observed in the embryogenesis time course (Figure 1; see Supplemental Figure 5 online). Thus, the observed developmental delay, rather than a generalized loss of transcriptional regulation, is the most likely cause of differences in gene expression between A. thaliana selfed seed and interspecies hybrids.

Comparison of the Transcriptomes of Control and Hybrid Crosses

We compared transcriptional profiles of Col-0 to Col-0 × A. arenosa hybrids and C24 to C24 × A. arenosa hybrids to identify differentially expressed genes. We found expression of 509 genes differed between Col-0 and Col hybrids, and 4206 were differentially expressed between C24 and C24 hybrids (see Supplemental Figure 6 online). Of the 509 differences between Col-0 and Col hybrids, 73% were also found among the set of differences between C24 and C24 hybrids. The numbers of differentially expressed genes in Col-0 and C24 hybrids was unexpected. It is unlikely that disparities between differentially expressed genes is attributed to biological variation within a genotype, as there is a high degree of correlation (≥91%) in expression values between replicates (see Supplemental Figure 7 online). Furthermore, if transcriptional dysregulation is the cause of incompatibility, one would expect the greater degree of transcriptional differences to be found between Col-0 and Col-0 hybrids, which exhibit stronger incompatibilities.

To characterize differentially expressed genes, Gene Ontology (GO) enrichment was assessed relative to the proportions of genes found in the A. thaliana genome (see Supplemental Table 2 online) and a set of 20,768 transcripts with coverage more than four reads in both 3 DAP Col-0 and C24, which we called the “seed genes” (see Supplemental Figure 5 and Supplemental Data Set 1 online). Genes that were differentially expressed in both hybrids (371 genes; see Supplemental Figure 6 online) were enriched for genes classified in cell killing, photosynthesis, cell wall modifications, and defense responses (see Supplemental Tables 2 and 3 and Supplemental Data Set 2 online). Very similar GO categories and fold enrichments were observed for both The Arabidopsis Information Resource and seed gene background sets (see Supplemental Tables 2 and 3 online). For example, we observed a fourfold enrichment for cell killing in developmentally expressed transcripts relative to all Arabidopsis genes and a 6.5-fold enrichment relative to seed transcripts (see Supplemental Tables 2 and 3 online). Genes that were specifically differentially expressed between Col-0 and Col-0 hybrid (138 genes; see Supplemental Figure 6 online) also displayed enrichment for cell killing genes and ribosomal protein genes, with both classes induced in hybrids (see Supplemental Tables 2 and 3 online). The genes annotated as “cell killing,” all low molecular weight Cys-rich peptides (CRPs) and defensin-like proteins, are expressed in a manner inconsistent with inducible hypersensitive defenses response since they were also moderately expressed in control seeds (see Supplemental Table 4 online). These results suggest that defense responses, or related pathways, could play a role in the abnormal development of hybrids.

Gene Misregulation Is Limited to Particular Seed Compartment

To assess the role of the hybridization-sensitive genes during seed development, we took the set of genes that were differentially regulated between Col-0 and Col-0 × A. arenosa and analyzed their patterns of tissue-specific expression at different stages of normal seed development using the Harada-Goldberg laser capture microdissection (LCM) seed expression data set (Figure 2). Of the 509 genes that were significantly up- or down-regulated between Col-0 and Col-0 hybrids, 396 were represented in the LCM set. Some of these genes clustered together as developmentally coexpressed in their normal regulation, and these clusters responded to hybridity in a consistent manner. For example, a set of genes expressed preferentially in the general seed coat with functions in secondary metabolism (primarily flavonoid biosynthesis) was downregulated in hybrids. By contrast, a set of genes expressed specifically in the chalazal end of the seed coat, which included extensins and the cell identity and layer specification factor SHORT ROOT (Levesque et al., 2006), were coordinately overexpressed in hybrids. The regulatory changes in seed coat gene expression in the hybrids confirm the intimate
Figure 2. Hybridization Disrupts Specific Spatio-Temporal Patterns of Seed Development.

The seed compartments in which hybridization sensitive genes were normally expressed were identified using the Harada-Goldberg LCM seed expression data set, and the expression patterns were used to cluster the 396 represented genes out of the 509 that were significantly up- or downregulated between Col-0 and Col-0 × A. arenosa (A) and between the incompatible Col-0 hybrid and the compatible C24 hybrid (B). The heat maps represent patterns of expression for genes (rows) across seed compartments (columns) at five developmental stages from 2 DAP (preglobular) to 10
connection between fertilization products and maternal tissue (Garcia et al., 2005; Dlikes et al., 2008). A set of genes preferentially expressed in the endosperm adjacent to the chalazal end were upregulated in hybrids and included transporters and endosperm transcription factors. Notably, three of these clusters that were upregulated in response to hybridization (such as photosynthetic genes) remain relatively constant in normal A. thaliana from preglobular to heart stage (Figure 2, heat map) inconsistent with a heterochronic malfunction resulting in hybrid lethality.

Hybrids Display Unique Seed Gene Expression Signatures Relative to Both A. thaliana and A. arenosa Parents

Nonadditive gene regulation was tested by comparing the transcriptionomes of the hybrids and seed mother to the seed transcriptionome of A. arenosa seed at 3 DAP. The quality of the latter RNA sequencing libraries was lower than those derived from selfed and hybridized A. thaliana (see Supplemental Table 1 and Supplemental Figure 7 online; see Methods), resulting in fewer high-quality reads. Nonetheless, the comparison revealed nonadditivity for a majority (293) of the 509 genes significantly regulated between A. thaliana versus Col-0 hybrid (see Supplemental Data Set 3 online). Of these 293, over 200 were regulated in the same direction (i.e., up in hybrid relative to both parents) and included endosperm transcription factors IKU1, HDG9, and HDG10. The overall response, which was enriched relative to our seed genes for cell killing (P < 0.0001), defense response (P = 0.004), and translation (P = 0.02), suggested that hybrids resembled the seed parent more than the pollen parent. Col-0 hybrid and C24 hybrid shared 84% of the differentially expressed genes between Col-0 hybrid and A. arenosa (3592 and 3975, respectively) with over half (2305/3592) upregulated in hybrid relative to A. arenosa, about twice more than those observed between Col-0 hybrid and Col-0.

To better understand the influence of each parent on the hybrid, we identified genes whose regulation was conserved between one parent and the hybrid, but differed from the other parent. Relative to seed genes, 3298 A. thaliana–like genes were predominately upregulated in the hybrid relative to the A. arenosa parent (2373 of 3298 genes, mean fold change 24×). These were enriched in secondary metabolic processes (100 genes, twofold enrichment, P < 0.0001), phenylpropanoid biosynthesis (47 genes, twofold enrichment, P < 0.0001), and aromatic compound biosynthesis (55 genes, twofold enrichment, P < 0.001). The 214 A. arenosa–like genes were not distinctly regulated (either up or down) relative to A. thaliana (see Supplemental Data Set 3 online). They contained many photosystem proteins (upregulation relative to Col-0, corresponding to those observed in cluster analysis (Figure 2). In common with the A. arenosa pollen parent, the hybrid appeared deficient for polycomb protein FERTILIZATION INDEPENDENT SEED2 (FIS2), imprinted transcription factor AGAMOUS-LIKE (AGL96) (Wolff et al., 2011), and SEEDSTICK.

The overall pattern of differential gene expression between hybrids and parents is consistent with a release of gene suppression in Col-0 hybrids. Complications introduced by mapping specificity and reduced read numbers in the A. arenosa libraries relative to the A. thaliana and hybrid data sets are unlikely to account for frequent upregulation in Col-0 hybrid relative to A. arenosa.

