Conservation of the Restricted Expression of Brassicaceae B_{sister}-Like Genes in Seeds Requires a Transposable Element in *Arabidopsis thaliana*

Clemens Roessner [1], Amey S. Bhide,[1,2] Andrea Hoffmeier,[3] Julian Schenk,[1,4] Thomas Groß,[1,5] Lydia Gramzow [3], Günter Theißen [3] and Annette Becker *[,1]

[1]Institute of Botany, Justus-Liebig-University Gießen, Gießen, Germany
[2]OrchidKraft, Satara, India
[3]Matthias Schleiden Institute/Genetics, Friedrich-Schiller-University Jena, Jena, Germany
[4]Service-Verband KVD, Dorsten, Germany
[5]SGS Institut Fresenius GmbH, TraitGenetics Section, Gatersleben, Germany

***Corresponding author:** E-mail: annette.becker@bot1.bio.uni-giessen.de.

Associate editor: Patricia Wittkopp

Abstract

Changes in transcription factor binding sites (TFBSs) can alter the spatiotemporal expression pattern and transcript abundance of genes. Loss and gain of TFBSs were shown to cause shifts in expression patterns in numerous cases. However, we know little about the evolution of extended regulatory sequences incorporating many TFBSs. We compare, across the crucifers (Brassicaceae, cabbage family), the sequences between the translated regions of *Arabidopsis* B_{sister} (*ABS*)-like MADS-box genes (including paralogous *GOA*-like genes) and the next gene upstream, as an example of family-wide evolution of putative upstream regulatory regions (PURRs). *ABS*-like genes are essential for integument development of ovules and endothelium formation in seeds of *Arabidopsis thaliana*. A combination of motif-based gene ontology enrichment and reporter gene analysis using *A. thaliana* as common *trans*-regulatory environment allows analysis of selected Brassicaceae B_{sister} gene PURRs. Comparison of TFBS of transcriptionally active *ABS*-like genes with those of transcriptionally largely inactive *GOA*-like genes shows that the number of in silico predicted TFBS) is similar between paralogs, emphasizing the importance of experimental verification for in silico characterization of TFBS activity and analysis of their evolution. Further, our data show highly conserved expression of Brassicaceae *ABS*-like genes almost exclusively in the chalazal region of ovules. The *Arabidopsis*-specific insertion of a transposable element (TE) into the *ABS* PURRs is required for stabilizing this spatially restricted expression, while other Brassicaceae achieve chalaza-specific expression without TE insertion. We hypothesize that the chalaza-specific expression of ABS is regulated by *cis*-regulatory elements provided by the TE.

Key words: **evolution of gene expression regulation, MADS-box gene, Brassicaceae, B_{sister} gene, GUS assay, transcription factor binding site (TFBS), transposable element (TE).**

Introduction

Since plants conquered land ~470 million years ago (Ma), they rapidly radiated into a morphologically highly diverse group, with the flowering plants (angiosperms) as the most successful clade in terms of species richness and soil coverage (Barba-Montoya et al. 2018). Several independent gene duplication and whole-genome duplication (WGD) events during land plant evolution (Clark and Donoghue 2018) followed by functional changes affecting one or both of the gene copies (Ohno 1970) are suggested drivers of the diversification and radiation of flowering plants (Edger and Pires 2009; Rensing 2014; Panchy et al. 2016).

An extraordinary example of a clade of regulatory genes whose members experienced a wide spectrum of fates after a duplication event, including gene loss, subfunctionalization, and neofunctionalization, is the B_{sister} genes (Hoffmeier et al. 2018). These genes got their name because of their evolutionary relation to class B floral homeotic genes and their preferential expression in female reproductive organs (Becker et al. 2002). After the split of B and B_{sister} genes, another duplication event occurred after the Ranunculales branched off from the most recent common ancestor (MRCA) of extant core eudicots, eventually leading to two B_{sister} genes in *Arabidopsis thaliana*, *ABS* (*Arabidopsis* B_{sister}, also called *TRANSPARENT TESTA 16* or *AGL32*) and *GOA* (*GORDITA*, also called *AGL63*) (Hoffmeier et al. 2018).

Like the paralogous group of class B genes, B_{sister} genes encode MIKC-type MADS-domain transcription factors

Article

(TFs) regulating reproductive development in seed plants. Class B genes specify petal and stamen identity of flowering plants (Coen and Meyerowitz 1991) and are also expressed in male cones of gymnosperms, where they very likely have a similar function in specifying male reproductive organs (Winter et al. 1999; Theißen and Becker 2004).

The expression patterns of *ABS* and *GOA* in *A. thaliana* differ: While *ABS* is mainly expressed in the endothelium, that is, the innermost layer of the inner integument focused on the micropylar and the chalazal end, *GOA* is expressed in the outer integument of the ovules but much weaker (Nesi et al. 2002; Prasad et al. 2010). In the following, we refer to *ABS*-like genes sensu lato (s.l.) when the clade including both *ABS* and *GOA* orthologs is meant and to *ABS*-like genes sensu stricto (s.s.) when only *ABS* orthologs are described.

