

Genetic variation and temperature affects hybrid barriers during interspecific hybridization

Katrine N. Bjerkan^{1,2}, Karina S. Hornslien¹, Ida M. Johannessen¹, Anders K. Krabberød¹, Yuri S. van Ekelenburg¹, Maryam Kalantarian¹, Reza Shirzadi¹, Luca Comai³, Anne K. Brysting^{1,2}, Jonathan Bramsiepe^{1,2} and Paul E. Grini^{1,*} 

¹EVOGENE, Department of Biosciences, University of Oslo, 0316, Oslo, Norway,

²CEES, Department of Biosciences, University of Oslo, 0316, Oslo, Norway, and

³Plant Biology and Genome Center, University of California, Davis, Davis, CA, 95616, USA

Received 5 August 2018; revised 31 July 2019; accepted 19 August 2019.

*For correspondence (e-mail paul.grini@ibv.uio.no).

SUMMARY

Genomic imprinting regulates parent-specific transcript dosage during seed development and is mainly confined to the endosperm. Elucidation of the function of many imprinted genes has been hampered by the lack of corresponding mutant phenotypes, and the role of imprinting is mainly associated with genome dosage regulation or allocation of resources. Disruption of imprinted genes has also been suggested to mediate endosperm-based post-zygotic hybrid barriers depending on genetic variation and gene dosage. Here, we have analyzed the conservation of a clade from the MADS-box type I class transcription factors in the closely related species *Arabidopsis arenosa*, *A. lyrata*, and *A. thaliana*, and show that *AGL36-like* genes are imprinted and maternally expressed in seeds of *Arabidopsis* species and in hybrid seeds between outbreeding species. In hybridizations between outbreeding and inbreeding species the paternally silenced allele of the *AGL36-like* gene is reactivated in the hybrid, demonstrating that also maternally expressed imprinted genes are perturbed during hybridization and that such effects on imprinted genes are specific to the species combination. Furthermore, we also demonstrate a quantitative effect of genetic diversity and temperature on the strength of the post-zygotic hybridization barrier. Markedly, a small decrease in temperature during seed development increases the survival of hybrid F1 seeds, suggesting that abiotic and genetic parameters play important roles in post-zygotic species barriers, pointing at evolutionary scenarios favoring such effects.

Keywords: Imprinting, endosperm, hybridization, post-zygotic barriers, *Arabidopsis thaliana*, *Arabidopsis arenosa*, *Arabidopsis lyrata*.

INTRODUCTION

Seed development is a sophisticated and highly regulated process that requires precise signaling events and interaction between many distinct cell types and tissues. It starts with fusion of the male and female gametes generated in the male and female gametophytes, giving rise to the embryo and endosperm that develop in parallel inside the protective seed coat. The process is initiated when a conspecific pollen grain lands on the stigma of the female reproductive organ and the pollen tube delivers two sperm cells to the female gametophyte. One sperm cell fertilizes the haploid egg cell which develops into the diploid embryo, while the other sperm cell fertilizes the homodiploid central cell generating the triploid endosperm. The endosperm is important for nutrient flow to the embryo

but also for coordinating growth of the developing seed (Nowack *et al.*, 2010).

The endosperm has two maternal genome copies and one paternal copy, and a specialized epigenetic phenomenon called genomic imprinting regulates parent-specific gene dosage during seed development, usually occurring in the endosperm (Gehring and Satyaki, 2017). Imprinting is manifested by expression of one parental allele, with concurrent silencing of the other allele. The main mechanisms for this process are DNA methylation and histone methylation (Berger *et al.*, 2006). The FERTILIZATION INDEPENDENT SEED-Polycomb Repressive Complex 2 (FIS-PRC2) mediates histone methylation while, in *A. thaliana*, DNA methylation mediated imprinting is

maintained by the DNA methyltransferase MET1 (Rodrigues and Zilberman, 2015).

A prominent gene family displaying frequent imprinting of its members is the MADS-box transcription factor (TF) family. The MADS-box TFs can be divided into type I and type II by evolutionary relationships. The type I TFs are further divided into $M\alpha$, $M\beta$, and $M\gamma$ phylogenetic subclasses and only share the highly conserved DNA-binding MADS (M) domain. The type II TFs have, in addition to the M domain, the Intervening (I), the Keratin (K) and the C-terminal (C) domains that are often referred to as the MIKC type (Parenicova *et al.*, 2003). The type II class is thought to have evolved from an ancient whole genome duplication, as orthologs are found in many other species and the genes are well distributed across all chromosomes in *A. thaliana*. The type I class TFs originate from more recent and smaller scale duplication events and, in *A. thaliana*, they are mainly concentrated on chromosomes I and V (Parenicova *et al.*, 2003; Airoidi and Davies, 2012). As a consequence, MADS-box type I orthologs are uncommon in other species (Masiero *et al.*, 2011). Imprinted genes occur frequently in the type I class, consistent with the hypothesis that recently duplicated genes are more often imprinted to regulate gene dosage (Yoshida and Kawabe, 2013). Imprinting is observed mainly in the $M\alpha$ and $M\gamma$ subgroups and, moreover, members of these two subclasses interact extensively in yeast two-hybrid assays, suggesting a common function as heterodimers (de Folter *et al.*, 2005).

Functional studies of the MADS-box type I TFs by genetic dissection, however, are hampered by genetic redundancy. Their roles have also been suggested to have restricted effect and may therefore be involved in a specific developmental processes (Nam *et al.*, 2004). Only a few type I genes have been studied phenotypically, including *AGAMOUS-LIKE (AGL) 23*, *AGL36*, *PHERES (PHE) 1 (AGL37)* and *PHE2 (AGL38)*, *DIANA (AGL61)*, *AGL62*, and *AGL80* (Köhler *et al.*, 2003; Köhler *et al.*, 2005; Bemer *et al.*, 2008; Colombo *et al.*, 2008; Kang *et al.*, 2008; Steffen *et al.*, 2008; Shirzadi *et al.*, 2011). To this end, the biological roles of many imprinted genes are still not known, but the role of imprinting is mainly associated with genome dosage regulation or allocation of resources (Haig and Westoby, 1989; Dilkes and Comai, 2004; Rodrigues and Zilberman, 2015).

Imprinting has previously been shown to be disrupted in hybrid crosses of *A. thaliana* and *A. arenosa*. The MADS-box TF *PHE1*, which is imprinted and only paternally expressed in *A. thaliana*, was upregulated in hybrid seeds and it was shown that the expressed *PHE1* was predominantly maternally expressed (Josefsson *et al.*, 2006). Disruption of the expression levels of co-adapted MADS-box TFs in hybrids may thus trigger genome-wide perturbations observed in hybrids (Roth *et al.*, 2019). Furthermore,

other MADS-box type I TFs have been shown to be highly upregulated in incompatible hybrid crosses between *A. thaliana* mothers and *A. arenosa* fathers. Using knock-out mutant lines of these genes as *A. thaliana* mother, increased viability in the incompatible hybrid seeds, suggesting that these MADS-box type I TFs partly constitute a genetic basis for the post-zygotic barrier (Walia *et al.*, 2009). Hence, investigation of the imprinting status of these genes and other known imprinted genes in *A. arenosa* and *A. lyrata* will shed light on the role and consequently the evolution of imprinting. It is disputed whether imprinting of specific genes is conserved, and whether the mechanisms behind the establishment and maintenance of imprinting between related and distant species are preserved (Waters *et al.*, 2013; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016; Chen *et al.*, 2018).

Diploid *A. arenosa* crossed as father to more than 50 accessions of *A. thaliana* displayed live seeds in the range of 1% to 30% (Burkart-Waco *et al.*, 2012). This suggests that the strength of the post-zygotic barrier can be modulated by genetic variation in accessions. Comparison of *A. arenosa* crossed to different *A. thaliana* accessions, thorough phenotyping (Burkart-Waco *et al.*, 2013) and sequencing of RNA from hybrid seeds (Burkart-Waco *et al.*, 2015), identified perturbation of the imprinting patterns of eight known paternally expressed genes. As these crosses were limited to a specific *A. arenosa* population, we hypothesize that the observed barrier is population dependent. Lafon-Placette *et al.* (2017) demonstrated that in crosses between *A. lyrata* and *A. arenosa*, the post-zygotic species barrier is due to endosperm cellularization failure. A similar study in the *Capsella* genus also indicated endosperm failure as the main seed defect in incompatible crosses (Rebernik *et al.*, 2015). A post-zygotic endosperm-based barrier has also been described for rice (Tonosaki *et al.*, 2018; Wang *et al.*, 2018) and tomato (Florez-Rueda *et al.*, 2016).

Here we have investigated the role of genetic variation in the establishment of post-zygotic endosperm-based hybrid barriers both in general, using accession and in a targeted manner, addressing specific MADS-box type I loci. We investigated the function and regulation of a conserved clade of MADS-box type I $M\gamma$ class (*AGL34*, *AGL36*, and *AGL90*) together with some of their interacting partners. To further elucidate function, we have analyzed the conservation of this clade in the closely related species *A. arenosa*, *A. lyrata*, and *A. halleri*, including the imprinting status of *AGL36-like* genes in *A. arenosa*, in *A. thaliana* crossed to *A. arenosa* and in the reciprocal cross of *A. arenosa* and *A. lyrata*. We find that *AGL36-like* genes are imprinted and maternally expressed in seeds of Arabidopsis species and in hybrid seeds between outbreeding species. In hybridizations between outbreeding and inbreeding species the paternally silenced allele of the

AGL36-like gene is reactivated in the hybrid, demonstrating that also maternally expressed imprinted genes are perturbed during hybridization and that such effects on imprinted genes are specific to the species combination.

Moreover, we investigated the role of temperature in hybridization of different genetic backgrounds and specific loci and find a significant positive correlation between lower temperatures and hybrid seed germination rate. We report that just a small change in temperature during seed development is sufficient to increase survival of hybrid F1 seeds, suggesting that abiotic parameters play an important role in post-zygotic, endosperm-based species barriers. Crossing mutants of the $M\gamma$ and $M\alpha$ clades, and their interacting partners to *A. arenosa* to further investigate the effect of these genes on the hybridization barrier identified that lack of *AGL35* significantly aggravated the *A. thaliana* *A. arenosa* hybrid barrier and that *AGL35* is involved in the temperature dependency of the hybrid barrier.

RESULTS

MADS-box type I $M\alpha$ and $M\gamma$ expression in seed development

In order to investigate the role of imprinted loci in the establishment of endosperm-based hybrid barriers, we analyzed MADS-box TFs that are closely related to the paternally silenced *AGL36* (Shirzadi *et al.*, 2011). We re-analyzed the phylogeny of the $M\alpha$ and $M\gamma$ classes and assembled them in two groups with several subclades (Figure 1a). In the $M\gamma$ group, *AGL36* constitutes a subclade together with *AGL90* and *AGL34*, and the latter genes may represent recent, local gene duplication events in *A. thaliana* as orthologs are not readily identified (Masiero *et al.*, 2011). *AGL36* and *AGL90* have been shown to be imprinted (Shirzadi *et al.*, 2011, Zhang *et al.*, 2018). The wider subclade includes known imprinted, paternally expressed genes *PHE1/AGL37* and *AGL92* (Wolff *et al.*, 2011), which cluster together with *PHE2/AGL38* and *AGL86*, respectively. *PHE2* expression has previously been demonstrated to be bi-allelic (Villar *et al.*, 2009). The most distant member of the subclade is *AGL35*, which is closely linked to and located between *AGL34* and *AGL36* on chromosome 5. *AGL34*, *AGL35*, *AGL36* and *AGL90* all map in a 100 kbp cluster on chromosome 5 (Parenicova *et al.*, 2003), and this clustering makes this subclade an especially interesting case to study evolution of imprinted genes. The function of *AGL90* and *AGL34* is not known, but *AGL36* interacts with two $M\alpha$ MADS-box TFs, *AGL28* and *AGL62* (de Folter *et al.*, 2005; Bemer *et al.*, 2010) where *AGL28* has been shown not to be imprinted (Zhang *et al.*, 2018) or to display accession dependent imprinting (Wolff *et al.*, 2011). *AGL62* is biparentally expressed and required for endosperm cellularization (Kang *et al.*, 2008), and for reason of functional study we have included *AGL62* and *AGL28* in our analysis.

Next, we investigated the relative expression of all MADS-box type I TFs at seed developmental stages ranging from 1 day after pollination (DAP) to 12 DAP (Figures 1b and S1). *AGL36* expression peaked at 4 DAP and coincided with the timing of endosperm cellularization (Shirzadi *et al.*, 2011), hence an RNA sequencing-based differential expression analysis relating all stages to 4 DAP was performed. $M\alpha$ and $M\gamma$ class TFs are overrepresented in the transcriptome of the developing seed compared with the $M\beta$ class. All 16 $M\gamma$ genes and two-thirds of the 25 $M\alpha$ genes are expressed, whereas less than half of the 21 $M\beta$ genes can be identified (Figure S1b). Ordering the $M\alpha$ and $M\gamma$ expression profiles according to the branching pattern displayed a general expression trend with a peak between 4 and 6 DAP (Figure 1b). In the $M\gamma$ *AGL36* subclade, *AGL35*, *AGL36* and *AGL90* display similar profiles with increasing or unchanged expression toward 4 DAP followed by a decline. *AGL34* can only be detected in a small developmental window, but the relative expression pattern is equivalent to *AGL36* and *AGL90* at these stages (Figures 1b and S1b). This supports findings by Zhang *et al.* (2018) and indicates that *AGL34* is not a pseudogene, as previously postulated (Bemer *et al.*, 2010). A similar pattern is found in $M\alpha$ subclades, including *AGL28*. A decrease of *AGL62* levels was observed after the expression maxima observed in $M\alpha$ and $M\gamma$ classes (Figure 1b). The $M\alpha$ *AGL62* is required for correct timing of endosperm cellularization (Kang *et al.*, 2008), and thus plays a putative role in the establishment of endosperm-based hybrid barriers (Lafon-Placette *et al.*, 2017). The $M\alpha$ MADS-box TF subclass is hypothesized to form dimers with the $M\gamma$ -type (de Folter *et al.*, 2005), and taken together, the co-occurring $M\alpha$ and $M\gamma$ expression patterns may indicate a possible role for these TFs in the establishment of cellularization-based post-zygotic hybrid barriers.