Aberrant Expression of Endosperm-Localized Transcription Factors

We investigated misregulated transcription factors that are involved in endosperm development. FIS2, HAiku1, SEEDSTICK/Agl11, SHORT ROOT, HOMEODOMAIN GLABROUS9 (HDG9), HDG10, and ZHOUPI/RETARDED GROWTH OF EMBRYO1 (Chaudhury et al., 1997; Luo et al., 2000; Kondo et al., 2008; Yang et al., 2008; Mizzotti et al., 2012) were all misregulated in comparison to control crosses. A subset of these genes, FIS2, HDG9, and HDG10, is coexpressed during development (Figure 2; see Supplemental Figure 8A online). The expression of both ZHOUPI and FIS2 was sixfold to 20-fold higher in A. thaliana control crosses and the expression of HDG9/10 and IKU1 was lower in control crosses (Figure 2; see Supplemental Figure 8B online). Comparison of hybrid FIS2 expression to our previous studies (Walia et al., 2009) and the Harada-Goldberg Ath1 data (Le et al., 2010) indicates that FIS2 expression remains relatively constant through early seed development (Figure 2; see Supplemental Figure 8C online). ZHOUPI, a factor required for embryonic cuticle formation and embryo-endosperm separation, is reduced 15-fold in Col-0 hybrids (P = 0.048) versus sixfold in C24 hybrids (P < 0.05). Correspondingly, its target ABNORMAL LEAF SHAPE1 (ALE1), which encodes a subtilisin–like protein required for cuticle formation, is reduced in Col-0 hybrids (10×, P = 0.06) and not in C24-hybrids (0.64×, P = 0.9). The observed reduction in expression of two important seed development factors, ZHOUPI and FIS2, may compromise seed function.

To explore the role of the transcription factors that underwent abnormal activation, such as IKU1, we tested whether maternal transmission of loss-of-function mutations affected seed survival.

Figure 2. (continued).

DAP (mature green; see definitions at top). The median fold change for each cluster [A], control/hybridization; [B], compatible/incompatible) is displayed on the right.

(A) Interspecific hybridization causes decreased expression of seed coat genes that specify the accumulation of secondary metabolites, mostly flavonoids. Other gene clusters are impacted in the opposite way; for example, those expressed preferentially in endosperm and eventually in the embryo and encoding photosynthetic genes. Two additional clusters, one preferentially expressed in chalazal endosperm and the other in chalazal seed coat, illustrate acceleration of the chalazal pole program. Interestingly, three well-known regulators in the chalazal endosperm diverged from the cluster trend and were poorly expressed in the hybrids (box in bottom left).

(B) Compatible ecotype C24 differs distinctly from incompatible ecotype Col-0, resembling the normal developmental pattern. For example, embryonic and micropylar gene clusters display enhanced response, while the chalazal pole overexpression induced by hybridity (A) is diminished.
Since the pollen-parent in our hybrid crosses was wild-type A. arenosa, only haploinsufficiency, additive, or maternal effects could be observed. Because of the effect of hybridity on maternal seed coat and genes involved in the flavonoid pathway, we included in the analysis the maternally expressed factor TRANS-PARENT TESTA GLABRA2 (TTG2) (Johnson et al., 2002; Garcia et al., 2005), which also affects interploidy hybrid lethality (Dilkes et al., 2008), and IKU2 because of its relationship to IKU1 and endosperm development (Garcia et al., 2003, 2005). The genetic backgrounds of these mutants, Col-0 and Landsberg erecta-0 (Ler-0), influence percentage of live seed. Col-0 × A. arenosa hybrids displayed strong incompatibility. Similarly to hybrids of ecotype C24, hybrids of ecotype Ler-0 had partial fertility. Comparisons were always made to the respective control. Our results are presented in Table 1 and Figure 3. We found no effects from the inactivation of several candidates (Table 1). However, loss of IKU1 in Ler-0 background significantly improved seed survival (iku1 × A. arenosa, 7.1% live seed versus 4% live seed in Ler-0 × A. arenosa control) and maternal TTG2 loss tripled seed survival (Figure 3A). However, T-DNA mutation in the gene IKU2 did not significantly improve overall seed survival. The double mutant of iku2 ttg2 was not significantly different from the single ttg2 mutant for %normal + green + viviparous (%NGV) seed, consistent with additive maternal effects at least for %NGV (Figure 3A), although the ratio of %G to %N is altered. The TTG2 and IKU1 phenotype confirms the role of maternal-filial communication in regulating incompatibility (Dilkes et al., 2008).

Natural Variation for Defense Response Genes

We previously identified quantitative genetic variation in seed survival between Col-0 hybrids and C24 hybrids (Burkart-Waco et al., 2012). To connect this variation and the observed differences in expression, we identified those genes differentially expressed between the two ecotypes in the 3 DAP seed and silique libraries (two replicates each). A set of 1228 genes was differentially expressed between the two hybrids (Padj < 0.05) seed libraries and 753 for siliques (see Supplemental Figure 9 online). In addition, 1291 genes were differentially expressed between the Col-0 and C24 control seed libraries (Padj < 0.05). Supplemental Tables 5 and 6 online report the top six most significant GO enrichment terms (Padj < 0.05) as well as predictive protein motif (for full GO characterization, see Supplemental Data Set 2 online). Genes that were differentially expressed between Col-0 and C24 were enriched for many different environmental responses as well as secondary metabolism and defense response genes. Natural variation between Col-0 and C24 has been previously observed for freezing tolerance (Korn et al., 2008) and defense response gene activation in response to cucumber mosaic virus infection (Postnikova et al., 2011). Like Col-0 and C24, Col-0 × A. arenosa and C24 × A. arenosa displayed differential defense response, including Leucine-rich-repeat (LRR) proteins (see Supplemental Tables 5 and 6 online). Of the 1291 differentially expressed genes between Col-0 and C24, 738 genes (57%) had higher expression in C24 and 553 genes in Col-0 (see Supplemental Figure 9A online). The ratio of high Col-0:high C24 was significantly different from that observed in hybrid crosses (see Supplemental Figure 9A online) where, out of 1228 differentially expressed genes, 537 genes (44%) exhibited higher expression in C24 × A. arenosa and 691 genes higher expression in Col-0 × A. arenosa (\(\chi^2\) test; \(P = 1.78E-22\)). This was also observed for two biological replicates of both Col-0 and C24 hybrid 3 DAP siliques (see Supplemental Data Set 4 and Supplemental Figure 9A online). Therefore, relative to controls, in both seeds and siliques there was a net derepression of Col-0 genes in the hybrid state for both seeds and siliques, in particular for GO-enriched transcript (see Supplemental Tables 5 and 6 online).

Immune response genes were prominent among those that displayed natural variation posthybridization. The 303 transcripts (see Supplemental Figure 9C online) that were upregulated in Col-0 (relative to C24) and Col-0 × A. arenosa (relative to C24 × A. arenosa) were highly enriched for RNAs encoding proteins associated with apoptosis and defense response GO categories (see Supplemental Data Set 2 online). However, these classifications may require refinement as many of the putative proteins associated with defense are uncharacterized nucleotide binding site (NBS)-LRR genes or Cys-rich proteins for which developmental roles are possible (Marshall et al., 2011). The 250 seed transcripts that were uniquely upregulated in Col-0 hybrid also displayed defense response and cell killing signatures (\(P < 0.0001\)). Although a defense response signature was observed for genes that were high in C24 hybrid (relative to Col-0 hybrid), C24 (relative to Col-0) displayed no enrichment (see Supplemental Data Set 2 online). The upregulation of defense and immune response genes was even more predominant in Col-0 hybrid siliques transcriptomes as there was a 13.5-fold enrichment in apoptotic factors (see Supplemental Data Set 2 online). Together, this indicates that Col-0 and C24 have differential defense responses and perhaps immune response genes are activated in Col-0 and to a lesser degree in C24 hybrids posthybridization (see Supplemental Data Set 2 online). Therefore, natural variation in defense response factors present in maternal parents is preserved after hybridization and may play a role in incompatibility.