The female reproductive organs of flowering plants include ovules, composed of the nucellus harboring the egg cell, which is surrounded by the inner and outer integument. The ovule is connected to the maternal plant at the chalazal end by a funiculus, while the opposite micropylar end provides the site for pollen entry required for fertilization of the egg cell. After fertilization, the integuments develop into the seed coat. The expression of B_{sister} genes occurs often in the integuments and nucellus of the ovules of flowering plants and female cones of gymnosperms (Coen and Meyerowitz 1991; Becker et al. 2002, 2003; Lovisetto et al. 2013).

The duplication event leading to the two B_{sister} genes in extant eudicots likely included the regulatory regions of the genes and hence changes in expression patterns after the duplication are probably due to mutations in *cis*-regulatory elements (CREs), many (if not most) of which are probably TF binding sites (TFBSs). Change in CREs is a possible driver of B_{sister} gene evolution. The alteration of CREs within a regulatory DNA region can, even without changes in the *trans*-acting transcription machinery, lead to modifications in expression patterns and subsequently may allow for neofunctionalization or nonfunctionalization (pseudogenization) (Kusters et al. 2015; Jiang and Rausher 2018). Here, we define the putative upstream regulatory region (PURR) as the genomic region upstream of the start codon of a gene of interest until the upstream gene's 3′ UTR end. The PURR includes the 5′UTR, immediate promoter (iP) required for basal transcription when alone, CREs, TFBSs, and DNA of undefined function.

ABS regulates cell shape of the inner integument in ovules. Furthermore, it affects seed pigmentation, endosperm proliferation, and degeneration of the nucellus during embryo development and the apoplastic barrier of the seed coat (Nesi et al. 2002; Ehlers et al. 2016; Xu et al. 2016, 2017). The seeds of *abs* mutants show a light brown coloration, owing to a lack of proanthocyanin pigmentation. *ABS* appears to be involved in the regulation of flavonoid biosynthesis in the seed coat, which provides the precursors for proanthocyanins. They are condensed tannins that are located at the endothelium layer of the seed and participate in seed longevity and dormancy (Nesi et al. 2002).

The functions of *GOA* are harder to pinpoint. It has been reported that loss of *GOA* function affects the longitudinal growth of the fruits (Prasad et al. 2010), but these results could not be reproduced by at least two other laboratories (Hoffmeier et al. 2018; our unpublished results). Protein sequences of *GOA*-like genes in other Brassicaceae exhibit several mutations in conserved domains leading to structural changes and less protein interaction. Moreover, their expression levels in different Brassicaceae are close to or even below the detection limit of quantitative reverse transcription PCR (RT-qPCR) (Erdmann et al. 2010; Hoffmeier et al. 2018). In line with this, a strong tendency of Brassicaceae species to lose *GOA*-like genes was recently reported (Hoffmeier et al. 2018). This ongoing loss of *GOA*-like genes in Brassicaceae may be caused by mutations in CREs or in their protein coding sequences. Whatever the cause, *GOA*-like genes are "marked for death" in the Brassicaceae (Hoffmeier et al. 2018).

Brassicaceae are divided into the sister lineages of core Brassicaceae and the tribe Aethionemeae (Guo et al. 2017; Nikolov et al. 2019; Walden et al. 2020); and the core Brassicaceae are separated into five lineages (I–V) (Nikolov et al. 2019). Of the species studied here, *Arabis alpina* belongs to lineage IV, *Eutrema salsugineum* belongs to lineage II, while *Capsella rubella* and the most prominent model organism within the Brassicaceae, *A. thaliana*, are members of lineage I (Walden et al. 2020). These species were selected for analysis because sufficient genome information was present and established lines were available, covering three of the five core Brassicaceae lineages.

In this study, we employ *GUS* reporter lines and motif-based gene ontology (GO) enrichment analyses to elucidate the conservation of *ABS*-like s.s. regulatory sequences. We observed their spatial and temporal activity in transformed *A. thaliana* plants in closely related but different heterologous environment. Previously, we showed that the *ABS*-like genes s.l. are predominantly expressed in buds, flowers, and siliques (Hoffmeier et al. 2018). Here, we show that the activity of closely related regulatory sequences in the heterologous system is largely consistent with the known expression pattern.

Results

The CREs of *ABS*-Like Genes s.s. Include Conserved and Diverged Sequence Blocks

We isolated the PURRs and genomic loci of *A. thaliana* and *C. rubella* (lineage I), *E. salsugineum* (lineage II), and *A. alpina* (lineage IV) as easily accessible representatives of Brassicaceae diversity. We first analyzed the degree of conservation of the regions immediate upstream of the translation start sites of *ABS*-like genes s.l. until the 3′ end of the next gene's 3′ UTR, assuming that conservation indicates CREs. CREs may include promoters, enhancers, silencers, and insulators; they may affect gene expression via

recruitment of *trans*-acting factors such as TFs, chromatin remodelers, and RNA polymerase II (Hajheidari and Huang 2022). It is reasonable to assume that many CREs are TFBSs when comparing the PURRs of *ABS*-like s.s. and *GOA*-like genes. The PURRs were similar in length with 2.8–2.9 kb for *ABS*-like s.s. sequences. Only the PURR of *A. thaliana* was significantly larger with ~3.6 kb due to the insertion of a transposable element (TE) (fig. 1*A*). We identified the TE in the PURR of *ABS* as a plant miniature inverted-repeat transposable element (MITE) of the family DTA_Art2, member of the hAT superfamily (Chen et al. 2014) ~1 kb in length.