Imprinting and regulation of $M\gamma$ and $M\alpha$ MADS-box genes

Dosage imbalance caused by imprinted genes has been proposed as a cause for hybrid failure in both plants and animals (Dilkes and Comai, 2004; Wolf *et al.*, 2014; Brekke *et al.*, 2016). To this end, we successively re-analyzed parent-of-origin expression of the wider *AGL36* subclade (Figure 1a), including the interacting $M\alpha$ *AGL28* (Figure 2). Using Col-0 and Tsu-1 accession-specific single nucleotide polymorphisms (SNP), we analyzed *AGL28*, *AGL35*, *AGL36* and *AGL90* in 4 DAP Col-0 Tsu-1 hybrid seeds (Figure 2a). Maternal bias from the seed coat could be excluded, as all transcripts were previously shown to be enriched >8-fold 4 DAP in the peripheral endosperm (*AGL36*, *AGL90*) compared to all other seed tissues or >8-fold and >5-fold enriched in the chalazal endosperm (*AGL35* and *AGL28*, respectively) (Belmonte *et al.*, 2013; Hornslien *et al.*, 2019). *AGL34* was not expressed at a sufficient level in 4 DAP Tsu-1 and was thus omitted. Gene-specific RT-PCR

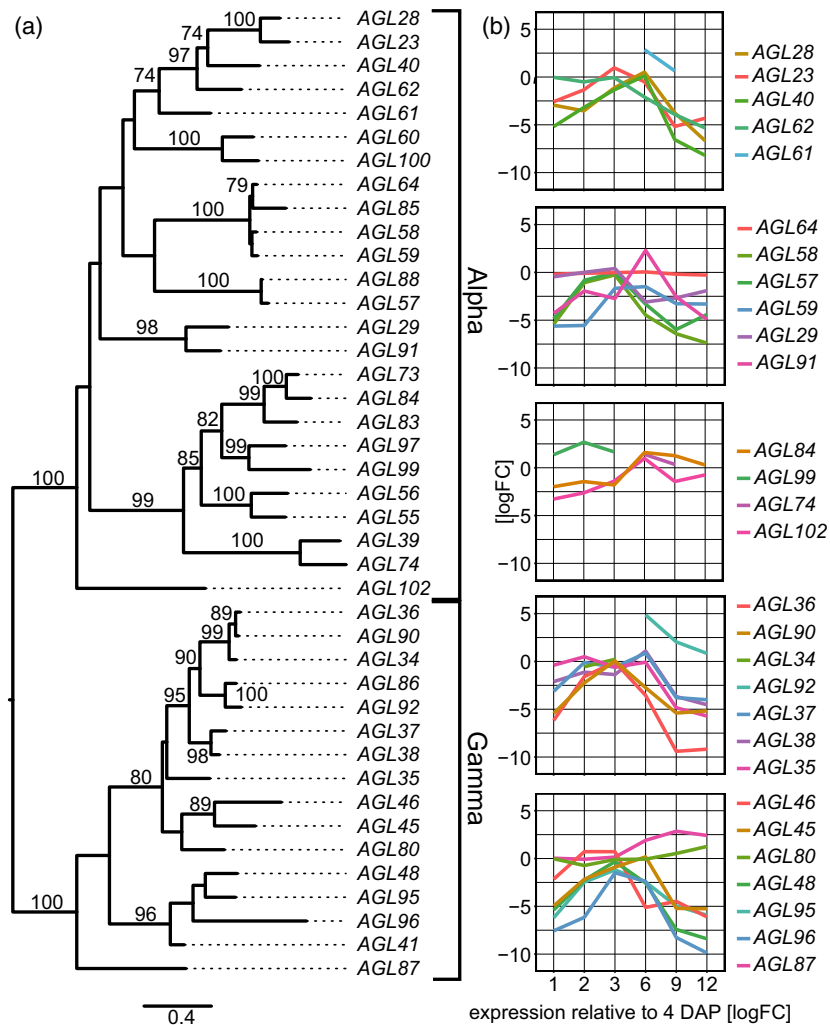
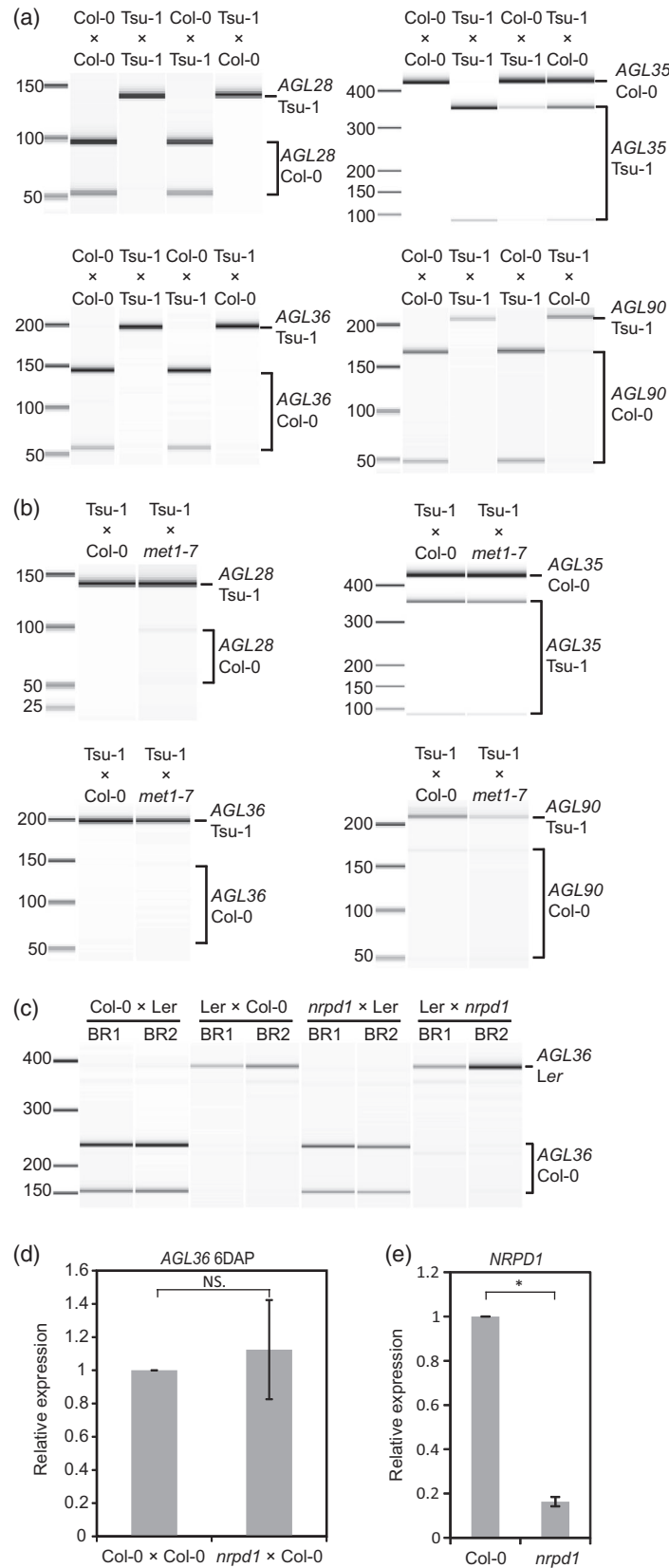


Figure 1. MADS-box type I transcription factors share similar expression profiles during seed development. (a) Maximum likelihood phylogeny of alpha and gamma MADS-box type I genes in *Arabidopsis thaliana*. The tree was inferred using the GTRGAMMA model on 41 genes with 532 unambiguously aligned nucleotides. Scale bar represents the mean number of nucleotide substitutions per site. Only bootstrap values above 65% are shown. (b) Gene expression profiles of alpha and gamma MADS-box type I genes were ordered in five groups according to the branching pattern. Transcript quantification and differential expression analysis was performed with RSEM and visualized using R. Gene expression profiles for stages ranging from one to 12 days after pollination (DAP) are relative to four DAP using a base-2 logarithmic scale (logFC). Two biological replicates with three technical replicates were analyzed. Note that genes within groups show similar gene expression profiles, with a common maximum reached between three and six DAP.

products from hybrid crosses were digested with SNP-specific restriction endonucleases (Table S1) and fragments analyzed on a Bioanalyzer 2000 as well as by Sanger sequencing (Figures 2a and S2). The M_γ genes *AGL36* and *AGL90* and the M_α *AGL28* were maternally expressed. The M_γ *AGL35* was biparentally expressed as previously reported (Zhang *et al.*, 2018). *AGL36* (Shirzadi *et al.*, 2011; Wolff *et al.*, 2011; Zhang *et al.*, 2018) and *AGL90* (Zhang *et al.*, 2018) has previously been shown to

be maternally expressed. Here, we show that *AGL90* is maternally biased in its expression but the paternal allele shows accession dependent imprinting and is not completely silenced from Col-0 pollen donors (Figures 2a and S2). A similar lack of silencing of the *AGL90* paternal Col-0 allele was recently also reported (Hornslien *et al.*, 2019). In contrast, *AGL28* was previously reported not to be maternally expressed (Zhang *et al.*, 2018), or to display accession dependent imprinting in hybrids (Wolff *et al.*, 2011).

Figure 2. Imprinting and epigenetic regulation of *AGL28*, *AGL35*, *AGL36* and *AGL90*. (a) Imprinting analysis of *AGL28*, *AGL35*, *AGL36* and *AGL90* using accession-specific restriction digest on single nucleotide polymorphisms (SNPs) in reciprocal crosses between accessions Tsu-1 and Col-0. For each panel, the accession-specific digestion pattern is indicated. Seeds were harvested for analysis 4 days after pollination (DAP). Bioanalyzer images of one of three biological replicates is shown. (b) SNP analysis of *AGL28*, *AGL35*, *AGL36* and *AGL90* in crosses with *met1-7+/-* (Col-0 background) pollen. Only *AGL28* displays paternal activation in the *met1-7+/-* mutant. Wild type (WT) crosses are duplicated from (a) for visualization. Crosses were harvested as (a). Bioanalyzer images of one of three biological replicates is shown. (c) SNP analysis of *AGL36* in reciprocal crosses between WT (Col-0) or *nripd1* (Col-0) and WT (*Le*r). Crosses were harvested six DAP. The *AGL36* imprinting pattern is not changed in the *nripd1* crosses compared with WT crosses. (d) Real-time PCR analysis of *AGL36* expression in three biological replicates the *nripd1* background at six DAP. The relative expression difference is not significant (NS, *t*-test; $P = 0.389$). (e) Real-time PCR verifying significant knock-down (*t*-test; $P = 0.015$, indicated by asterisk) of *NRPD1* in three biological replicates of *nripd1* homozygous background. Error bar indicates standard deviation (SD).



To address the regulation of imprinted genes we contrasted 4 DAP parental seed expression from crosses between Tsu-1 and Col-0 accessions versus Tsu-1 and a hemizygous *met1-7+/-* in a Col-0 background to determine if MET1 is involved in maintaining silencing of the paternal copy of *AGL28*, *AGL36* and *AGL90* (Figure 2b, Table S1, Figure S2). The paternal copy of *AGL28* was shown to be expressed using *met1-7^{+/+}* as pollen donor, suggesting that MET1 is required for silencing of the paternal *AGL28* allele. In our experimental settings, however, lack of *MET1* did not reactivate the paternal copy of *AGL36* and *AGL90*, whereas the *AGL35* biparental control remains unchanged (Figure 2b). In contrast with these findings, the paternal allele of *AGL36* was previously shown to be reactivated in crosses with homozygous and hemizygous *MET1* mutant pollen using the *met1-4+/-* allele (Saze *et al.*, 2003; Shirzadi *et al.*, 2011). Lack of *MET1*, both in homozygous and heterozygous mutants, leads to DNA hypomethylation and eventually the accumulation of epimutations; we attributed the previously observed paternal expression to such effects in the *met1-4+/-* background. The *met1-7+/-* allele used in our study was kept hemizygous through repeated outcrosses and therefore more likely devoid of such effects (Hornslien *et al.*, 2019).

According to a report investigating the role of small interfering (si) RNA and RdDM in interploidy crosses, several MADS-box type I genes, including *AGL36*, were deregulated in diploid crosses with NUCLEAR RNA POLYMERASE D1 (NRPD1) mutant mothers, deficient in the largest subunit of RNA polymerase IV, a key component of canonical RdDM (Lu *et al.*, 2012). Using *nuclear rna polymerase d1 (nrpd1)* mutant mothers, *AGL36* was upregulated more than 20-fold (Lu *et al.*, 2012). We have previously shown that imprinting of *AGL36* do not require paternal DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) or ARGONAUTE4 (AGO4), both part of the RdDM pathway (Shirzadi *et al.*, 2011). The data from Lu *et al.* (2012) suggested that an RdDM-dependent mechanism maintained the expression level of *AGL36* or is active in maintaining the silencing of the paternal *AGL36* allele after fertilization. To test the latter hypothesis, we analyzed parental expression from 6 DAP seeds, using *nrpd1* both as a maternal and paternal contributor in crosses to wild type (Figures 2c and S2). The *AGL36* imprinting pattern was not affected in any cross direction, suggesting that reactivation of the paternal allele is not causing elevated levels of *AGL36*. In contrast with the previous report (Lu *et al.*, 2012), we could also not detect any significant upregulation of *AGL36* by real-time PCR (Figure 2d) in crosses using a homozygous *nrpd1* knock-out allele (Figure 2e) as maternal cross partner. We concluded that neither the MET1 nor the PolIV RdDM pathway is sufficient to silence the paternal allele of *AGL36*.