Suppressing Basal Defense Response Partially Rescues Hybrid Crosses

Natural variation for defense response between Col-0 and C24 led to the hypothesis that differential activation of defense response genes affects hybrid seed survival. We used mutants to explore the role of major defense pathways. Mutant loci included both master regulators and downstream pathogen effectors. NONEXPRESSOR OF PR GENES (NPR1), which controls systemic acquired resistance through the expression of pathogenesis-related proteins, was a likely candidate gene as it is activated in response to stress and is known to promote cell death via salicylic acid signaling (Mukhtar et al., 2009; Spoel et al., 2009). Seed survival after hybridization of npr1 mutants is presented in Figure 3, Table 1, and Supplemental Table 7 online. Both npr1-1 and npr1-5 mutants produced more viviparous seed than their corresponding control crosses (Figure 3B; see Supplemental Table 7 online). Inactivation of SALICYLIC ACID INDUCTION-DEFICIENT2 (SID2), which encodes the enzyme isochorismate synthase (Wildermuth et al., 2001) required for salicylic acid synthesis, also had a beneficial effect on seed survival, a twofold increase, consistent with the npr1 mutant phenotype (Table 1; see Supplemental Table 7 online). Taken together, these results are consistent with a negative effect.
of defense and stress response activation in response to hybridization across ecotypes and a potential role for salicylic acid–dependent defense signaling in *Arabidopsis* hybrid seed death.

The Activity of Genes That Are Normally Associated with Polar Compartments Differentiates Compatible from Incompatible Responses

To connect seed tissue differentiation to the incompatibility response, we clustered the 1228 genes that differed in expression between the hybrids according to the tissue and temporal series provided by the Harada-Goldberg LCM study (Figure 2). Interestingly, clusters of genes defined by common regulation in normal seed development responded consistently. C24 hybrids displayed ~8 times higher expression of genes associated with micropylar endosperm, suspensor and embryo, averaging above the mean expression of self-cross controls, while Col-0 hybrids averaged below the mean expression of controls. These genes encoded potential regulators AGL48, AGL57, AGL96, and AUXIN RESPONSE FACTOR12 (ARF12), ARF13, ARF21, FLOWERING WAGENINGEN FWA, SUVH8, and WRKY DNA-BINDING PROTEIN10 WRKY10. Genes whose normal expression was associated with the chalazal region had an opposite response, displaying higher expression in Col-0 hybrids. Therefore, relative overexpression of genes normally localized to the chalazal pole was associated with the worst outcome.

**Investigation of AGL Proteins**

AGL proteins, such as AGL62, play an important role in endosperm development (Kang et al., 2008; Shirzadi et al., 2011), and differential regulation of AGL genes is associated with deleterious effects in interspecific-interploidy *Arabidopsis* hybrid fitness (Josefsson et al., 2006; Walia et al., 2009). We found that 10 AGLs, including the previously documented endosperm cluster, were differentially expressed between the Col-0 hybrid and C24

### Table 1. Characterization of Candidate Genes

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene</th>
<th>Locus</th>
<th>Mean&lt;sup&gt;a&lt;/sup&gt; Col-0</th>
<th>Mean C24</th>
<th>Mean Aa</th>
<th>Mean Col-0 X Aa</th>
<th>Mean C24 X Aa</th>
<th>Seed Phenotype At x Aa&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed transcription factors</td>
<td>FIS2</td>
<td>AT2G35670</td>
<td>115.4</td>
<td>222.5</td>
<td>2.2</td>
<td>13.3</td>
<td>11.1</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>PHERES1</td>
<td>AT1G65330</td>
<td>303.6</td>
<td>90.9</td>
<td>0.0</td>
<td>551.1</td>
<td>1388.1</td>
<td>9.9&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>TTG2</td>
<td>AT2G37260</td>
<td>771.1</td>
<td>705.0</td>
<td>84.5</td>
<td>758.5</td>
<td>614.3</td>
<td>15.28&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>IKU1</td>
<td>AT2G53230</td>
<td>200.0</td>
<td>171.0</td>
<td>44.5</td>
<td>488.0</td>
<td>519.6</td>
<td>7.1&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>IKU2</td>
<td>AT3G19700</td>
<td>16.1</td>
<td>8.0</td>
<td>0.0</td>
<td>5.2</td>
<td>6.7</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL2</td>
<td>AT1G15800</td>
<td>10.3</td>
<td>9.8</td>
<td>8.9</td>
<td>13.8</td>
<td>19.1</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL25</td>
<td>AT5G10140</td>
<td>12.2</td>
<td>56.9</td>
<td>211.3</td>
<td>51.9</td>
<td>35.4</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL12</td>
<td>AT1G71692</td>
<td>102.3</td>
<td>62.4</td>
<td>2.2</td>
<td>26.7</td>
<td>54.8</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL27</td>
<td>AT1G77080</td>
<td>330.6</td>
<td>3.4</td>
<td>522.5</td>
<td>115.3</td>
<td>118.9</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL29</td>
<td>AT2G34440</td>
<td>24.6</td>
<td>8.9</td>
<td>44.5</td>
<td>43.5</td>
<td>53.4</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL31&lt;sup&gt;b&lt;/sup&gt;</td>
<td>AT5G65050</td>
<td>546.1</td>
<td>107.5</td>
<td>84.5</td>
<td>223.3</td>
<td>257.8</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL42</td>
<td>AT5G62165</td>
<td>22.0</td>
<td>4.0</td>
<td>88.9</td>
<td>1.8</td>
<td>7.0</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL48&lt;sup&gt;b&lt;/sup&gt;</td>
<td>AT2G40210</td>
<td>48.1</td>
<td>43.2</td>
<td>0.0</td>
<td>104.1</td>
<td>167.8</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL62</td>
<td>AT5G60440</td>
<td>147.6</td>
<td>83.0</td>
<td>2.2</td>
<td>322.2</td>
<td>617.0</td>
<td>10.9&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>AGL67</td>
<td>AT1G22590</td>
<td>50.8</td>
<td>107.7</td>
<td>2.2</td>
<td>85.4</td>
<td>131.3</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>AGL96&lt;sup&gt;b&lt;/sup&gt;</td>
<td>AT5G06500</td>
<td>109.5</td>
<td>96.5</td>
<td>0.0</td>
<td>147.0</td>
<td>224.9</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>VRN2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>AT4G16845</td>
<td>366.7</td>
<td>139.9</td>
<td>244.6</td>
<td>182.1</td>
<td>169.7</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>VIM5</td>
<td>AT1G57800</td>
<td>95.8</td>
<td>90.6</td>
<td>11.1</td>
<td>707.2</td>
<td>947.4</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>HDG9</td>
<td>AT3G17320</td>
<td>189.2</td>
<td>229.5</td>
<td>6.7</td>
<td>816.1</td>
<td>1179.9</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>HDG10</td>
<td>AT1G43650</td>
<td>67.2</td>
<td>10.2</td>
<td>17.8</td>
<td>222.1</td>
<td>241.6</td>
<td>–</td>
</tr>
<tr>
<td>Defense response</td>
<td>PCC1</td>
<td>AT3G22231</td>
<td>1.4</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>NPR1</td>
<td>ATIG64280</td>
<td>43.4</td>
<td>18.2</td>
<td>37.8</td>
<td>18.1</td>
<td>30.5</td>
<td>1.8&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>NBS-LRR</td>
<td>AT3G44630</td>
<td>152.6</td>
<td>28.0</td>
<td>6.7</td>
<td>59.0</td>
<td>31.8</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>RPS1</td>
<td>AT3G07040</td>
<td>30.6</td>
<td>0.0</td>
<td>13.3</td>
<td>4.7</td>
<td>2.7</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>EDS1</td>
<td>AT3G48090</td>
<td>29.2</td>
<td>12.3</td>
<td>100.0</td>
<td>19.1</td>
<td>14.6</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>CDR2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>AT5G33340</td>
<td>32.7</td>
<td>6.4</td>
<td>4.4</td>
<td>84.4</td>
<td>109.0</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>NDR1</td>
<td>AT3G26060</td>
<td>10.9</td>
<td>8.0</td>
<td>26.7</td>
<td>4.7</td>
<td>3.5</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>SID2</td>
<td>AT1G74710</td>
<td>78.3</td>
<td>67.3</td>
<td>229.0</td>
<td>125.6</td>
<td>156.8</td>
<td>ns</td>
</tr>
</tbody>
</table>

<sup>a</sup>The mean number of read counts (normalized to library size) from two biological replicates.

<sup>b</sup>Benjamini-Hochberg adjusted P value from differential expression analysis of Col-0 hybrid versus C24 hybrid in DESeq.

<sup>c</sup>The percentage of live seed at maturity when crossed to *A. arenosa*.

<sup>d</sup>Significant difference in live seed in comparison to Col-0 X Aa (0.8% live seed, *n* = 3893) in a Student’s *t* test at P < 0.05.

<sup>e</sup>Significant difference in live seed in comparison to Ler-0 X Aa (4.9% live seed, *n* = 832) in a Student’s *t* test at P < 0.05.
hybrid seed (see Supplemental Table 8 online). AGL expression was also confirmed in hybrid silique analysis for five of 10 genes. Only two of these genes were differentially expressed between control crosses (at \( \text{Padj} < 0.05 \)). However, the magnitude of expression in hybrid and control crosses was similar and in the same direction for some of these genes (see Supplemental Table 8 online), suggesting that natural variation between Col-0 and C24 may underlie the differences in \( \text{padj} \) (Table 1; see Supplemental Table 9 online), suggesting lack of maternal effect or redundancy in interspecific seed survival. Decreasing expression of apparently activated transcription factors improves seed survival. Additional, \( \text{np}1-1 \), \( \text{sid}2 \), and \( \text{np}1-5 \) mutants (\( \text{padj} \)) were different from Col-0 for mean \%NGV, as determined by Student’s \( t \) test at \( P < 0.05 \). Additionally, \( \text{np}1-1 \) and \( \text{np}1-5 \) had significantly higher \%viviparous seed (\( P < 0.05 \)). The ecotype background of wild-type control and \( \text{np}1-5 \) are the same. Line 1/8E, described as Nossen-0 (Shah et al., 1997), was determined in this study to be an unknown ecotype after original seed stocks were obtained and genotyped. The mean and \( \sigma \) are given for each cross. Letters indicate lines that were significantly different from control in a Student’s \( t \) test at \( P < 0.05 \).

**Figure 3.** Endosperm Transcription Factors and Defense Response Regulators Affect Hybrid Seed Survival. (A) \( \text{Ler}0 \) and mutants in \( \text{Ler}0 \) background (\( \text{x} \) axis), \( \text{iku}1 \), \( \text{iku}2 \), \( \text{phe}1 \), and \( \text{ttg}2 \) and \( \text{ttg}2 \) were crossed as seed parents to \( \text{A. arenosa} \), and the %NGV (i.e., live) was scored for each genotype (\( \text{y} \) axis). \( \text{Ler}0 \) (\( n = 832 \) seeds), \( \text{iku}1 \) (\( n = 1617 \) seeds), \( \text{iku}2 \) (\( n = 2836 \) seeds), \( \text{phe}1 \) (\( n = 1920 \) seeds), \( \text{ttg}2 \) (\( n = 2404 \)), and \( \text{ttg}2 \) were different from Col-0 for mean %NGV, as determined by Student’s \( t \) test at \( P < 0.05 \). Additionally, \( \text{ip}1-1 \) and \( \text{ip}1-5 \) had significantly higher %viviparous seed (\( P < 0.05 \)). The ecotype background of wild-type control and \( \text{np}1-5 \) are the same. Line 1/8E, described as Nossen-0 (Shah et al., 1997), was determined in this study to be an unknown ecotype after original seed stocks were obtained and genotyped. The mean and \( \sigma \) are given for each cross. Letters indicate lines that were significantly different from control in a Student’s \( t \) test at \( P < 0.05 \).

**Transposon Activation Is Sporadic in Hybrid 3 DAP Seed**

TE genes are reported to be induced in incompatible hybrid crosses (McClintock, 1984; Josefsson et al., 2006). To address the potential role of transposable elements and the connected natural variation in hybrid seed death, we surveyed relative expression level in the early seed transcriptome. We considered both 3900 genes annotated as transposable element genes (TEGs; such as AT1G37170) as well as 31,189 \( \text{A. thaliana} \) TEs (such as AT1TE21600) in the TAIR10 genome release, to which we aligned reads that did not map to the genes. Our findings are presented in Supplemental Table 10 and Supplemental Figure 10 online. We found that ~3000 TEs were expressed in both hybrid and \( \text{A. thaliana} \) seed libraries. TE expression was low for both hybrid and control crosses (mean nine to 26 reads for TE and four to 36 reads for TEG) relative to the mean read count per gene in our libraries, which ranged from 300 to 400 reads and median 22 to 56 reads (see Supplemental Table 10 online). To account for variability between TE features expression, we summed read counts between replicates instead of averaging them and then analyzed fold change in Col-0 hybrid relative to Col-0. We identified relatively few TEs that were differentially expressed between hybrid and \( \text{A. thaliana} \) controls. Expression of these TEs was equally probable to be high in Col-0 or in Col-0 hybrid (see Supplemental Figure 10 online). This trend was also identified for three differentially expressed TE genes that were part of the 509 differentially expressed genes Col-0 versus Col-0 hybrid; two out of three were high in Col-0 (see Supplemental Table 11 online).

To determine how natural variation affected transposon expression, we compared Col-0 and C24 to identify differentially expressed members of the TE set. We found 56 TEG among the 1228 differentially expressed genes identified during hybridization to \( \text{A. arenosa} \). In self crosses, 39 TEG could be identified among the 1291 differentially expressed genes. There was a large overlap between the differentially expressed hybrid and control TEG (33 out of 39 control TEG found in hybrid set), suggesting that some transposons maintain their normal pattern of regulation in the hybrid state.

We were concerned that our transposon analysis might be affected by mapping bias potentially underreporting TEs of \( \text{A. arenosa} \). If TEs were too divergent to map on the reference...
Hybrid incompatibility subverts regulation of endosperm and seed coat

Hybrid incompatibility has been attributed to multiple mechanisms, including immune responses, genomic shock resulting in TEs activation, and regulatory disruption of normal developmental pathways. To help discriminate between these mechanisms, we investigated the early regulatory changes associated with hybrid incompatibility and manifested in developing seed by comparing selfed and hybrid seed in two ecotypes differing by the degree of incompatibility. Embryonic delay in comparison to maternal self crosses was obvious at 3 DAP, while endosperm development appeared normal or nearly normal in compatible hybrids. This contrasts with the endosperm phenotype of incompatible interpollody crosses, which displayed accelerated proliferation (Scott et al., 1998; Dilkes et al., 2008). Endosperm overproliferated relative to the cognate embryo but not relative to seed age (Figure 1; see Supplemental Figure 3 online). These observations could have three explanations: First, the hybrid embryo displays additive regulation developing at a rate intermediate between the parental ones. Second, hybrid incompatibility is manifested predominantly in the embryo. Third, hybrid incompatibility affects early endosperm function but not its growth. Physiological impairment of the endosperm results in delayed embryo development. To distinguish these three hypotheses, we undertook analysis of the transcriptome.

Comparison of transcriptomes in control and hybrids revealed dramatic changes in genes normally associated with the endosperm and with the seed coat (Figure 2). Furthermore, it indicated that genes that are normally localized in the chalazal compartment undergo a distinct regulatory fate, suggesting that the chalazal compartment responds differently from the rest of the seed. Regulatory changes in the seed coat were indicated by the suppression of a gene cluster specifying proteins of secondary metabolism that are preferentially expressed during normal development in the general seed coat (Figure 2). At the same time, a cluster of coregulated genes expressed preferentially in the chalazal end seed coat was upregulated in incompatible Col-0 relative to controls. This suggests that the transcriptional response observed in hybrids is affected along the polarity axis of the seed.