Next, we analyzed the effect of PRC2 on MADS-box type I genes. The endosperm cellularization defect observed in

Arabidopsis interspecies hybrid seeds is highly reminiscent to the failure of endosperm cellularization phenotype observed in mutants of *FIS-PRC2* (Lafon-Placette *et al.*, 2017). We therefore compared the RNA-seq relative expression of all MADS-box type I genes between a *FIS-PRC2* mutant and wild type at seed developmental stages ranging from one DAP to 12 DAP (Figures 3a and S3). Clustering of transcript profiles revealed four main patterns of regulation, ranging from highly regulated to no effect (Figure S3). Overall, the M β class as a group was significantly less regulated by the PRC2 *medea (mea)* mutation than the M α and M γ classes, and also displayed the least variation (Figure 3b). This is in accordance with previous observations, that the M β class TF are mainly expressed at low levels or in female gametophytic stages (Bemer *et al.*, 2010).

In the deregulated classes of transcript profiles, consisting of mainly but not exclusively M α and M γ , the wild type profiles are generally characterized by increasing expression that decreases after a peak (Figure S3, left panels). In the *mea* cross, both this pattern and the peak are shifted toward higher expression levels and later developmental stages. In certain cases, decrease is not observed within the analyzed developmental time-frame (Figure S3, right panels). A distinct shift in transcript profiles could also be observed between the profile clusters in *mea*, with one class derepressed before 6 DAP, while a second class started at 9 DAP and the third class at 12 DAP (Figure 3a, three top clusters). M γ dominates the two former classes of transcript profiles together with M α whereas the latter constitutes of M α and M β genes. In a recent report, Zhang *et al.* (2018) analyzed MADS-box type I deregulation in a *swinger (swi) mea* double mutant. These authors identified two major expression clusters (C1 and C2) based on difference in temporal expression patterns both in the wild type and in the PRC2 double mutant. The latter cluster was distinguished by the an upregulation of the expression pattern in the mutant and could be further divided in two clusters (C2.1 and C2.2) based on the timing of downregulation in the wild type (Zhang *et al.*, 2018). The three clusters described in our study (Figure 3a, three top clusters) are well in line with the real-time-PCR-based study of Zhang *et al.* (2018). Eight out of 12 genes in the C2.1 cluster are also found in our top cluster, starting deregulation at the earliest stage (Figure 3a, top cluster), whereas three genes are found in our second cluster (Figure 3a, second top cluster) together with all *mea swi* upregulated genes identified in the C2.2 cluster. This also includes *AGL91*, *AGL49* and, importantly, *AGL34* that are upregulated in our study, whereas no upregulation was identified by Zhang *et al.* (2018). In contrast, we could not detect any upregulation for *AGL64*, as reported by the other study (Zhang *et al.*, 2018).

We conclude that the *AGL36* subclade, including *AGL34* and *AGL90*, as well as the *AGL36* interacting M α *AGL28* are commonly repressed by MEDEA from 4-6 DAP. The *AGL36*

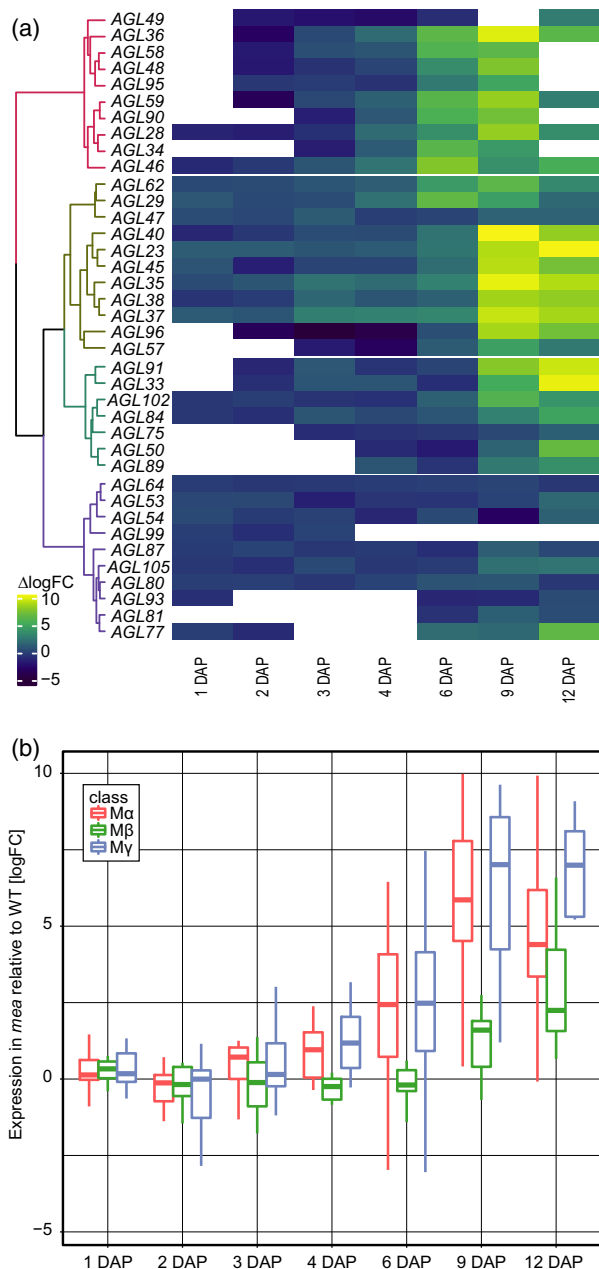


Figure 3. PRC2 dependent transcriptional repression of MADS-box type I transcription factors during seed development. (a) Heat-map clustering of gene expression profiles of MADS-box type I transcription factors (TFs) in the Polycomb Repressive Complex 2 (PRC2) *MEDEA* mutant seeds (*mea*) compared to wild type (WT). Expression profile-based clustering, transcript quantification and differential expression analysis was performed for three biological replicates using RSEM and visualized using R. Differential gene expression profiles for stages ranging from one to 12 days after pollination (DAP) are shown using a base-2 logarithmic scale (logFC). Note strong expression and differential change in three out of four expression clusters. (b) Box plot showing class specific expression of MADS-box type I TFs at stages from 1 to 12 DAP. The MEA-dependent repression is shared by the alpha ($M\alpha$) and gamma ($M\gamma$) class TFs, while the beta ($M\beta$) class is weakly affected in the *mea* background. Relative expression changes are displayed in a base-2 logarithmic scale. Points indicate outliers.

and also the AGL90 interacting $M\alpha$ AGL62 are upregulated in *mea* at 9 DAP, in accordance with the role of AGL62 in endosperm cellularization (Kang *et al.*, 2008).

Conservation of Arabidopsis AGL36-like imprinting in hybrid seeds is species dependent

Having analyzed the expression and regulation of MADS-box type I genes in *A. thaliana*, we turned our focus to the expression and role of these genes in the genus Arabidopsis. MADS-box type I genes are often less conserved between model species. For instance, no orthologs of AGL36 were identified in rice or maize (Masiero *et al.*, 2011). AGL36-like genes can be found when analyzing more closely related species such as in the genus Arabidopsis (Figures 4a and S4). Two genomic loci of AGL36-like genes were identified in *A. arenosa* by assembling online resources (see Experimental procedures). Both were verified in various individuals from two *A. arenosa* populations using PCR amplification (MJ09-1, MJ09-4) (Jørgensen *et al.*, 2011; Lafon-Placette *et al.*, 2017). The two genes differ in length (1050 bp versus 1008 bp). The shorter gene does not have a continuous open reading frame and most likely harbors an intron based on two open reading frames spaced by an 88-bp sequence. The 1050-bp locus, but not the 1008-bp AGL36 locus, was confirmed to produce a transcript in 9 DAP seeds, corresponding to the globular-embryo seed stage. Online genome sequencing resources of *A. lyrata* subsp. *lyrata* suggest one AGL36-like gene (Figure 4a). In the subspecies *A. lyrata* subsp. *petrea*, two loci have been indicated (Yoshida and Kawabe, 2013), but by performing Sanger sequencing from the *A. lyrata* subsp. *petrea* population MJ09-11 (Jørgensen *et al.*, 2011) combined with online resources, we concluded that *A. lyrata* contains only one AGL36-like locus (Figures 4a and S4). Notably, although two AGL36-like loci are present in *A. arenosa* and *A. halleri*, our analysis indicates that the two duplication events creating the AGL36 subclade (AGL34, AGL36, AGL90) do not exist outside *A. thaliana*.

To analyze imprinting of *A. arenosa* AGL36-like (*AaAGL36-like*), we screened natural populations (MJ09-4 and MJ09-1) for SNPs that could be used to distinguish the parental alleles. We identified one individual that had a SNP in *AaAGL36-like* (I, cf. Figure 4a) that also allowed SNP detection with restriction enzymes (Table S1c). Seed RNA was harvested from reciprocal crosses at 9 DAP followed by RT-PCR of the SNP-containing regions from *AaAGL36-like* (Table S2). The PCR products were digested with SNP-specific enzymes (Table S1c) and fragments analyzed (Figure 4b). Only maternal expression was found, suggesting that *A. arenosa* AGL36-like (*AaAGL36* I) is an imprinted maternally expressed gene.

Next, we analyzed AGL36-like imprinting in hybrids of *A. arenosa* and *A. lyrata*. Amplifying AGL36-like (Table S1c)