Disruption of seed coat function is accompanied by increase in gene clusters that are preferentially expressed in endosperm or in chalazal endosperm. Some of these, such as the set of photosynthetic genes preferentially expressed in endosperm during seed development (Figure 2; see Supplemental Tables 2 and 3 online), could depend on pollen-parental influence. Underrepresentation of embryonic cells in the seed probably hindered detection of embryonic changes. Taken together with the rest of the regulatory changes, these data provide a molecular explanation for the time-proven practice of embryo rescue to achieve interspecific hybridization (Sharma et al., 1996). A probable scenario is one in which regulatory dysfunction in the endosperm affects both the maternal seed coat, which downregulates the secondary metabolite gene cluster, and the embryo, whose growth is compromised.

A Pathway Mediating Maternal-Zygotic Communication in the Seed Affects Hybrid Incompatibility

The gene encoding HA1K1, an endosperm factor affecting seed size, was upregulated in hybrids. HA1K1 interacts with HA1K2 and maternal factor TTG2 to regulate seed size through endosperm growth (Luo et al., 2005; Wang et al., 2010). TTG2 variation affects seed death in interpollody crosses within A. thaliana (Dilkes et al., 2008). We show that both HA1K1 and TTG2 knockouts have beneficial effects on hybrid seed survival (Table 1, Figure 3),

DISCUSSION

Hybrid Incompatibility Subverts Regulation of Endosperm and Seed Coat

Hybrid incompatibility and manifested in developing seed by comparing
The expression levels of FIS2 and ZHOUPI, which encode seed factors required for normal seed development, were significantly suppressed in both hybrids. For Col-0 and C24, control (self-cross) expression of ZHOUPI was 16 and 6 times that of the hybrid, respectively; expression of FIS2 was 20 and 11 times (see Supplemental Figure 8 online). ZHOUPI encodes a basic helix-loop-helix endosperm transcription factor that regulates development of the embryo-endosperm interface through targets, including ALE1. Loss of either ALE1 or ZHOUPI results in shriveled seed whose embryos lack a proper cuticle (Yang et al., 2008). ALE1 expression was significantly reduced in Col-0 hybrids, but not in C24 hybrids, consistent with ZHOUPI regulation. FIS2 encodes a subunit of the POLYCOMB REPRESSIVE COMPLEX (PRC). Its reduction might explain the previously reported alterations in the expression of Polycomb-regulated PHERES (Josefsson et al., 2006) and perhaps HDG9 and IKU1 misexpression (see Supplemental Figure 8C online). Interestingly, low FIS2 and ZHOUPI expression in A. arenosa could potentially explain its low expression in the hybrid. The A. arenosa expression pattern could result from its different developmental strategy in early seed development, namely, rapid endosperm proliferation coupled with delayed embryo growth (see Supplemental Figure 1 online; Bushell et al., 2003). Alternately, the mechanism governing ZHOUPI and FIS2 misexpression may also be responsible for HDG10 and IKU1 derepression. A candidate pathway is the one responsible for DNA methylation, as both FIS2 and HDG9 are imprinted via DNA methylation and maternally expressed (Julien et al., 2006; Gehring et al., 2009). The relationship between expression level and function in the seed is unknown. Reduction of the ZHOUPI target ALE1 in Col-0 hybrids and not in C24 hybrids could contribute to their differential response. However, this hypothesis will have to be evaluated by transgenic supplementation to establish whether these changes are causal to the incompatibility syndrome and whether the stronger suppression in Col-0 is significant.

### Critical Endosperm Factors Are Downregulated

The expression of the hybrid endosperm transcription factor that regulates development of the embryo-endosperm interface through targets, including ALE1. Loss of either ALE1 or ZHOUPI results in shriveled seed whose embryos lack a proper cuticle (Yang et al., 2008). ALE1 expression was significantly reduced in Col-0 hybrids, but not in C24 hybrids, consistent with ZHOUPI regulation. FIS2 encodes a subunit of the POLYCOMB REPRESSIVE COMPLEX (PRC). Its reduction might explain the previously reported alterations in the expression of Polycomb-regulated PHERES (Josefsson et al., 2006) and perhaps HDG9 and IKU1 misexpression (see Supplemental Figure 8C online). Interestingly, low FIS2 and ZHOUPI expression in A. arenosa could potentially explain its low expression in the hybrid. The A. arenosa expression pattern could result from its different developmental strategy in early seed development, namely, rapid endosperm proliferation coupled with delayed embryo growth (see Supplemental Figure 1 online; Bushell et al., 2003). Alternately, the mechanism governing ZHOUPI and FIS2 misexpression may also be responsible for HDG10 and IKU1 derepression. A candidate pathway is the one responsible for DNA methylation, as both FIS2 and HDG9 are imprinted via DNA methylation and maternally expressed (Julien et al., 2006; Gehring et al., 2009). The relationship between expression level and function in the seed is unknown. Reduction of the ZHOUPI target ALE1 in Col-0 hybrids and not in C24 hybrids could contribute to their differential response. However, this hypothesis will have to be evaluated by transgenic supplementation to establish whether these changes are causal to the incompatibility syndrome and whether the stronger suppression in Col-0 is significant.

### Early Deregulation of Select AGL Genes

Analysis of natural variation between Col-0 and C24 hybrids identified 1228 genes that were differentially expressed. To evaluate the contribution of natural variation to interspecific seed survival, we prioritized candidate genes by seed compartmentalization and potential role in other incompatibility syndromes. We identified 10 AGL genes, preferentially expressed in the endosperm (Figure 2)
that displayed accession-dependent expression. Loss-of-function mutations in these factors did not impact seed survival (Table 1; see Supplemental Table 8 online), although is difficult to know whether this is caused by lack of maternal effect or haploinsufficiency or because AGL31 and VRN2, which are known to affect vernalization and flowering, are not required for seed development (Ratcliffe et al., 2003; Yan et al., 2004). Other AGLs (for example AGL96 and AGL40), which display differential regulation relative to \textit{A. thaliana} crosses and polarity in compartmentalization, remain to be investigated, in particular endosperm-localized and imprinted AGL96 (Köhler et al., 2012), whose chromosomal localization is consistent with one of the incompatibility QTL described by Burkart-Waco et al. (2012) (Table 2) and appears to be deficient like FIS2 protein.

**A Defense-Like Response Involving NPR1 Mediates Hybrid Seed Death**

Defense genes are activated in response to hybridity (see Supplemental Tables 2 and 3 online). Stress-related responses have been reported previously from the analysis of somatic tissue such as leaves (Chen and Ni, 2006), but cause and effect was not clear from these observations. The genes classified in this response encompass genes encoding NBS-LRR and defensin/CRPs. A few of these proteins, including NPR1, are confirmed defense factors. Our expression analysis uncovered natural variation for these stress-related genes' expression and this variation is maintained after hybridization being evident at 3 DAP (see Supplemental Tables 5 and 6 online). Deregulation of defense response pathways contributes to seed death as suppressing defense responses, either through inactivation of the sensor NPR1 or of the salicylic acid biosynthetic enzyme encoded by SID2, increases seed survival (Figure 3B; see Supplemental Table 7 online). NPR1 and SID2 activity was deleterious for hybrid survival, but we found no evidence that this response was mediated by the negative interaction of two factors resulting in hybrid necrosis such as observed by Bomblies et al. (2007). The contribution of NPR1 to A. \textit{thaliana} × A. \textit{arenosa} seed survival is modest, consistent with the finding that at least seven QTL explain variation between C24 and Col-0 hybrids (Burkart-Waco et al., 2012).

Notwithstanding the GO classification suggesting activation of apopotic pathways (see Supplemental Tables 2 and 3 online), no induction of pathogenesis-related proteins or of bona fide apoptosis markers was detected. Studies in maize (\textit{Zea mays}) indicate that ORPs are involved in nutrient transfer between endosperm and embryo (Magnard et al., 2000; Balandin et al., 2005; Marshall et al., 2011), suggesting a similar developmental role in \textit{Arabidopsis}. It is possible that the induced NBS-LRR, including NPR1, may serve a regulatory role during seed development. NPR1 is known to interact with TTG2 in tobacco (\textit{Nicotiana tabacum}; Li et al., 2012); perhaps in \textit{Arabidopsis} it senses disrupting events in the cell and signals the maternal seed coat or endosperm to subvert development.

**TEs Do Not Undergo Widespread Activation in the Early Incompatibility Response but Are Involved in Regulatory Changes**

If epigenetic remodeling of heterochromatin leading to catastrophic activation of TEs triggers seed death, increased expression of TEs should be manifest at an early stage. We failed to find such a smoking gun. Mean and distribution of expression of TEs in controls and hybrid crosses were essentially the same (see Supplemental Tables 10 and 11 and Supplemental Figure 10 online). Alignment-independent motif analysis indicated that there is not an overrepresentation of Copia or Athila motifs in 3 DAP hybrid seed (see Supplemental Table 12 online), making it unlikely that we failed to detect changes derived from expression of the paternal genome. We identified modest differences in transposon suppression between treatments (23 TEG that were differentially expressed between hybrids but not controls), indicating that some TEs were differentially regulated after hybridization. The outcome of such divergence affected selected TEs, but it also had repercussions on gene expression because

<table>
<thead>
<tr>
<th>Table 2. QTL Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>Seed transcription factors</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Defense response genes</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AGL proteins</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Chr, chromosome.<br><sup>b</sup>Gene start position.<br><sup>c</sup>Gene end position.<br><sup>d</sup>Candidate gene based on position and mutant hybrid seed survival.<br><sup>e</sup>QTL associated with natural variation for interspecific seed between Col-0 and C24 (described in Burkart-Waco et al., 2012).
regulated genes were physically closer to TEs than expected (Figure 4).

A Summary View of Genetic and Molecular Events Observed during Postzygotic Incompatibility

Collectively, considerable information has been gained on the A. thaliana × A. arenosa postzygotic incompatibility syndrome from developmental, molecular, and genetic analyses (Comai, 2000; Bushell et al., 2003; Josefsson et al., 2006; Walia et al., 2009). Genetically, the largest effect on hybrid seed survival is mediated by maternal dosage, whereby doubling the maternal contribution rescues the cross yielding 70 to 90% live seeds. This indicates that a maternal factor or factors is deficient in the diploid cross studied here. Deficiency in this unknown factor (factors) affects endosperm 3 DAP. Possible candidates are P4-siRNAs, a 24-nucleotide small RNA class dependent on RNA Pol IV, that are contributed maternally and are candidate factors for this (Mosher et al., 2009; Lu et al., 2012) as they are associated with AGL and TE induction (Lu et al., 2012). Here, we find (Figure 5) that at 3 DAP, well before endosperm overproliferation, Athila retroelement activation, and seed shrinking, there is ZHOUPI and FIS2 downregulation. This downregulation provides a potentially attractive explanation for the shrunken embryo phenotype, which may depend on the ZHOUPI-ALE1 pathway, and for endosperm overproliferation, which may depend on AGL regulation by FIS2 (Kang et al., 2008; Walia et al., 2009). A role for FIS2, and its connection to the hypothetical loss of PRC function, is consistent with its location within the FOE incompatibility QTL (Burkart-Waco et al., 2012). An additional FOE QTL candidate is IKU1 (Table 2), a factor affecting endosperm and seed size, whose role is supported by the increased compatibility of the hybrid cross in the iku1 mutant background. The function of IKU1 in incompatibility is further supported by the comparable effect of a mutation in TTG2, a maternal factor required for function of the IKU1 pathway. The activity of the TTG2-IKU1 pathway in interspecific incompatibility is supported by loss in expression of secondary metabolite genes in the seed coat. This alteration of the expression program in the maternal tissue must result from a signal emanating from the hybrid zygotes, most likely the endosperm, and underscores the similarity to intraspecific incompatibility (Dilkes et al., 2008). The complexity of the syndrome is underscored by the function of the NPR1-SID2 pathway, the activation of defense-like responses and the parental genotype–dependent activation of AGL genes. It is possible that all these responses may be caused by a single regulatory effect.

Figure 5. Summary of Developmental and Molecular Events during Normal Development and Interspecific Hybridization.

The top part of the figure illustrates normal seed development in the early steps (days 1 to 6) after pollination. The chalazal end of the seed is organized in chalazal seed coat (maternal) and chalazal endosperm (zygotic), which are thought to form a conduit for nutrient transmission from the seed mother to the zygotes. The bottom part of the figure illustrates the main developmental and molecular changes observed in this study (3 DAP) or reported in the literature (4, 5, and 6 DAP). The results, summarized in this figure, demonstrate that regulatory disruption is evident well before endosperm proliferation and activation of Athila retroelements.

[See online article for color version of this figure.]
event, perhaps the loss of FIS2 regulation. According to this hypothesis, the incompatibility phenotype may result from PRC2-dependent disruption of downstream pathways, including TGCIKU1, NPR1-SID2, and certain AGLs. Modulation of these pathways, either through genetic manipulations or through natural variation could explain the variation in phenotype, such as observed between Col-0 and C24. Alternatively, the dosage dependent factor(s) deficient in the seed mother may have multiple independent effects that contribute to incompatibility. However, our 3 DAP transcriptome analysis found no evidence of early widespread TE activation. This is inconsistent with the hypothesis that parental imbalance has a primary and frequent effect on heterochromatin maintenance (Dilkes and Comai, 2004; Madlung et al., 2005; Martienssen, 2010; Lu et al., 2012).

In summary, our analysis of seed development in this incompatible system has revealed the effect on seed survival of disruption in early communication between the zygote(s) and the maternal tissues enveloping it. Subversion of multiple key regulators in the endosperm is accompanied by the activation of photosynthetic and ribosomal genes, and a cluster of NBS-LRR, one of which, NPR1, acts through the salicylate pathway to suppress seed survival. The decreased expression of a critical PRC2 member, FIS2, suggests future paths of investigation.

METHODS

Plant Growth

All plants were grown under long-day conditions (16 h light at 21°C and 8 h dark at 18°C) in a controlled environment facility at the University of California, Davis, CA. For time-course and expression analysis, the female Col-0 (seed stock ID CS6673) and C24 (seed stock ID CS22620) diploid ecotypes of Arabidopsis thaliana were hemizygous for a male sterility construct (Paul et al., 1992) and were gifts from Rod Scott (University of California, Davis, CA). For time-course and expression analysis, the female Col-0, n = 12 seed at 2 DAP and n = 26 seed at 3 DAP; C24, n = 10 seed at 2 DAP and n = 22 seed at 3 DAP; Col-0 hybrid, n = 8 seed at 2 DAP and n = 19 seed at 3 DAP; C24 hybrid, n = 7 seed at 2 DAP and n = 7 seed at 3 DAP). Development images were taken on a Leica DM-6000 (Leica Microsystems) with Nomarski optics using a DFC 350 FX camera (Leica). Images were processed with Leica Application Suite V3 (Leica Microsystems) and Illustrator CS4 (Adobe).

Tissue Preparation for Transcriptome Profiling

Two biological replicates were prepared for each condition for transcriptome analysis (see Supplemental Figure 4A online). Immature seeds or siliques were harvested at 2, 3, 4, 5, and 6 DAP and dissected with a hypodermic needle under a stereo-microscope. Seeds were cleared using Hoyer's medium (cholar hydrate: distilled water:glycerol [80:30:10 )), mounted, and imaged using differential interference contrast microscopy. Peripheral endosperm nuclei were imaged from up to seven optical sections through whole mounted seeds. Endosperm nuclei were counted in a section of the seed (about half the peripheral endosperm area as determined by pixel counts in Adobe Photoshop CS3). The number of nuclei was then scaled to reflect the total endosperm area, which was also determined graphically in Photoshop. Seed coat measurements, from Photoshop CS3 (Adobe) pixel counts, were made on one optical section of mounted seed (Col-0, n = 12 seed at 2 DAP and n = 26 seed at 3 DAP; C24, n = 10 seed at 2 DAP and n = 22 seed at 3 DAP; Col-0 hybrid, n = 8 seed at 2 DAP and n = 19 seed at 3 DAP; C24 hybrid, n = 7 seed at 2 DAP and n = 7 seed at 3 DAP). Development images were taken on a Leica DM-6000 (Leica Microsystems) with Nomarski optics using a DFC 350 FX camera (Leica). Images were processed with Leica Application Suite V3 (Leica Microsystems) and Illustrator CS4 (Adobe).