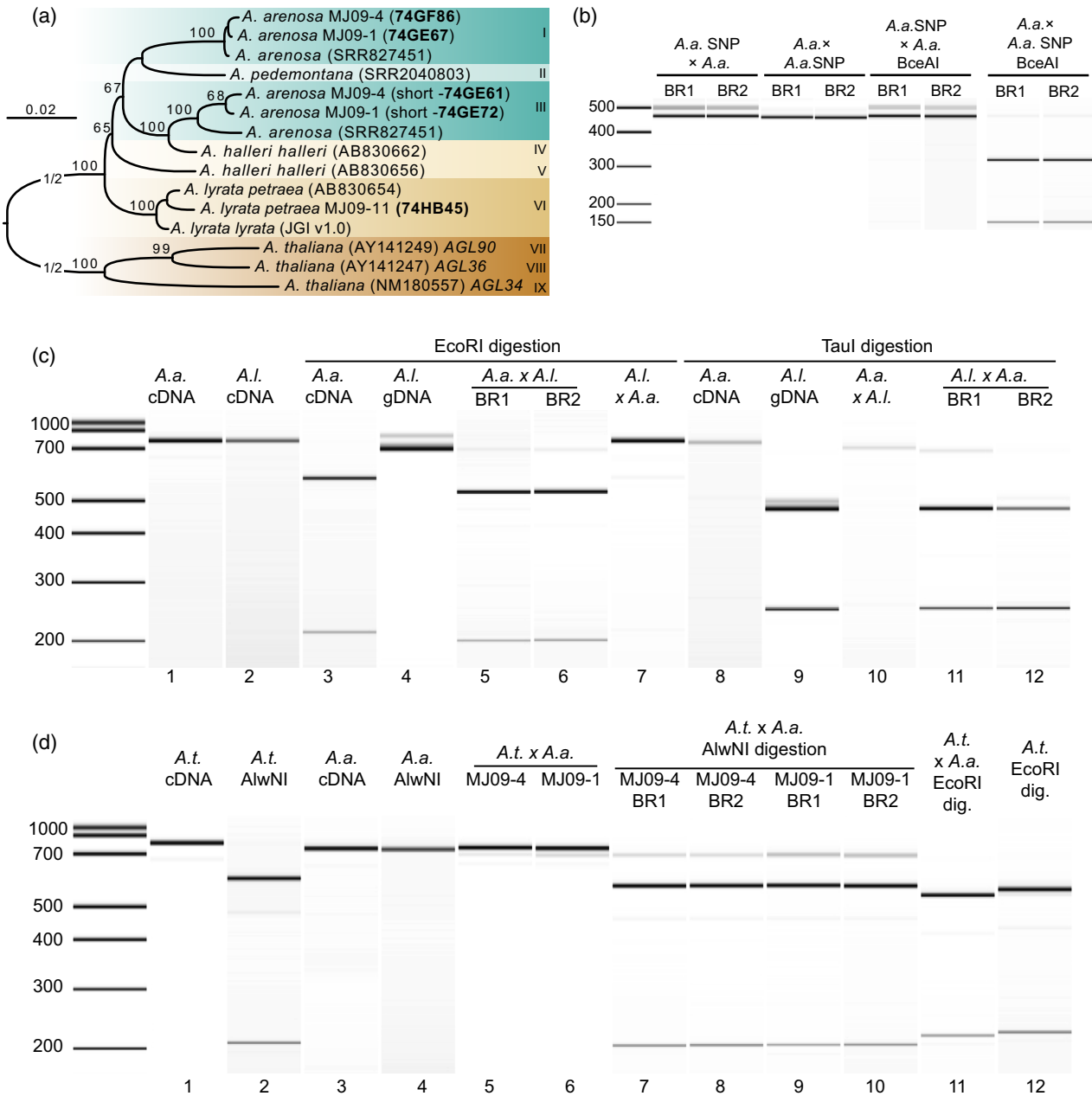


Figure 4. Conservation and imprinting of *AGL36*-like genes in *Arabidopsis* and in selected hybrid crosses. (a) Maximum likelihood phylogeny of *AGL36*-like genes in *A. arenosa*, *A. pedemontana*, *A. halleri*, *A. lyrata* and *A. thaliana*. The tree was inferred using the GTRGAMMA model on 15 sequences with 938 unambiguously aligned nucleotides. Because of space limitations branches marked ½ are shortened to half their original length. Only bootstrap values above 65% are shown. Scale bar represents the mean number of nucleotide substitutions per site. *Arabidopsis* species are indicated by colors. Roman numerals (right) indicate distinct genes in the respective species and is used as reference in (b–d). (b) SNP analysis of *A. arenosa* *AGL36* (I) in MJ09-4 background at 9 days after pollination (DAP). Left half, undigested; right half, digested with *Bce*Al. Cross plants in which *AaAGL36* is digested by *Bce*Al are indicated ‘A.a.’. Cross plants in which *AaAGL36* is not digested by *Bce*Al are indicated ‘A.a. SNP’. The shorter 475-bp fragment is *AaAGL36*. *AaAGL36* (I) is maternally expressed in *A. arenosa* as only the undigested *AGL36* fragment is present when *A.a.* SNP is the mother, whereas completely digested fragments result when *A.a.* is the mother. (c) Imprinting and maternal expression of *AGL36* is conserved in reciprocal crosses between *A. arenosa* and *A. lyrata* at 9 DAP. *AaAGL36* (I) and *AlAGL36* (VI) fragments are both 768 bp of length (lanes 1–2). *Eco*RI digests *AaAGL36* (I) only (lanes 3 and 4) and *Taul* digests *AlAGL36* (VI) only (lanes 8 and 9). In hybrid crosses, only the maternal fragments are digested, and no or very weak undigested fragments are left (lanes 5–7, 10–12). (d) Imprinting of *AGL36* is lifted in crosses between *A. thaliana* and *A. arenosa*. The undigested fragments of *AtAGL36* (VIII) and *AaAGL36* (I) are 819 and 768 bp, respectively (lanes 1 and 3). An *Alw*NI restriction site is only present in *AtAGL36* (VIII) (lanes 2 and 4). In hybrid crosses, both fragments are visible (lanes 5 and 6) and *Alw*NI restriction digest only the maternal *AtAGL36* fragment. Note that the paternal *AaAGL36* (I) allele is expressed (lanes 7–10). *Eco*RI digests both *AaAGL36* and *AtAGL36* leaving only digested fragments (lanes 11 and 12). BR, biological replicate; *A.t.*, *A. thaliana*; *A.a.*, *A. arenosa*; *A.l.*, *A. lyrata*. All crosses shown are in the order female × male. Two biological replicates represent results from three biological replicates tested. DAP stages used are four for *A.t.*, seven for *A.t.* × *A.a.* and 9 DAP for *A.a.*, *A.l.* and the *A.a.* × *A.l.* reciprocal cross.

from cDNA of reciprocal crosses of *A. arenosa* and *A. lyrata* resulted in one fragment because *AGL36-like* (I and VI, respectively in Figure 4a) from both species are the same length (Figure 4c). When *A. arenosa* is crossed as mother to *A. lyrata*, *AaAGL36-like* is successfully digested by *EcoRI* (Figure 4c, Table S1c). Using *A. lyrata* as a mother crossed to *A. arenosa*, no *EcoRI* digestion fragments occur and thus only expression of *AIAGL36-like* was observed (Figure 4c). As a reciprocal control, we also used *A. lyrata*-specific digestion by *TauI*, giving only digestion fragments in crosses with *A. lyrata* mothers (Figure 4c). To verify the identity of the amplified fragments and the maternal expression pattern, the undigested PCR products were Sanger sequenced and identified as only maternally contributed (Figure S5). In conclusion, only the maternal allele is expressed in reciprocal interspecies crosses between *A. arenosa* and *A. lyrata* indicating that imprinting of *AGL36-like* is preserved in *A. arenosa* × *A. lyrata* hybrid seeds.

Finally, we investigated *AGL36/AGL36-like* imprinting in *A. thaliana* × *A. arenosa* hybrid seeds, using *A. arenosa* as the paternal cross partner. Upon amplification of cDNA, both maternally expressed *AtAGL36* and paternally expressed *AaAGL36-like* fragments could be identified (Figure 4d, lanes 5–6, see figure legend). Restriction with *AlwNI* digest the maternally contributed cDNA (Figure 4d and Table S1c). The paternal *AaAGL36-like* fragments remained undigested in hybrid crosses employing two independent *A. arenosa* populations (Figure 4d, lanes 7–10). The paternal cDNA was verified by Sanger sequencing and comparison with the *A. arenosa* control (Figure 4d). We thus conclude that in hybrid *A. thaliana* × *A. arenosa* seeds, and in contrast with hybrid *A. lyrata* × *A. arenosa* seeds, the silenced paternal *A. arenosa* allele is reactivated, demonstrating differential action by the maternal species in the hybrid. Furthermore, this finding demonstrates that the paternal alleles of maternally expressed imprinted genes are deregulated in hybrid crosses, and not limited to paternally expressed genes as described in previous studies (Josefsson *et al.*, 2006; Burkart-Waco *et al.*, 2015).

Genetic and environmental factors influence post-zygotic hybrid barriers

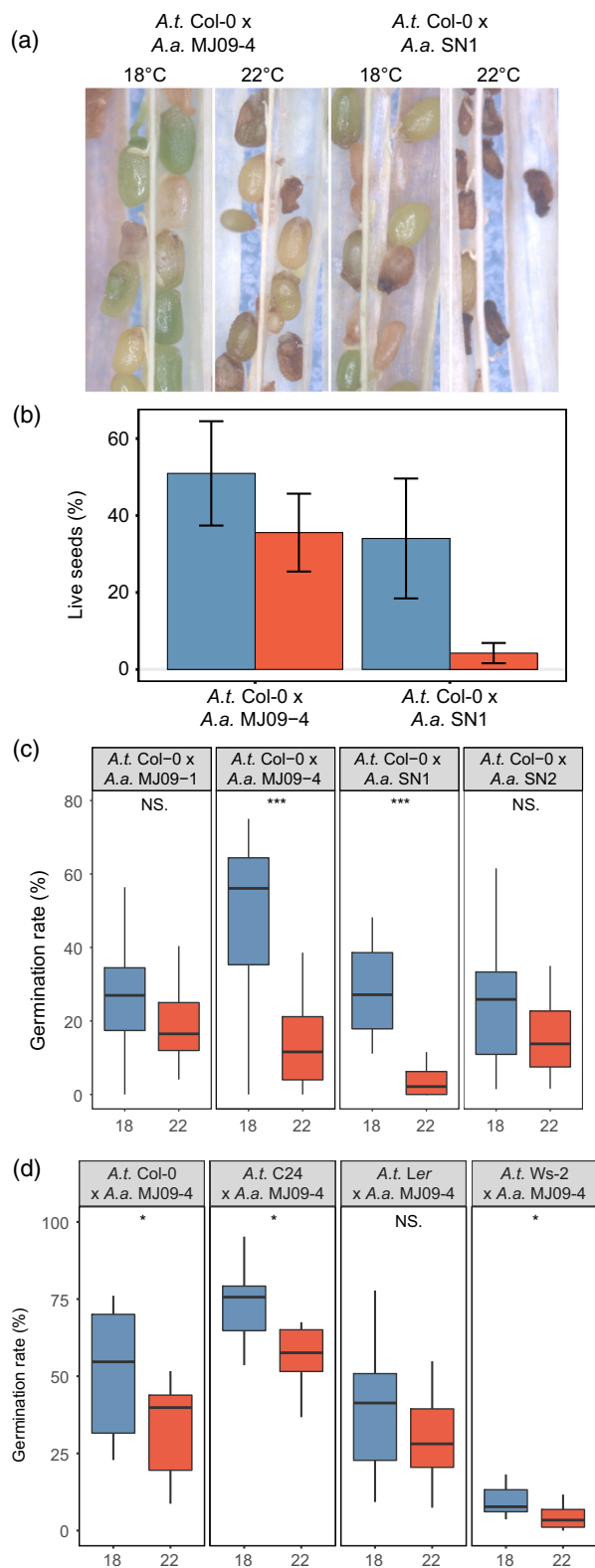
Deregulation of MADS-box type I TF genes has been implicated in setting up the post-zygotic hybridization barrier in incompatible hybrid *A. thaliana* × *A. arenosa* (Strecno1) seeds, and *A. thaliana* mutation of some of these genes could partially rescue the severe late seed phenotype in the same hybrid cross from approximately 1 to 10% (Walia *et al.*, 2009). In order to systematically examine if the MADS-box type I TF genes analyzed in this work influence the strength of the hybrid barrier we noted that previous analyses to investigate hybrid barriers in Arabidopsis have been performed under slightly different temperature regimes (Josefsson *et al.*, 2006; Walia *et al.*, 2009; Burkart-

Waco *et al.*, 2012; Lafon-Placette *et al.*, 2017) and in line with this, the major hypotheses to explain hybrid barriers are centered on genetic factors, not taking environmental variation into account. In rice, however, it has been demonstrated that temperature affects cellularization of the endosperm (Folsom *et al.*, 2014) and that type I MADS-box TF genes are deregulated during moderate heat stress (Chen *et al.*, 2016). To rule out a temperature effect in our experiments, we therefore repeated crosses first performed by Walia *et al.* (2009) with Strecno1 and Col-0 using the original temperature regime (22°C) and at slightly lower temperatures (18°C) and also included the *A. arenosa* accession used in this study, MJ09-4. Surprisingly, both the difference in temperature and genetic variation between *A. arenosa* populations had a major effect on the strength of the post-zygotic hybrid barrier (Figure 5a).

To quantify this observation, we first investigated seed survival in the same crosses and temperature conditions. Using *A. arenosa* accessions MJ09-4 and Strecno1 (SN1) in crosses to *A. thaliana* Col-0 at both 18°C and 22°C, a substantial increase in the survival of hybrid seeds at 18°C for both accessions was observed (Figure 5b) with 18°C MJ09-4 replicates showing up to 60% live seeds while at the same time also obtaining the same results as Walia *et al.*, 2009 when crossing Col-0 with Strecno1 at 22°C (live seed count 1%, $N = 162$) (Figure 5b).

In hybrid seed germination experiments, the temperature dependency of the strength of the hybrid barrier became even more evident for the accessions MJ09-4 and SN1 (Figure 5c; $P < 0.001$). Interestingly, when comparing two other *A. arenosa* accessions, MJ09-1 and SN2, in crosses to Col-0, these were found to be insensitive to the temperature change tested here although they display a higher seed survival rate than SN1 crossed to Col-0 (Figure 5c). These accessions may still be affected at larger differences in temperature due to genetic variation and different adaptation. Even though the variation between replicates is high, especially in 18°C crosses, a clear bypass effect of low temperature on the post-zygotic barrier is observed. Furthermore, germination of both 18°C and 22°C crosses demonstrated an increased germination rate of hybrid seeds involving the *A. arenosa* MJ09-4 population as a paternal cross partner compared with crosses with *A. arenosa* SN1, SN2 or MJ09-1 (Figure 5c). Control interspecies crosses in *A. thaliana* accessions and *A. arenosa* MJ09-4 at 18°C and 22°C displayed no significant difference in germination between the temperatures (Figure S6). Furthermore, tetraploid *A. thaliana* mothers have been shown to alleviate the hybridization barrier and to exclude this scenario we verified the diploidy of *A. thaliana* Col-0 accessions, the *A. arenosa* MJ09-4 population and the Strecno lines (SN1 and SN2) using flow cytometry (Figure S7).

Our findings indicated that genetic variation between *A. arenosa* populations also influences the success rate of hybridization, as previously demonstrated for different *A.*



thaliana genotypes (Burkart-Waco *et al.*, 2012). We therefore further investigated the temperature dependency of the hybrid barrier by varying the maternal *A. thaliana*

Figure 5. Temperature has a significant effect on the hybrid barrier between *A. thaliana* and *A. arenosa*. (a) Micrographs of siliques with *A. thaliana* crossed to *A. arenosa* hybrid F1 seeds grown at 18°C and 22°C 20 days after pollination (DAP). The crosses were made using two different *A. arenosa* lines, MJ09-4 and Strečno1 (SN1). Live seeds are green, collapsed seeds are brown or pale green. (b) Graph showing percentage live *A. thaliana* × *A. arenosa* hybrid F1 seeds from crosses in (a). Three biological replicates were tested for each temperature for both *A. arenosa* lines MJ09-4 and SN1 ($N = 174, 163, 175, 162$, respectively). (c) Germination rate of *A. thaliana* × *A. arenosa* hybrid F1 seeds. Four *A. arenosa* accessions were crossed to *A. thaliana* at 18°C and 22°C, MJ09-1, MJ09-4, SN1 and SN2 (18°C: MJ09-1 $N = 18$ BR (931 seeds), MJ09-4 $N = 18$ BR (986 seeds), SN1 $N = 9$ BR (524 seeds) and SN2 $N = 9$ BR (475 seeds). 22°C: MJ09-1 $N = 12$ BR (612 seeds), MJ09-4 $N = 36$ BR (1482 seeds), SN1 $N = 36$ BR (1544 seeds), SN2 = 12 BR (673 seeds)). (d) Germination rate of *A. thaliana* Col-0, Ler, C24 and Ws-2 crossed to *A. arenosa* MJ09-4 at 18°C and 22°C (18°C: Col-0 $N = 12$ BR (572 seeds), C24 $N = 8$ BR (451 seeds), Ler $N = 12$ BR (772 seeds), Ws-2 $N = 12$ BR (622 seeds). 22°C: Col-0 $N = 24$ BR (1212 seeds), C24 $N = 12$ BR (462 seeds), Ler $N = 12$ BR (751 seeds), Ws-2 $N = 12$ BR (727 seeds)). A.t., *A. thaliana*; A.a., *A. arenosa*. Blue color: 18°C, red color: 22°C. Outliers are plotted as large points. Dots indicate single BR. BR, biological replicate. Significance is indicated for the comparison of lines at 18°C and 22°C (Wilcoxon rank-sum test: NS: $P > 0.05$; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$). Error bar indicates standard deviation (SD).

accession used in the hybrid cross. Burkart-Waco *et al.* (2012) crossed 56 accessions using the Strečno1 line at 22°C to investigate the effect of the genetic variation on the hybrid barrier and could demonstrate a weaker barrier when using C24, producing 17% normal seeds, while using Ler-1 and Ws-2 resulted in 5.2% and 3.5% normal seeds, respectively. Using Col-0 they obtained 1.7% normal seeds. Here, we demonstrate that crossing Col-0, Ler-1, C24 and WS-2 accessions to the *A. arenosa* accession MJ09-4 give the effect of elevated seed survival for all accessions except Ws-2 (Burkart-Waco *et al.*, 2012) when comparing to the previous report using SN1 at 22°C (Figure 5d). In addition, the accessions Col-0, C24 and Ws-2 have a significant increase in seed survival when decreasing the temperature to 18°C (Figure 5d; $P < 0.05$). Ler-1 appeared to be insensitive to the temperature change, similar to the observations using *A. arenosa* accessions MJ09-1 and SN2 in combination with Col-0 (Figure 5c).

Previously published data reported that the embryo does not make the transition to the heart stage in crosses between diploid *A. thaliana* Col-0 and *A. arenosa* Strečno1 at 22°C (Burkart-Waco *et al.*, 2013), which we also could confirm for Strečno1 under our laboratory conditions at 22°C (Figure 5). Our analysis at 18°C, however, demonstrated that most seeds develop past this point (Figure 6a–i). There was a clear correlation between the severity of the hybrid barrier in *A. arenosa* accessions and the timing of endosperm cellularization in hybrid seeds (Figure 6a–i) suggesting that endosperm cellularization is the major mechanism for setting up the barrier.

The role of MADS type I loci in the *A. thaliana* × *A. arenosa* hybrid barrier

In order to investigate the specific effect of selected MADS type I loci in establishing or bridging the *A. thaliana* × *A.*

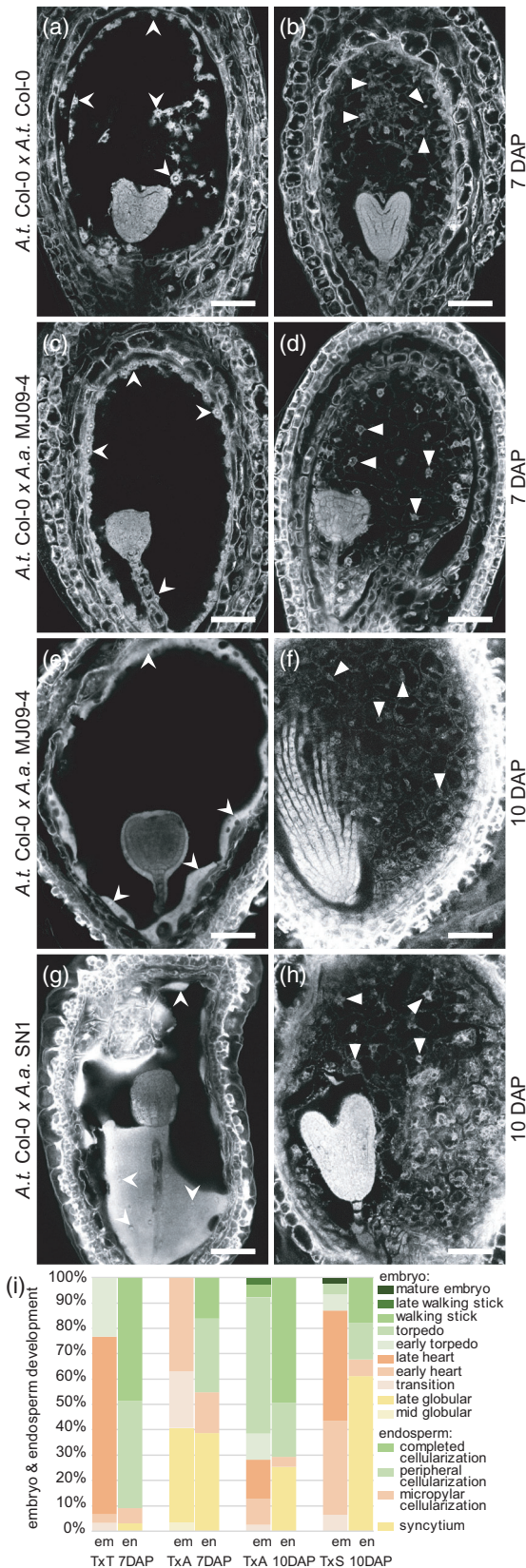


Figure 6. Variation in endosperm cellularization between *A. thaliana* and *A. arenosa* hybrids. (a–h) Confocal scanning laser micrographs of endosperm cellularization in hybrid seeds visualized by Feulgen staining. For all crosses, both non-cellularized and cellularized endosperm is observed and micrographs representative for each class are presented in the left and right panels respectively. Open arrowheads point to syncytial endosperm nuclei while closed arrowheads point to cellularized endosperm nuclei. Scale bar = 50 μ m. (a and b) *A. thaliana* control 7 days after pollination (DAP) typically at the embryo late heart stage in which most seeds display complete endosperm cellularization (b). (c, d) *A. thaliana* \times *A. arenosa* MJ09 hybrid seeds at 7 DAP. Embryo development is slower compared with *A. thaliana* controls. Both non-cellularized (c) and cellularized endosperm (d) was frequently observed. (e, f) *A. thaliana* \times *A. arenosa* MJ09 hybrid seeds at 10 DAP. Only a few seeds fail to cellularize (e) and most seeds exhibit completed endosperm cellularization (f). (g, h) *A. thaliana* \times *A. arenosa* SN1 hybrid seeds at 10 DAP. A higher fraction of seeds display syncytial stage endosperm (g) compared with *A. arenosa* MJ09 hybrid seeds (e, f), but some have completed endosperm differentiation (h). (i) Quantification of the described embryo and endosperm stages. All crosses are indicated as female \times male. T \times T, *A. thaliana* seeds, $N = 34$; T \times A, *A. thaliana* \times *A. arenosa* MJ09 hybrid seeds, $N = 81$; T \times S, *A. thaliana* \times *A. arenosa* SN1 hybrid seeds, $N = 98$; em, embryo stages; en, endosperm stages.

arenosa hybrid barrier, we analyzed insertional mutant alleles of the selected candidate genes (Figure S8a and Table S3). Homozygous mutants could be obtained for all investigated loci except as previously described for *AGL62* (Kang *et al.*, 2008), suggesting no vital requirement in seed development. Significantly reduced transcript levels were demonstrated in all lines with the exception of *AGL34* in which transcript levels were significantly elevated (Figure S8b). Segregation analysis could not detect reduced transmission of the mutant alleles suggestive of a requirement in male or female gametophytes or a recessive effect in embryo or endosperm (Figure S8c and Table S4). We also inspected seed size, seed germination and flowering time. For this analysis, *agl28-1* was omitted due to mixed Ws-2 Col-0 accession background (Yoo *et al.*, 2006). Only minor differences were observed in seed size and flowering time (Figure S9a,b) and no difference in germination of mutant seeds was observed (97–100%, $N = 200$). Finally, seed developmental phenotypes in single and higher order mutants were investigated, scoring live, aborted and unfertilized seeds (Figure S9c). Notably, a heterozygous *agl62-1* mutation in a double homozygous *agl28-1 agl36-1* background did not differ from single *agl62* mutants. We concluded that a thorough analysis of seed development in single, double and triple mutants of *AGL34*, *AGL35*, *AGL36* and *AGL90* including their interaction partners *AGL28* and *AGL62* did not result in any obvious seed developmental phenotypes (Figures S8 and S9).

In the case of *agl28-1* a mixed Ws-2 Col-0 accession background did not allow a direct comparison of hybrid seed barrier strength effects, due to the strong effect of the Ws-2 accession (Figure 5d, right panel). Indeed, single *agl28-1* mutants as well as double or triple mutant combinations with *agl36*, *agl62* and *agl90* crossed with *A. arenosa* all produced significantly lower seed germination

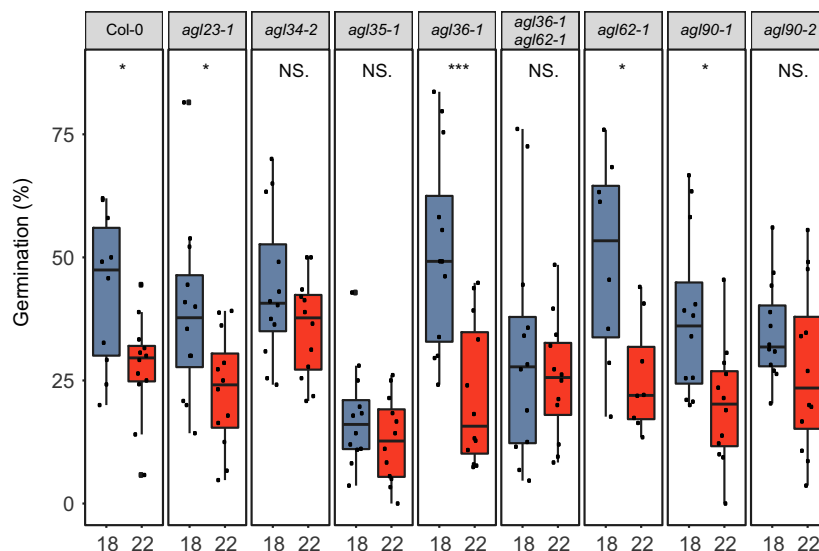


Figure 7. Genetic and environmental parameters influence the F1 hybrid barrier. (a) Germination rate of seeds from *A. arenosa* MJ09-4 crossed as pollen to the *A. thaliana* (Col-0), single mutants *agl23-1*, *agl34-2*, *agl35-1*, *agl36-1*, *agl62-1*, *agl90-1*, and *agl90-2* and the double mutant *agl36-1 agl62-1* at 18°C and 22°C. Box plot contains scattered dots representing germination rates observed per silique. Outliers are plotted as large points. Significance is indicated for the comparison of the mutant lines between 18°C and 22°C (Kruskal–Wallis test: $P < 2.2 \times 10^{-16}$; Wilcoxon rank-sum test: NS: $P > 0.05$; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$). 18°C: Col-0 $N = 10$ BR (536 seeds), *agl23-1* $N = 12$ BR (294 seeds), *agl34-2* $N = 12$ BR (707 seeds), *agl35-1* $N = 12$ BR (641 seeds), *agl36-1* $N = 12$ BR (704 seeds), *agl36-1 agl62-1* $N = 12$ BR (532 seeds), *agl62-1* $N = 8$ BR (442 seeds), *agl90-1* $N = 12$ BR (568 seeds), *agl90-2* $N = 12$ BR (753 seeds). 22°C: Col-0 $N = 12$ BR (578 seeds), *agl23-1* $N = 12$ BR (407 seeds), *agl34-2* $N = 12$ BR (610 seeds), *agl35-1* $N = 12$ BR (571 seeds), *agl36-1* $N = 12$ BR (635 seeds), *agl36-1 agl62-1* $N = 12$ BR (498 seeds), *agl62-1* $N = 8$ BR (403 seeds), *agl90-1* $N = 12$ BR (514 seeds), *agl90-2* $N = 12$ BR (635 seeds). BR, biological replicate.

rates in nine out of 10 cross combinations (Figure S10a). In a direct comparison using *Ws-2* wild type as the control, compared with Col-0 and *agl28-1* when crossed with *A. arenosa* MJ09-4, revealed no difference in seed germination rate between *Ws-2* and *agl28-1* (Figure S10b, left). In order to further investigate the role of *agl28-1* we generated an *AGL28* genomic rescue construct that was transformed into the homozygous *agl28-1* mutant background. Six homozygous T2 lines were crossed to *A. arenosa* MJ09-4 and scored for germination (Figure S10b, right). None of the rescue lines was significantly different from *agl28-1* or the *Ws-2* control, suggesting that the observed reduced germination is caused by the *Ws-2* background alone. As the *Ws-2* background effect co-segregated with the *agl28* mutation through repeated introgression to Col-0, we hypothesized a major part of the genetic variation causing the strong *Ws-2* *A. arenosa* hybrid barrier to be linked to the *AGL28* locus. In order to genetically map the effect, we backcrossed a Col-0 introgressed hemizygous *agl28-1* to Col-0 and genotyped the progeny for the presence of *agl28-1*, and crossed the two types of plants resulting with *A. arenosa* MJ09-4 pollen donors (Figure S10c, see legend for detail). Plants wild type for the *AGL28* locus had a high probability to be Col-0 in the *AGL28* region, and did also regain Col-0 germination rates in crosses with *A. arenosa* (Figure S10c; NS, not significantly different). In addition, plants hemizygous for *agl28-1* having a high probability to be *Ws-2* Col-0 heterozygous in the *AGL28*

region, displayed intermediate germination levels and were still significantly different from Col-0 (Figure S10c; $P < 0.05$). This strongly suggested that the strong *Ws-2* effect on the hybrid barrier was linked to the *AGL28* region on top of chromosome 1.

Finally, various single and double *A. thaliana* MADS-box type I mutants were crossed with the *A. arenosa* MJ09-4 population at the two temperatures established (18°C and 22°C). First, we wanted to investigate the influence of the mutated loci on the strength of the hybrid barrier, as measured by germination rate. Secondly, we wanted to explore if the identified temperature effect on the strength of the hybrid barrier was modulated by the mutated loci, as observed using in both *A. thaliana* and *A. arenosa* accessions (Figure 5c, d).

Hybrid seed phenotypes were inspected for some representative crosses at 18°C (Figure S11a–c). We observed the seed classes normal non-collapsed seeds, collapsed brown seeds and viviparous seeds (Figure S11a). The frequency of non-collapsed seeds in a silique and seed germination frequency were well correlated (Figure S11b). Seed size measurements also displayed variation but no strong effect of the mutant mother background (Figure S11c). The ploidies of *A. thaliana* × *A. arenosa* hybrids with both Col-0 and *agl36-1 agl90-2* as mother were verified using flow cytometry (Figure S7).