Illumina Sequencing

Approximately 10 μg of total RNA was processed in preparation for Illumina Sequencing using a homemade kit. All enzymatic reactions followed manufacturer's protocol unless otherwise noted, and a list of reagents is provided in Supplemental Table 14 online. mRNA was purified from total RNA using Dynabeads (Invitrogen), which purify poly(A)-containing transcripts. Then, cDNA was synthesized with random priming using Superscript III (Invitrogen) followed by heat inactivation for 5 min at 85°C. Second-strand cDNA was then synthesized as follows: We combined 5' amplification primer (primers 5'-AGAGGAGACATTGTTATACTCT-3' and 5'-GACAGGACGGCTGCTACTTC-3'), and codominant cleaved amplified polymorphic sequence analysis was used to determine the genotype at the NPR1 locus (Sha et al., 1997, 1999). The npr1-1 mutant lacks NiaI restriction site, and the npr1-7 allele lacks NiaIV as described (Cao et al., 1994; Sha et al., 1997, 1999). All other mutant lines were genotyped with standard primers as provided by the SIGNAL database (signal.salk.edu). After genotyping, unopened buds were emasculated and allowed to develop for 24 h. Mutants and controls (same ecotype as mutant line) were then pollinated with A. arenosa pollen, and seed phenotypes were assessed as previously described (Burkart-Waco et al., 2012). Each cross was performed in triplicate unless otherwise noted. Mutant effect on seed phenotype was assessed by comparison of mean between the wild type crossed to A. arenosa and mutant hybrids with the Student’s t test.

Microscopy

To score embryo and endosperm stage, siliques were harvested at 2, 3, 4, 5, and 6 DAP and dissected with a hypodermic needle under a stereo-microscope. Seeds were cleared using Hoyer’s medium (cholar hydrate: distilled water:glycerol [80:30:10 ]), mounted, and imaged using differential interference contrast microscopy. Peripheral endosperm nuclei were imaged from up to seven optical sections through whole mounted seeds. Endosperm nuclei were counted in a section of the seed (about half the peripheral endosperm area as determined by pixel counts in Adobe Photoshop CS3). The number of nuclei was then scaled to reflect the total endosperm area, which was also determined graphically in Photoshop. Seed coat measurements, from Photoshop CS3 (Adobe) pixel counts, were made on one optical section of mounted seed (Col-0, n = 12 seed at 2 DAP and n = 26 seed at 3 DAP; C24, n = 10 seed at 2 DAP and n = 22 seed at 3 DAP; Col-0 hybrid, n = 8 seed at 2 DAP and n = 19 seed at 3 DAP; C24 hybrid, n = 7 seed at 2 DAP and n = 7 seed at 3 DAP). Development images were taken on a Leica DM-6000 (Leica Microsystems) with Nomarski optics using a DFC 350 FX camera (Leica). Images were processed with Leica Application Suite V3 (Leica Microsystems) and Illustrator CS4 (Adobe).
A-base overhangs, which are needed for adaptor ligation, were added with Klenow (NEB). End repair and A-base addition were both followed by AMPure cleanup. Barcoded Illumina adapters for controls and regular Illumina adapters for controls (see Supplemental Table 15 online) were ligated at room temperature using NEB Quick Ligase. To remove adapter contamination in library sequencing, we size-selected our libraries using AMPure. Changing the amount of AMPure buffer changes fragment size selection. We used 0.1:1 (v/v), AMPure initial reaction volume to obtain fragments (cDNA insert + adaptor) greater than 300 bp and eluted in 20 µL Qiagen elution buffer. Ten microliters of each library was enriched using 15 µL Phusion 2x HF master mix (NEB), 3 µL water, and 2 µL 5 µM premixed PE primers (see Supplemental Table 15 online) with the following reaction conditions: 30 s at 98°C; 14 cycles of 10 s at 98°C, 30 s at 65°C, and 30 s at 72°C; final extension with 5 min at 72°C. Enriched libraries were purified with AMPure (0.8:1 v/v; AMPure library), and quality and quantity were assessed using the Agilent BioAnalyzer. Libraries were sequenced according to the manufacturer’s instructions. The hybrids were sequenced using Illumina’s HiSeq2000 (100-bp paired-end reads), and the controls were sequenced using Illumina’s HiSeq2000 (100-bp paired-end reads) or Illumina’s Genome Analyzer II (40-bp single-end reads).

Computational Analysis

For an overview of the bioinformatics pipeline, see Supplemental Figure 4B online.

Sequence Preprocessing

All sequence preprocessing was done with custom Python scripts available at the Comai lab wiki (http://comailab.genomecenter.ucdavis.edu/index.php/Barcoded_data_preparation_tools). We chose to process forward reads only, so as not to inflate count scores. Sequences were trimmed using four separate scripts to remove adaptors, remove barcodes, convert Illumina 1.5+ fastq to sanger, and remove ambiguous bases. Reads that were shorter than 35 bp or discarded along with sequence with base quality score lower than Phred 20.

Alignment

To assess gene expression, each biological replicate was aligned separately to both TAIR10 cDNA (available for download at ftp://ftp.Arabidopsis.org/home/tair/Sequences/blastd_datasets/TAIR10_blastsets/TAIR10_cDNA_20101214) using Burrows-Wheeler Aligner version 0.5.8c (Li and Durbin, 2009). We also aligned unmapped reads to a set of ~31K transposons annotated in the TAIR10 genomic gff file (available at http://www.Arabidopsis.org). After alignment, read counts were generated using a custom Python script, Sam Gene Counter (http://comailab.genomecenter.ucdavis.edu/index.php/Hybrid_Transcriptome). Sequences that mapped to more than one gene model of the same gene were saved and counted as one read count, while sequences that mapped to more than one gene were excluded. See Supplemental Table 1 online for mapping efficiency and Supplemental Figure 7 online for the high degree of correlation in read counts between biological replicates. Read count/locus tables for both genes and TE are available at the Gene Expression Omnibus (GSE42957).

A. arenosa Processing

A. arenosa libraries were processed as described in the alignment section. After alignment, we found a high degree of repetitive reads (reads with same start position) even if all mid-procedure quality control steps were consistent with good progress. Because of the small library size, we assumed this was because of reduced RNA complexity. To avoid count inflation, we ran an overamplification detection script (http://comailab.genomecenter.ucdavis.edu/index.php/Hybrid_Transcriptome) to remove all but one copy of repeated reads. Then, counts from the filtered FASTAQ files were generated as described above. The low quality of these libraries likely depends on some intrinsic feature of the A. arenosa seed: Multiple attempts employing different methods of RNA purification and performed by different individuals resulted in severely to moderately flawed results (as evaluated by sequencing) while virtually all other RNA sequencing libraries yielded satisfactory results.

Expression Analysis

Differential expression was assessed in R (version 2.14.1) using the R package DESeq (version 1.6.1; http://bioconductor.org/packages/release/bioc/html/DESeq.html), which is designed to test differential expression of count data from two biological replicates taking into account both technical and biological variability (Anders and Huber, 2010). A negative binomial distribution with parametric fit was used to assess differential expression based on variance estimates. All read counts were normalized to a library size factor calculated by DESeq, which takes into account the read count for each library and then scales read counts up or down to normalize between all libraries. We set a threshold of Benjamini-Hochberg (Benjamini, 1995; Benjamini and Yekutieli, 2005) adjusted P < 0.05 to call differentially expressed genes.