In general, none of the single or higher order *A. thaliana* mutants had a significant effect to alleviate the hybrid

barrier when crossed to the *A. arenosa* MJ09-4 accession (Figure 7). This is conflicting with previous data reported for *agl62* and *agl90* crossed with an *A. arenosa* Strecno1 population at 22°C (Walia *et al.*, 2009). In the Walia study, mutants of *agl62* and *agl90* used as mothers in the cross increased the germination frequencies from 1% in crosses with Col-0 to 10% in crosses with *agl62* and *agl90*. In our experiment, the average germination frequencies of *agl62-1* at 22°C, and the average germination frequencies of *agl34-2* at 22°C were indeed slightly higher than the Col-0 control crossed to *A. arenosa*, but the difference is not significant. We could therefore not reproduce the findings of Walia *et al.* (2009) in crosses using the *A. arenosa* MJ09-4 accession. The reason for this discrepancy may be genetic differences in the *A. arenosa* accessions used (MJ09-4 in this study versus Strecno1 in Walia *et al.*, 2009). As demonstrated here, different *A. arenosa* accessions can modulate the strength of the hybrid barrier (Figure 5c).

We did however see a significant effect of the single mutant *agl35-1* to aggravate the hybrid barrier when crossed to the *A. arenosa* MJ09-4 accession (Figure 7). Significant reduction was observed at both temperatures tested. When the homozygous *agl35-1* mutant was crossed to *A. arenosa* at 18°C, a significant reduction from average values of close to 50% germination to less than 20% was observed (Figure 7; $P < 0.001$), and in the 22°C experiment we found a reduction from more than 25% to close to 15% (Figure 7; $P < 0.01$). These findings suggest that AGL35 may play a role to relieve and bypass the hybrid barrier, or that lack of AGL35 disrupts or lowers the threshold for disruption of endosperm cellularization.

It is noteworthy that the effect of lower temperature to alleviate the hybrid barrier is also bypassed by mutation of AGL35. We analyzed if any of the loci investigated played a role in establishing the observed temperature effect on the hybrid barrier strength (Figure 7). The Col-0 control and most mutant crosses, including *agl23*, *agl36*, *agl62* and *agl90*, displayed significant differences in germination frequencies between the two temperatures examined (Figure 7). Crosses with the single mutants *agl34* and *agl35* and the *agl36 agl62* double mutant, however, displayed no significant differences in its germination rate between 18° and 22°C (Figure 7), similar to the previous observation in the *Ler-1* accession cross (Figure 5d). The temperature immunity seen in *agl35* appears to be the most prominent due to low variation in the observations, and may suggest that the AGL35 links the hybrid block to the temperature effect.

DISCUSSION

The role and regulation of imprinted genes

In this report, we have systematically analyzed the function and conservation of imprinting a subset of MADS-box type

I TFs in hybrid crosses in the context of biological and environmental variance. In our expression analysis, we observed that $M\alpha$ and $M\gamma$ class type I TFs are highly represented in the transcriptome of the developing seed, and that both classes contain correlated transcript profiles that peak around the onset of endosperm cellularization. The expression peak occurs at a crucial developmental time point when the endosperm switches from nutrient sink to source for the developing embryo (Lafon-Placette and Köhler, 2016), which is suggestive of a function in this process. Defects in endosperm cellularization are also the mechanistic basis for a post-zygotic reproductive barrier between Arabidopsis species and thus support a role for these genes in a hybrid scenario. The $M\alpha$ and $M\gamma$ class co-regulation is consistent with the notion that $M\alpha$ forms dimers with the $M\gamma$ -type MADS-box TF class (de Folter *et al.*, 2005). Furthermore, since imprinted genes interact with biparentally expressed genes this would favor the dosage hypothesis for the selection of imprinted genes (Dilkes and Comai, 2004). According to this hypothesis imprinting is a means to optimize the expression level of a gene, opposed by the parental conflict theory, in which parental conflict over resources leads to the selection of genes that promotes or restricts resource allocation (Haig and Westoby, 1991).

It has been postulated that maternally expressed imprinted genes are regulated by the release of DNA methylation in the central cell by central cell specific action of the DNA glycosylase DEMETER (DME). We could, however, demonstrate that the lack of DNA methylation maintenance in the pollen germline does not activate all silenced paternal alleles. There is a discrepancy between the lack of activation of paternal AGL36 observed in this work and previously published data (Shirzadi *et al.*, 2011). However, the two studies were done using different accessions and also different *met1* mutant alleles. Accession-specific effects are less likely as we see paternal reactivation of AGL28 using the same accessions. However, as the history of zygosity is not known for the *met1* allele used in the previous study, accumulated hypomethylation may explain the inconsistency. Nevertheless, even though the paternal allele of AGL28 reactivated in our study, paternal silencing of AGL90 and AGL36 was not lifted. In this study, *met1-7* was held as heterozygous, thus avoiding parental demethylation. This suggests that maternally expressed imprinted genes may be regulated by other mechanisms or have different regulatory requirements.

Evolution of silencing of the paternal allele of maternally expressed imprinted genes has been associated with global methylation patterns generated toward suppressing transposons (Kim and Zilberman, 2014; Anderson and Springer, 2018). *De novo* methylation of transposons is mainly performed by the action of RdDM, and in the lack of a mechanistic scenario for the imprinting of maternally

expressed genes this pathway is an obvious candidate. Several $M\alpha$ and $M\gamma$ class type I TFs including *AGL36* have also been suggested to be upregulated in the lack of RdDM (Lu *et al.*, 2012). For *AGL36* we could, however, not observe any change in paternal silencing in reciprocal crosses with a PollV mutant (*nprp1*) that is blocking canonical RdDM. Furthermore, we were also unable to verify the previously reported upregulation of *AGL36* in the same crosses and time points (Lu *et al.*, 2012). To determine the global effect, however, a systematic elucidation of the role of RdDM in the regulation of maternally expressed imprinted genes is required.

Although it does not regulate imprinting of maternally expressed genes, we showed that repression by PRC2 is specifically targeted toward $M\alpha$ and $M\gamma$ class type I TFs and acts to repress gene activity concurrent with and at post-cellularization stages. In the absence of MEA, only the maternal allele of *AGL36* is upregulated (Shirzadi *et al.*, 2011), indicating that the mode of regulation does not interfere with the actual imprinting mechanism. We hypothesize that DNA methylation of the paternal allele protects from PRC2 repression, as suggested for the paternally expressed imprinted gene *PHE1* (*AGL37*) (Köhler *et al.*, 2003; Makarevich *et al.*, 2008; Villar *et al.*, 2009). In such a scenario the paternal allele of paternally expressed genes should be targeted by PRC2 at cellularization, but the mechanism here remains to be explored. We observed, however, that $M\alpha$ and $M\gamma$ class type I TFs are regulated by PRC2 in distinct clusters from 6 to 12 DAP and that maternally and paternally expressed genes are co-regulated in the same cluster. While the meaning of the observed gradual repression by FIS-PRC2 in the context of $M\alpha$ and $M\gamma$ role in seed development remains open to speculation, it seems clear that FIS-PRC2 acts through specific sets of genes at successive stages in seed development, rather than exerting a global effect in the seed concurrent with endosperm cellularization.

Species-dependent deregulation of imprinting

We identified orthologs of *AGL36* in *A. arenosa* and *A. lyrata* and studied imprinting in *A. arenosa* and hybrids of *A. arenosa*, *A. lyrata* and *A. thaliana*. *AGL36* is imprinted in *A. thaliana* and maternally expressed, and here we prove that one of the two *AGL36-like* genes in *A. arenosa* is also exclusively maternally expressed, demonstrating conservation of imprinting. The maternal allele of the paternally expressed imprinted genes have previously been shown to be reactivated in the hybrid of *A. thaliana* and *A. arenosa* (Josefsson *et al.*, 2006; Walia *et al.*, 2009; Burkart-Waco *et al.*, 2015). We show that, in hybrid *A. thaliana* \times *A. arenosa* seeds, the paternal *AGL36-like* allele is reactivated, suggesting that deregulation in hybrid crosses is not limited to paternally expressed genes as described in previous studies (Walia *et al.*, 2009). This loss of regulation is in

strong contrast with the *A. lyrata* \times *A. arenosa* hybrid cross in which the imprinting of *AGL36-like* is maintained.

The variation or perturbation of the pattern of imprinting might play a role in the endosperm hybridization barrier between species (Florez-Rueda *et al.*, 2016), and in the case described here, different mechanisms may act in the same species depending on the cross partner. Previous imprinting analyses involved crosses between inbreeders and outbreeders, fueling speculation that the mechanisms regulating imprinted genes may differ depending on the mating system (Josefsson *et al.*, 2006; Burkart-Waco *et al.*, 2012; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016). However, intraspecific variation of imprinting within *A. thaliana* has previously been demonstrated (Waters *et al.*, 2013; Pignatta *et al.*, 2014). Consistent with the notion that imprinting can vary independently of mating systems, reactivation of the normally imprinted paternal *A. arenosa* *AGL36* allele depended on the maternal species: it was observed only in crosses to *A. thaliana* but not to *A. lyrata*.

Temperature- and population-dependent quantitative barrier phenotype

We have identified an important role of temperature in the establishment of endosperm-based post-zygotic hybrid barriers. This opens for speculation and a multifaceted scenario emerges identifying several parameters; both intrinsic variation in genetic pathways in both parents and external abiotic factors such as temperature may act in concert to mediate the generation of post-zygotic species barriers.

The discovery that lowering the temperature by only 4°C from 22°C to 18°C during the fertilization and development of the *A. thaliana* \times *A. arenosa* hybrid seed increases survival was surprising. Nevertheless, incompatibility between diverged individuals can be sensitive to temperature as, for example, in seedling hybrid necrosis (Bomblies and Weigel, 2007). In this case, however, appropriate temperatures can ameliorate acute incompatibilities manifested only during early seed development and, therefore, may play a significant role in reducing interspecific hybridization barriers. High environmental sensitivity may also explain inconsistencies with previous studies noted in Results.

Crossing *A. thaliana* wild type Col-0 and mutants to different *A. arenosa* accessions demonstrated variation depending on the pollen donor and no significant increase in seed viability could be observed by using the mutants of the MADS-box type I TFs as mothers. The variability caused by natural variation has been demonstrated in *A. thaliana* by using different accessions. Burkart-Waco *et al.* (2012) used the diploid *A. arenosa* accession Strecno-1 and crossed it as the pollen donor to 56 *A. thaliana* accessions and scored seed viability ranging from 30% normal seeds to close to 2%. Burkart-Waco *et al.* (2013) compared seed

development of *A. thaliana* accessions Col-0 and C24 crossed to *A. arenosa*, which respectively produced 0 to 1% and ~17% live hybrid seeds. Hybrid embryos at all stages beyond 2 DAP were developmentally delayed and similar between Col-0 and C24, with the exception of a few C24 × *A. arenosa* hybrid embryos displaying developmental progression to heart stage by 6 DAP, whereas no Col-0 embryos made this transition. This is in clear contrast with our findings where most seeds survive this stage. Using different *A. thaliana* accessions in crosses to the *A. arenosa* accession MJ09-4, we obtained highest seed germination rates in crosses with C24. However, the cross combination with the Ws-2 accession produced similarly low germination rates as the *A. arenosa* accession Strecno-1 crossed with Col-0. This indicates that rather than a specific accession having specific effects, it is the combination of accessions that determines the strength of the hybrid barrier. Although we could not observe elevated seed viability when single or multiple MADS-box type I mutants were used in crosses to *A. arenosa*, deficiency of *AGL35* resulted in significant reduction of the germination frequency. Notably, mutants of *AGL35* crossed to *A. arenosa* display two effects: first, it increase lethality; second, it decrease or eliminate the temperature effect, suggesting a critical role for this gene in mediating the strength of the hybrid block and the temperature effect.

Weakening of hybridization barriers at lower temperatures might increase fitness of a self-incompatible colonizer by broadening mate choice: few viable seeds are better than no seed. This mechanism could, for instance, have been important for recolonization after the Pleistocene glaciations, a period that was characterized by secondary contact and high amount of hybridization with or without genome duplication (Stebbins, 1984; Brochmann *et al.*, 2004). Although *A. thaliana* is a self-crosser, the mechanism might be ancestral and predate *A. thaliana*. Its occurrence should be investigated by hybridizing more species under varying temperature. In northern Europe and the Fennoscandian region, *A. thaliana* with unreduced gametes has most likely hybridized with pollen from tetraploid *A. arenosa* to create the allotetraploid *A. suecica* on multiple hybridization events (Novikova *et al.*, 2017). Making a synthetic tetraploid *A. thaliana* and crossing it to *A. arenosa*, can make viable, although unstable, hybrids (Comai *et al.*, 2000). Environmental stress such as heat or cold may increase the rate of unreduced gamete formation (De Storme and Mason, 2014) and facilitate hybridization between diploids and tetraploids (Vallejo-Marin and Hiscock, 2016). Such instances of genome doubling, however, did not occur in our experiments as flow cytometry of sampled hybrids indicated genome content consistent with reduced gametes of the diploid parents *A. thaliana* and *A. arenosa*. Formation of 2N gametes may be under different regulation in Arabidopsis. The temperature-sensitive mechanism affecting the endosperm-

based barrier and its dependency on AGL function remain an open area of investigation.

In rice, it has been shown that temperature affects cellularization of the endosperm (Folsom *et al.*, 2014) and that type I MADS-box TF genes are deregulated during moderate heat stress (Chen *et al.*, 2016). The temperature stress tested were much higher than tested here for Arabidopsis, but the different species have different temperature adaptations in general. The rice MADS-box TF OsMADS87 is a heat-sensitive imprinted gene which is associated with syncytial stage endosperm and regulates rice seed size (Chen *et al.*, 2016). OsMADS87 is a putative ortholog of Arabidopsis *PHE1*. Mutants of OsMADS87 have accelerated endosperm cellularization and lower sensitivity to a moderate heat stress in terms of seed size (Folsom *et al.*, 2014; Chen *et al.*, 2016). Our results indicated that some of the *Arabidopsis thaliana* MADS-box mutants hybridized to *A. arenosa*, display a lowered heat sensitivity that may indicate a temperature-sensing role for the MADS-box genes in the endosperm of *A. thaliana* as well. Considering that many of the type I MADS-box genes are regulated by the PRC2 complex and that the finding that OsFIE1 is imprinted and temperature sensitive in rice seeds (Folsom *et al.*, 2014), this proposes an epigenetic regulation during hybrid seed development which is altered during environmental perturbations (Folsom *et al.*, 2014). Given the importance of overcoming post-zygotic isolation for the early stages of neo-hybridization (Vallejo-Marin and Hiscock, 2016), this temperature effect can be a useful tool for investigating the endosperm-based post-zygotic barrier and early speciation.

EXPERIMENTAL PROCEDURES

Plant material and cultivation

A. thaliana accessions and mutant plant lines were obtained from the Nottingham Arabidopsis Stock Center (NASC) unless specified otherwise. For details on MADS-box type I mutant lines see Table S1. The *met1-7* and *nrrpd1* accession numbers are SALK_076522 and SALK_083051, respectively. The *mea/fis1* mutant was kindly supplied and described in Chaudhury *et al.* (1997). The *A. arenosa* populations MJ09-1 and MJ09-4 and the *A. lyrata* MJ09-11 descended from natural populations in central Europe as described, respectively, by Jørgensen *et al.* (2011) and Lafon-Placette *et al.* (2017). *A. arenosa* populations Strecno1 (SN1) and Strecno2 (SN2) were kindly supplied by Kirsten Bomblies (Hollister *et al.*, 2012). Seeds were surface sterilized either by washing steps with 70% ethanol, 20% bleach and wash solution (0.001% Tween20) or by over-night chlorine gas sterilization (Lindsey *et al.*, 2017), sown out on 0.5 Murashige and Skoog (MS) plates (Murashige and Skoog, 1962) supplemented with 2% sucrose and appropriate antibiotics for selection of mutant lines. The seeds were then stratified over-night (*A. thaliana*) or 1 to 3 weeks (*A. arenosa* and *A. lyrata*) at 4°C before transferring to growth chambers with either 18°C or 22°C under long-day conditions (16 h light, 160 μmol/m²/sec, relative humidity 60–65%). *A. thaliana* × *A. arenosa* F1 hybrid seeds were stratified at 4°C for 4–6 days before being placed in growth chambers for scoring of germination as seedling root protruding from the seed regardless

of survival at later stages. Germinated seedlings were transferred to soil and grown under long-day conditions at 18°C or 22°C. *A. arenosa* and *A. lyrata* plants were vernalized at 8°C under short-day conditions (10 h light) for 4–5 weeks to stimulate flowering. To avoid self-pollination, controlled crosses were performed by emasculating unopened flower buds followed by hand pollination after 2 days. Developing or mature seeds were harvested for designated purposes at defined time points. See Table S5 for an overview of interspecies crosses. For flowering time analysis, flowering time was scored as day after stratification and the average number of leaves at stem emergence from the rosette.

Tissue handling, DNA and RNA extraction and cDNA synthesis

Tissue was harvested directly in liquid nitrogen and DNA was isolated using E.Z.N.A. Plant DNA kit (Omega) according to manufacturer's instructions. For total RNA isolation, seeds were hand dissected from siliques directly into pre-chilled tubes with MagNA Lyser Green Beads (Roche) and ground in lysis buffer (Sigma Plant total RNA kit) using a MagNA Lyser Instrument (Roche). Isolated RNA was treated with DNase I (Sigma) and cDNA synthesized with oligo(dT) and Superscript III reverse transcriptase (Invitrogen). Samples were cleaned using a QIAquick PCR purification kit (Qiagen). DNA or RNA concentration was measured using a NanoDrop1000 spectrophotometer or RNA was measured with a Qubit 3 fluorometer (ThermoFisher) using the Qubit RNA BR Assay kit (Invitrogen). All kits were used according to the manufacturers' instructions.

RNA Sequencing and sequence analysis

Total RNA was isolated from dissected seeds at 1, 2, 3, 4, 6, 9 or 12 DAP from *Ler* crossed to Col-0 and from *meal/fis1* crossed to Col-0 in two biological replicates as described previously (Shirzadi *et al.*, 2011). RNA samples were DNase treated before quality checked using an Agilent 2100 Bioanalyzer. Total RNA was prepared to a strand-specific TruSeq™ RNA-seq library and all 28 samples sequenced over three lanes on an Illumina HiSeq 4000, 150 bp paired end reads. Differential expression analysis was performed with RSEM (Li and Dewey, 2011) using the edgeR software package (McCarthy *et al.*, 2012). The expression profiles were analyzed and visualized using the Tidyverse 1.2.1, ComplexHeatmap 1.17.1 (Gu *et al.*, 2016), Dendextend 1.7.0 (Galili, 2015), Viridis 0.5.0 packages in R version 3.4.3.

Molecular cloning and genotyping

All T-DNA mutant lines were genotyped using specific primers (Tables S1 and S2). Due to high sequence similarities between MADS-box genes and AGL36-like genes in different species, primers were optimized to ensure specific amplification and fragments were sequenced for confirmation of identity. The sequencing and characterization of AGL36-like genes from MJ09 *A. arenosa* and *A. lyrata* lines were performed by PCR amplification with designated primers (Table S2) using KOD Hot Start DNA polymerase (Sigma) according to the manufacturer's instructions with 1.5 mM of MgSO₄ and PCR program with 95°C denaturation, 55°C annealing, and 70°C extension for 35 cycles. The amplified fragment was subsequently cloned into a TOPO Blunt pCR Zero vector (Invitrogen) according to the manufacturers' instructions. The AGL28 genomic complementation construct was created by nested PCR using primers described in Table S2 containing *att* sites for Gateway cloning (Invitrogen) according to the manufacturer's instructions. The genomic AGL28 fragment was 3531 bp

including 2000 bp upstream of the start codon and 500 bp downstream of the stop codon and was cloned into the destination vector pMDC99. The construct was transformed into *Agrobacterium tumefaciens* strain GV3101 pMP90RK which was used to transform *agl28-1* mutant using the floral dip method (Clough and Bent, 1998) and transformants were selected for by the appropriate resistance encoded in the inserted T-DNA, hygromycin. All sequences from *A. arenosa* and *A. lyrata* generated for this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/>) with accession numbers MN380433 to MN380437.

Phylogenetic analysis

All alpha and gamma MADS-box genes were extracted from the *A. thaliana* genome (TAIR10 at <https://www.arabidopsis.org/>). The genomic sequence was translated into amino acids with the AUGUSTUS gene prediction program (Stanke and Morgenstern, 2005) using the *A. thaliana* gene model. Nucleotides from the coding regions were aligned based on the protein sequence with PAL2NAL (Suyama *et al.*, 2006). Positions with more than 80% gaps and ambiguously aligned positions were removed from the alignment. A phylogenetic tree was inferred from the resulting alignment using the GTRGAMMA model and the automatic bootstrapping criteria MRE (option -I autoMRE) in RAxML v8.0.26 (Pattengale *et al.*, 2011; Stamatakis, 2014). The genes AGL34, AGL36 and AGL90 from *A. thaliana* were used as queries in BLAST against the non-redundant nucleotide database at NCBI (blast.ncbi.nlm.nih.gov/) to find all homologous genes in the Arabidopsis genus. In addition, all available Arabidopsis Illumina whole genome sequence data from Sequence Read Archive (SRA) was employed in the phylogenetic analysis. These libraries were assembled with rna-spades (Bankevich *et al.*, 2012) and AGL36 related genes were identified with BLAST. All AGL36 related genes were aligned and phylogenetic trees were inferred as for the MADS-box genes. In the final analysis only one copy of the gene was kept for each species.

Real-time quantitative PCR

Real-time PCR was performed on a LightCycler 96 instrument using FastStart Essential DNA Probes Master protocol (Roche) and FastStart Essential DNA Green Master protocol (Roche) using designated primers (Table S2). Relative expressions were calculated according to Pfaffl (2001) and are average values of at least two biological replicates. In reactions with low relative expression an E-value of 2.0 was used. All PCR products were sequenced to verify identity of the product amplified.

Single nucleotide polymorphism analysis

Seed tissue was sampled at 4 DAP from reciprocal crosses from *A. thaliana*, at 7 DAP from crosses between *A. thaliana* and *A. arenosa*, and at 9 DAP from reciprocal crosses of *A. arenosa* and *A. lyrata*. cDNA was amplified by PCR using designated primers (Table S2) and digested using SNP-specific enzymes (Table S4) analyzed on an Agilent 2100 Bioanalyzer using the DNA-1000-LabOnChip system (Agilent Technologies). Images for figures were assembled using Illustrator software (Adobe).

Dry seed phenotyping using ImageJ

For *A. thaliana* *A. arenosa* hybrid crosses, mature seeds were harvested one silique at a time before imaging using a Nikon D90 and analysis using ImageJ to determine seed size (mm²). Viviparous seeds were excluded from the analysis. MADS-box type I mutants were grown and harvested at the same time and

conditions. Seed size was measured and analyzed using ImageJ to determine the average seed size (mm^2) per plant. The significance differences between plant lines were tested using the Kruskal–Wallis test: $P = 0.0047$ and in pairwise comparison with wild type (Wilcoxon rank-sum test).

Microscopy and Feulgen staining of seeds

Tissue was harvested from *A. thaliana* and *A. thaliana* \times *A. arenosa* at 7 DAP and 10 DAP (seeds from three siliques per biological replicate). The seeds were stained with Schiff's reagent (Sigma-Aldrich S5133) following fixation and embedding in LR White (London Resin) as described by Braselton *et al.* (1996). An Olympus FluoView 1000 Confocal laser scanning microscope (BX61WI) with an excitation of 488 nm and emission from 500 to 600 nm was used for imaging.

Ploidy measurements of adult hybrids as well as crossing parents

Ploidy was measured for a selection of the plants with two of the *A. arenosa* parents (from MJ09-4 population), two Col-0 plant individuals, 14 *A. thaliana* Col-0 \times *A. arenosa* F1 hybrids and 12 *A. thaliana* *agl36-1 agl90-2* \times *A. arenosa* F1 hybrids. Also, the *A. arenosa* lines Strecno1 and Strecno2 were included to confirm ploidy. One rosette leaf and one inflorescence were analyzed for all samples except for *agl36-1 agl90-2* \times *A. arenosa* F1 hybrids in which only rosette leaves were analyzed. The ploidy was assessed by establishing the genome content by estimating the relative fluorescence intensities by flow cytometry (FCM) and the two-step methodology according to Dolezel *et al.* (2007). The reference standards for the raw cytometric analysis were *Solanum pseudocapsicum* for the *A. thaliana* and *A. arenosa* hybrid comparison with *A. thaliana*, and *Carex acutiformis* for the *A. arenosa* Strecno1 ($N = 10$) and Strecno2 ($N = 12$). The samples and the internal reference were chopped with a razor blade in 0.5 ml ice-cold Otto I buffer (0.1 M citric acid, 0.5% Tween 20). This was then filtered through a nylon mesh (loop size 0.42 μm), incubated at room temperature for 5 min before being stained with 1 ml of Otto II buffer (0.4 M $\text{Na}_2\text{HPO}_4 \cdot 12 \text{H}_2\text{O}$) supplemented with AT-selective fluorescent dye DAPI (4',6-diamino-2-phenylindol) and 2-mercaptoethanol in final concentrations of 4 $\mu\text{g}/\text{ml}$ and 2 $\mu\text{l}/\text{ml}$, respectively. After about 5 min of incubation at room temperature, the relative fluorescence intensity for a minimum of 3000 nuclei was recorded using a Partec Space flow cytometer (Partec GmbH, Münster, Germany) equipped with an UV-LED chip (365 nm). The FCM results are the fluorescence intensities relative to unit fluorescence intensity of the internal reference standard.

ACCESSION NUMBERS

All sequences generated in this study have been deposited in the National Center for Biotechnology Information SRA (<https://www.ncbi.nlm.nih.gov/sra/>) with project number PRJNA562212.

ACKNOWLEDGMENTS

We thank Kirsten Bomblies for supplying *A. arenosa* Strecno1 and Strecno2 populations. Jason Miller and Harinder Singh are acknowledged for RNA-seq analysis. This work was supported by Fri prosjektstøtte grants 214052 and 262247 from the Norwegian Research Council to PEG and AKB.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

KNB and PEG conceived the project and wrote the article with contributions of all the authors. KNB designed the experiments. KNB, JB, KSH and IMJ performed experiments. KNB, JB and KSH analyzed the data. KSH, IMJ, AKK, YSE, MK and RS provided technical assistance. KNB, JB and KSH made figures. JB, AKB and LC contributed to project design and discussion. PEG supervised the project and agrees to serve as the author responsible for contact and ensures communication.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Phylogeny and expression of MADS-box type I transcription factors during seed development.

Figure S2. Imprinting analysis of *AGL28*, *AGL35*, *AGL36* and *AGL90*.

Figure S3. Clustering of MADS-box type I genes based on expression pattern in wild type and *mea* mutant seeds.

Figure S4. *AGL36*-related genes from all available Arabidopsis genomes in Sequence Read Archive, GenBank and Phytozome.

Figure S5. Imprinting analysis of *AGL36-like* in the reciprocal cross of *A. arenosa* and *A. lyrata*.

Figure S6. Germination rate in self crosses of *A. thaliana* accessions and *A. arenosa* MJ09-4 is not affected by temperature.

Figure S7. Ploidy measurement of Arabidopsis populations and hybrids.

Figure S8. Genetic analysis of selected MADS-box type I transcription factors.

Figure S9. Phenotypic characterization of MADS-box transcription factor mutants.

Figure S10. Characterization of a genetic background effect in mixed *A. thaliana* accessions crossed to *A. arenosa*.