Paternal Gene Analysis

To assess paternal gene expression, we aligned each replicate to an A. arenosa cDNA reference using a set of 40,205 novel ESTs. The tags were assembled from Illumina reads obtained by sequencing cDNA. The cDNA was prepared from RNA isolated from flowering plants of A. arenosa using Veloet, CLC, and CAP3 (Huang and Madan, 1999). Most contigs (~28K) had predicted A. thaliana orthologs with 90% identity across >400 bp. The remaining contigs did not have a good BLAST match. All reads that did not map to A. thaliana cDNA were mapped to Ae-cDNA to maximize mapping efficiency. We found that ~5% of reads mapped to A. arenosa contigs from each condition (Col-0 = 397,885 reads; C24 = 389,359 reads; Col-0 hybrid = 736,394 reads; C24 hybrid = 1,000,406 reads) and 66% of the mapping matches were to genes that lacked an orthologous assignment. Because of the small proportion of total reads that mapped to A. arenosa contigs and the observation that A. thaliana and A. thaliana hybrids mapped to A. arenosa cDNA, we decided that A. thaliana mapping was sufficient for expression analysis.

GO and Reference Gene Set Comparisons

To identify the set of genes that are normally expressed in preglobular and globular stage seed, we looked at the set overlap between our Burrows-Wheeler Aligner alignments in comparison to genes identified by the Goldberg and Harada laboratories using the Affymetrix array (Ath1, GEO Platform GPL198). All the Ath1 annotated genes are present in Arabidopsis TAIR10 cDNA, so comparisons between Harada-Goldberg and our gene set were possible without annotation conversions. To assess biological significance, we looked at GO enrichment using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) Bioinformatics Resources version 6.7 (http://david.abcc.ncifcrf.gov/) (Huang et al., 2009a, 2009b), which is a bioinformatics tool to functionally characterize gene enrichment based on annotations and functional domains. We used (1) the full Arabidopsis gene set background gene set and (2) a set of genes expressed in both Col-0 and C24 3 DAP seed with coverage more than four reads (see Supplemental Data Set 1 online) with maximum enrichment threshold of Padj (Benjamini) < 0.05. Significant predictive protein signatures from InterPro (http://www.ebi.ac.uk/interpro) identified by DAVID were also taken into account in the
functional analysis. To assess gene coexpression during development, we use the data set available at AtGenExpress (http://www.weigelworld.org/resources/microarray/AtGenExpress) (Schmid et al., 2005) to obtain expression values during development and JMP 9.0 (SAS) statistical suite to perform multivariate analysis.

Transposon Analysis

Transposons in the A. thaliana genome are annotated in two different ways: 3900 TEGs are included in the gene set and thus listed in the TAIR10_GFF3_gene.gff file. In addition, a 31K set of TEs is classified by the notation AT0TE00000 found in the TAIR10_GFF3_genes_transposons.gff. Both GFF files are available for download from The Arabidopsis Information Resource (http://www.Arabidopsis.org).

To assess proximity of differentially expressed genes to TEs, coordinates for each gene in the Arabidopsis genome (28,496 genes) as well as TE map position (for median) were obtained from The Arabidopsis Information Resource (http://www.Arabidopsis.org) using the TAIR10 GFF file. For all annotated genes that were not TEGs, we calculated the minimum distances in base pairs between their 5’ and 3’ ends (with respect to transcription) and the closest TE (in either direction) TE. This file is available at (http://comlab.genomecenter.ucdavis.edu/index.php/Hybrid_Transcriptome). We determined distance information for seed genes, 19,820 genes out of the 20,768 having more than four reads in both Col-0 and C24 (see Supplemental Figure 5 online), and differential expressed genes between hybrid (Col-0 hybrid versus C24 hybrid) and control sets (Col-0 versus C24). The null hypothesis that mean 5’ distance and mean 3’ distance of differentially expressed gene was not closer to TE was tested with a Student t test using a random sample of seed genes with equal to the number of differentially expressed genes as a control gene set. This random sampling was replicated 10 times to reduce sampling effects; only one of the 10 samples is displayed in Figure 4.

Motif analysis was employed to determine if common or specific TE motifs (see Supplemental Table 12 online) were overrepresented in any library. Peptide motifs common to many elements were selected from alignments of Arabidopsis copia or gypsy sequences. The motifs were tested for specificity by BLASTP analysis against Arabidopsis protein. To identify motifs in the RNA sequencing data, the first 30 bases of each read, regardless of its alignment behavior, were translated into six possible frames, and the occurrence of the target motifs was scored with a custom Python script. Control motifs were derived by inverting the real ones, such as the ELVISLIVES motif producing a SEVILISVE control. The counts were normalized to account for the total number of reads in the scored library, which ranged from 9M to 23M.

Accession Numbers

Raw and processed sequence data can be found at the Gene Expression Omnibus repository (GSE42957). A list of Arabidopsis accession numbers is provided in Supplemental Table 13 online.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Comparison of A. arenosa to Hybrid and A. thaliana Embryos at 5 DAP.

Supplemental Figure 2. Phenotypic Variability of Shriveled Hybrid Seed.

Supplemental Figure 3. Peripheral Endosperm Nuclei of 3 DAP Arabidopsis Seed.

Supplemental Figure 4. Experimental Design for Differential Expression Analysis.

Supplemental Figure 5. Expression Profiles from Transcriptome Analysis of High and Low Compatibility Hybrids.

Supplemental Figure 6. Comparison of Differential Expression between Hybrids and Controls.

Supplemental Figure 7. Correlation between Read Counts in Biological Replicates.

Supplemental Figure 8. Misregulation of Endosperm Transcription Factors Posthybridization.

Supplemental Figure 9. Characterization of Natural Variation for Gene Expression.

Supplemental Figure 10. Expression Analysis of Transposable Elements during Hybridization.

Supplemental Table 1. Library Statistics.

Supplemental Table 2. Gene Ontology Enrichment of Differentially Expressed Genes between Control and Hybrids Relative to TAIR10.

Supplemental Table 3. Gene Ontology Enrichment of Differentially Expressed Genes between Control and Hybrids Relative to Seed Genes.

Supplemental Table 4. Gene Expression Level of Genes with Predictive “Cell Killing” Function.

Supplemental Table 5. Gene Ontology Enrichment of Differentially Expressed Genes in 3 DAP Seed between Col-0 and C24 Relative to TAIR10.

Supplemental Table 6. Gene Ontology Enrichment of Differentially Expressed Genes in 3 DAP Seed between Col-0 and C24 Relative to Seed Genes.

Supplemental Table 7. Interspecific Seed Phenotypes of Defense Response Mutant Lines.

Supplemental Table 8. Expression Analysis of AGAMOUS-LIKE Genes.

Supplemental Table 9. Interspecific Seed Phenotypes of Mutant AGAMOUS-LIKE Genes in T-DNA Lines.

Supplemental Table 10. Transposable Element Read Count Statistics.

Supplemental Table 11. Differential Expression Analysis of Transposable Element Genes.

Supplemental Table 12. Kmer Analysis of Transposable Element Expression (Hits/10M Reads).

Supplemental Table 13. Mutant Line Accession Number.

Supplemental Table 14. Reagents for Library Preparation.

Supplemental Table 15. Adaptor and Barcode Sequences.

Supplemental Data Set 1. Seed Genes.

Supplemental Data Set 2. Gene Ontology Data Sets.

Supplemental Data Set 3. A. arenosa Comparisons.

Supplemental Data Set 4. Shared DESeq Genes between Seeds and Siliques.

ACKNOWLEDGMENTS

Illumina sequencing was primarily performed by UC Davis Genome Center DNA Technologies Core Facility. We thank Henriette O’Geen for technical support and assistance with sequencing. We also thank
AUTHOR CONTRIBUTIONS

D.B.-W. planned and performed research, analyzed data, and wrote the article. K.N. contributed new computational tools, analyzed data, and edited the methods section. B.D. performed research and edited the article. C.J. performed research and edited the article. L.C. designed research, contributed new computational tools, analyzed data, and wrote the article.

Received December 19, 2012; revised May 28, 2013; accepted June 10, 2013; published June 28, 2013.

REFERENCES