Figure S11. F1 hybrid seed phenotypes from *A. arenosa* crosses to *A. thaliana* Col-0, *agl36-1*, *agl90-1*, *agl90-1* and *agl36-1 agl90-2* at 18°C.

Table S1. Schematic overview of restriction digest set-up

Table S2. Oligonucleotide name, sequence and description

Table S3. Characterization of mutants used in this study

Table S4. Segregation of the mutant alleles *agl28-1*, *agl34-2*, *agl35-1*, *agl36-1*, *agl90-1* and *agl90-2* in self crosses and in reciprocal crosses to wild type (Col-0)

Table S5. Crossing scheme for various *A. thaliana* to *A. arenosa* experiments

OPEN RESEARCH BADGE



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA562212>. All sequences generated in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>) with project number PRJNA562212.

REFERENCES

- Airoldi, C.A. and Davies, B. (2012) Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genom.* **39**(4), 157–165.
- Anderson, S.N. and Springer, N.M. (2018) Potential roles for transposable elements in creating imprinted expression. *Curr. Opin. Genet. Dev.* **49**, 8–14.
- Bankevich, A., Nurk, S., Antipov, D. *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**(5), 455–477.
- Belmonte, M.F., Kirkbride, R.C., Stone, S.L. *et al.* (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl Acad. Sci. USA*, **110**, E435–E444.
- Bemer, M., Wolters-Arts, M., Grossniklaus, U. and Angenent, G. (2008) The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *Plant Cell*, **20**(8), 2088–2101.
- Bemer, M., Heijmans, K., Airoldi, C., Davies, B. and Angenent, G.C. (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. *Plant Physiol.* **154**(1), 287–300.
- Berger, F., Grini, P.E. and Schnittger, A. (2006) Endosperm: an integrator of seed growth and development. *Curr. Opin. Plant Biol.* **9**, 664–670.
- Bombliès, K. and Weigel, D. (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat. Rev. Genet.* **8**(5), 382–393.
- Braselton, J.P., Wilkinson, M.J. and Clulow, S.A. (1996) Feulgen staining of intact plant tissues for confocal microscopy. *Biotech. Histochem.* **71**(2), 84–87.
- Brekke, T.D., Henry, L.A. and Good, J.M. (2016) Genomic imprinting, disrupted placental expression, and speciation. *Evolution*, **70**(12), 2690–2703. <https://doi.org/10.1111/evo.13085>.
- Brochmann, C., Brysting, A.K., Alsos, I.G., Borgen, L., Grundt, H.H., Scheen, A.C. and Elven, R. (2004) Polyploidy in arctic plants. *Biol. J. Lin. Soc.* **82** (4), 521–536.
- Burkart-Waco, D., Josefsson, C., Dilkes, B., Kozloff, N., Torjek, O., Meyer, R., Altmann, T. and Comai, L. (2012) Hybrid incompatibility in Arabidopsis is determined by a multiple-locus genetic network. *Plant Physiol.* **158**(2), 801–812.
- Burkart-Waco, D., Ngo, K., Dilkes, B., Josefsson, C. and Comai, L. (2013) Early disruption of maternal-zygotic interaction and activation of defense-like responses in Arabidopsis interspecific crosses. *Plant Cell*, **25** (6), 2037–2055.
- Burkart-Waco, D., Ngo, K., Lieberman, M. and Comai, L. (2015) Perturbation of parentally biased gene expression during interspecific hybridization. *PLoS ONE*, **10**(2), e0117293.
- Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S. and Peacock, W.J. (1997) Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **94**(8), 4223–4228.
- Chen, C., Begcy, K., Liu, K., Folsom, J.J., Wang, Z., Zhang, C. and Walia, H. (2016) Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol.* **171**(1), 606–622. <https://doi.org/10.1104/pp.15.01992>.
- Chen, C., Li, T., Zhu, S. *et al.* (2018) Characterization of imprinted genes in rice reveals conservation of regulation and imprinting with other plant species. *Plant Physiol.* **177**(4), 1754–1771.
- Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.
- Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M. and Colombo, L. (2008) AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *Plant J.* **54**(6), 1037–1048.
- Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y. and Byers, B. (2000) Phenotypic instability and rapid gene silencing in newly formed Arabidopsis Allotetraploids. *Plant Cell*, **12**(9), 1551–1568.
- De Storme, N. and Mason, A. (2014) Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* **1**, 10–33.
- Dilkes, B.P. and Comai, L. (2004) A differential dosage hypothesis for parental effects in seed development. *Plant Cell*, **16**(12), 3174–3180.
- Dolezel, J., Greilhuber, J. and Suda, J. (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**(9), 2233–2244.
- Florez-Rueda, A.M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U. and Stadler, T. (2016) Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol. Biol. Evol.* **33**(11), 2935–2946.
- Folsom, J.J., Begcy, K., Hao, X., Wang, D. and Walia, H. (2014) Rice fertilization-independent Endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol.* **165**(1), 238–248. <https://doi.org/10.1104/pp.113.232413>.
- de Folter, S., Immink, R.G.H., Kieffer, M. *et al.* (2005) Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell*, **17**(5), 1424–1433.
- Gallii, T. (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, **31**(22), 3718–3720.
- Gehring, M. and Satyaki, P.R. (2017) Endosperm and imprinting, inextricably linked. *Plant Physiol.* **173**(1), 143–154.
- Gu, Z., Eils, R. and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32** (18), 2847–2849.
- Haig, D. and Westoby, M. (1989) Parent-specific gene-expression and the triploid endosperm. *Am. Nat.* **134**(1), 147–155.
- Haig, D. and Westoby, M. (1991) Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidy levels of the same species, and its implications for the evolution of Apomixis. *Phil. Trans.: Biol. Sci.* **333**(1266), 1–13.
- Hatorangan, M.R., Laenen, B., Steige, K.A., Slotte, T. and Köhler, C. (2016) Rapid evolution of genomic imprinting in two species of the Brassicaceae. *Plant Cell*, **28**(8), 1815–1827.
- Hollister, J.D., Arnold, B.J., Svedin, E., Xue, K.S., Dilkes, B.P. and Bombliès, K. (2012) Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* **8**, e1003093.
- Hornslie, K.S., Miller, J.R. and Grini, P.E. (2019) Regulation of parent-of-origin allelic expression in the endosperm. *Plant Physiol.* **180**(3), 1498–1519. <https://doi.org/10.1104/pp.19.00320>.
- Jørgensen, M.H., Ehrlich, D., Schmickl, R., Koch, M.A. and Brysting, A.K. (2011) Interspecific and interplodid gene flow in Central European Arabidopsis (Brassicaceae). *BMC Evol. Biol.* **11**(1), 346.
- Josefsson, C., Dilkes, B. and Comai, L. (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* **16**(13), 1322–1328.
- Kang, I.-H., Steffen, J.G., Portereiko, M.F., Lloyd, A. and Drews, G.N. (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell*, **20**(3), 635–647.
- Kim, M.Y. and Zilberman, D. (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci.* **19**(5), 320–326.
- Klosinska, M., Picard, C.L. and Gehring, M. (2016) Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. *Nat. Plants*, **2**, 16145.
- Köhler, C., Hennig, L., Spillane, C., Pien, S., Grisse, W. and Grossniklaus, U. (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.* **17**(12), 1540–1553.
- Köhler, C., Page, D.R., Gagliardini, V. and Grossniklaus, U. (2005) The *Arabidopsis thaliana* MEDEA polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat. Genet.* **37**(1), 28–30.
- Lafon-Placette, C., Johannessen, I.M., Hornslie, K.S. *et al.* (2017) Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc. Natl Acad. Sci. USA*, **114**(6), E1027–E1035.
- Lafon-Placette, C. and Köhler, C. (2016) Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol. Ecol.* **25**(11), 2620–2629. <https://doi.org/10.1111/mec.13552>.
- Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
- Lindsey, B.E. 3rd, Rivero, L., Calhoun, C.S., Grotewold, E. and Brkljacic, J. (2017) Standardized method for high-throughput sterilization of Arabidopsis seeds. *J. Vis. Exp.* (128), 56587. <https://doi.org/10.3791/56587>
- Lu, J., Zhang, C., Baulcombe, D.C. and Chen, Z.J. (2012) Maternal siRNAs as regulators of parental genome imbalance and gene expression in

- endosperm of Arabidopsis seeds. *Proc. Natl Acad. Sci. USA*, **109**(14), 5529–5534.
- Makarevich, G., Villar, C.B.R., Erilova, A. and Köhler, C.** (2008) Mechanism of PHERES1 imprinting in Arabidopsis. *J. Cell Sci.* **121**(6), 906–912.
- Masiero, S., Colombo, L., Grini, P.E., Schnittger, A. and Kater, M.M.** (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*, **23**(3), 865–872.
- McCarthy, D.J., Chen, Y. and Smyth, G.K.** (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**(10), 4288–4297.
- Murashige, T. and Skoog, F.** (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.
- Nam, J., Kim, J., Lee, S., An, G., Ma, H. and Nei, M.** (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc. Natl Acad. Sci. USA*, **101**(7), 1910–1915.
- Novikova, P.Y., Tsuchimatsu, T., Simon, S. et al.** (2017) Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol. Biol. Evol.* **34**(4), 957–968.
- Nowack, M.K., Ungru, A., Bjerkan, K.N., Grini, P.E. and Schnittger, A.** (2010) Reproductive cross-talk: seed development in flowering plants. *Biochem. Soc. Trans.* **38**, 604–612.
- Parenicova, L., de Folter, S., Kieffer, M. et al.** (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*, **15**(7), 1538–1551.
- Pattengale, N.D., Aberer, A.J., Swenson, K.M., Stamatakis, A. and Moret, B.M.** (2011) Uncovering hidden phylogenetic consensus in large data sets. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**(4), 902–911.
- Pfaffl, M.W.** (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**(9), e45.
- Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W. and Gehring, M.** (2014) Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife*, **3**, e03198.
- Rebernik, C.A., Lafon-Placette, C., Hatorangan, M.R., Slotte, T. and Köhler, C.** (2015) Non-reciprocal interspecies hybridization barriers in the Capsella genus are established in the endosperm. *PLoS Genet.* **11**(6), e1005295.
- Rodrigues, J.A. and Zilberman, D.** (2015) Evolution and function of genomic imprinting in plants. *Genes Dev.* **29**(24), 2517–2531. <https://doi.org/10.1101/gad.269902.115>.
- Roth, M., Florez-Rueda, A.M. and Stadler, T.** (2019) Differences in effective ploidy drive genome-wide endosperm expression polarization and seed failure in wild tomato hybrids. *Genetics*, **212**(1), 141–152. <https://doi.org/10.1534/genetics.119.302056>.
- Saze, H., Mittelsten Scheid, O. and Paszkowski, J.** (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **34**, 65–69.
- Shirzadi, R., Andersen, E.D., Bjerkan, K.N. et al.** (2011) Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of AGAMOUS-LIKE36. *PLoS Genet.* **7**(2), e1001303.
- Stamatakis, A.** (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9), 1312–1313.
- Stanke, M. and Morgenstern, B.** (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467.
- Stebbins, G.** (1984) Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach. *Bot. Helv.* **94**, 1–13.
- Steffen, J., Kang, I., Portereiko, M., Lloyd, A. and Drews, G.** (2008) AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. *Plant Physiol.* **148**(1), 259–268.
- Suyama, M., Torrents, D. and Bork, P.** (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612.
- Tonosaki, K., Sekine, D., Ohnishi, T., Ono, A., Furuumi, H., Kurata, N. and Kinoshita, T.** (2018) Overcoming the species hybridization barrier by ploidy manipulation in the genus *Oryza*. *Plant J.* **93**(3), 534–544.
- Vallejo-Marin, M. and Hiscock, S.J.** (2016) Hybridization and hybrid speciation under global change. *New Phytol.* **211**(4), 1170–1187.
- Villar, C.B., Erilova, A., Makarevich, G., Trosch, R. and Köhler, C.** (2009) Control of PHERES1 imprinting in Arabidopsis by direct tandem repeats. *Mol. Plant*, **2**(4), 654–660.
- Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J. and Comai, L.** (2009) Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.* **19**(13), 1128–1132.
- Wang, L., Yuan, J., Ma, Y., Jiao, W., Ye, W., Yang, D.L., Yi, C. and Chen, Z.J.** (2018) Rice interploidy crosses disrupt epigenetic regulation, gene expression, and seed development. *Mol. Plant*, **11**(2), 300–314.
- Waters, A.J., Bilinski, P., Eichten, S.R., Vaughn, M.W., Ross-Ibarra, J., Gehringer, M. and Springer, N.M.** (2013) Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc. Natl Acad. Sci. USA*, **110**, 19639–19644.
- Wolf, J.B., Oakey, R.J. and Feil, R.** (2014) Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity (Edinb)*, **113**, 167–175. <https://doi.org/10.1038/hdy.2014.11>.
- Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M. and Köhler, C.** (2011) High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. *PLoS Genet.* **7**, e1002126.
- Yoo, S.K., Lee, J.S. and Ahn, J.H.** (2006) Overexpression of AGAMOUS-LIKE 28 (AGL28) promotes flowering by upregulating expression of floral promoters within the autonomous pathway. *Biochem. Biophys. Res. Commun.* **348**, 929–936.
- Yoshida, T. and Kawabe, A.** (2013) Importance of gene duplication in the evolution of genomic imprinting revealed by molecular evolutionary analysis of the type I MADS-box gene family in Arabidopsis species. *PLoS ONE*, **8**, e73588.
- Zhang, S., Wang, D., Zhang, H., Skaggs, M.I., Lloyd, A., Ran, D., An, L., Schumaker, K.S., Drews, G.N. and Yadegari, R.** (2018) FERTILIZATION-INDEPENDENT SEED-polycomb repressive complex 2 plays a dual role in regulating type I MADS-box genes in early endosperm development. *Plant Physiol.* **177**, 285–299.