Phylogenetic shadowing (fig. 1*A*) revealed two highly conserved regions (evolutionarily conserved region 1 [ECR1] and 2 [ECR2]) in the PURRs of *ABS*-like genes s.s., positioned ~1 and 3 kb upstream of the start codons. ECR1 is ~200 bp and ECR2 is ~100 bp long, and they are separated by a nonconserved region. ECR1 and ECR2 are conserved in *ABS*-like s.s. PURRs among all Brassicaceae analyzed (supplementary fig. S1, Supplementary Material online) with a sequence identity of ~90%. In addition to ECR1 and ECR2, a ~500-bp-long region directly adjacent to the 5′ UTR, including a predicted TATA-box, showed a sequence similarity of over 70%, as expected for the iP. The overall sequence identity of *ABS*-like s.s. PURRs, including the ECRs based on pairwise distances among Brassicaceae varied between 58% for *E. salsugineum* to *Capsella grandiflora* and 90% for *C. grandiflora* to *C. rubella* and several other Brassicaceae (supplementary Table S1, Supplementary Material online). The average sequence identity between the PURRs of ABS of *A. thaliana* (pABS), *C. rubella* (pCrABS), *A. alpina* (pAaABS), and *E. salsugineum* (pEsABS) is 61%.

In addition to the large insertion in the PURR of *A. thaliana*, we identified an insertion at the proximal part of pAaABS and another deletion in the more distal region of pEsBS (fig. 1*C*) when compared with pABS. Longer stretches of sequence identity between pABS and pEsBS in the distal part of their PURRs were not observed. Several lineage-specific insertions/deletions are shown when the PURRs of *ABS*-like genes s.s. from *C. rubella*, *A. alpina*, and *E. salsugineum* are compared (fig. 1*A*).

In comparison with the PURR of *ABS*-like genes s.s., the PURRs of *GOA*-like genes were shorter (~1.8 kb [pGOA] for *A. thaliana*, ~1.5 kb for *C. rubella* [pCrGOA], and ~2.4 kb for *A. alpina* [pAaGOA]). *Eutrema salsugineum* has lost its *GOA*-like gene (Hoffmeier et al. 2018). Phylogenetic shadowing carried out for *GOA*-like sequences (fig. 1*B*) showed a weak conservation of PURRs. They did not include regions conserved between lineage I and lineage II representatives, only shorter stretches of sequence identity between pGOA and pCrGOA, which are both members of lineage I (fig. 1*D*). Interestingly, also the levels of sequence conservation in the exons and introns were weaker than those in *ABS*-like genes s.s. This is in line with previous reports on the lack of domain structure conservation of GOA-like proteins (Erdmann et al. 2010; Hoffmeier et al. 2018).

ABS-Like Gene's PURRs Share CREs

We were then interested to learn if the ABS-like gene s.s. PURRs have shared TFBSs. We thus carried out in silico searches for putative TFBSs in *ABS*-like gene s.l. PURRs using PlantPan 3.0 (Chow et al. 2019) with different settings. For the purpose of this analysis, we assume no divergence in *trans*-regulators between *A. thaliana* and the other Brassicaceae analyzed here. The output lists we retrieved contained information on TFBSs (motifs), the TFs binding to the TFBS, and the position of the TFBS within the input PURR. Because outside *A. thaliana* information on CRE binding is unavailable, we used the information available for *A. thaliana* to infer binding of TFs to CREs in the other three analyzed species, meaning we searched for TFBSs in PURRs of *C. rubella*, *Arabidopsis lyrata*, and *E. salsugineum* using the background settings for *A. thaliana* of PlantPan. All further steps were carried out using lists of TFs retrieved from PlantPan. First, we included all predicted TFs into our analysis irrespective of their TF binding being experimentally demonstrated. We identified 520 (pABS), 653 (pCrABS), 468 (pEsABS), and 447 (pAaABS) TFs with predicted binding motifs in the PURRs (fig. 1*E*). The majority of 324 CREs are shared between all sequences. The number of shared CREs between *GOA*-like sequences was surprisingly high (394). Between 485 (pCrGOA) and 606 (pAaGOA), CREs were found in the individual PURRs of *GOA*-like genes, and 247 CREs were shared between all B_{sister} sequences (fig. 1*E*). These numbers appeared very high, which was due the fact that a single TFBS might be linked to several different TFs and that there are TFs that only need a four base TFBS. Since we do not know which of these TFs bind in planta we considered all of them.

We then utilized a database of chromatin immunoprecipitation sequencing (ChIP-Seq) experiments, carried out in *A. thaliana*, to identify those CREs that might be functionally relevant. We only had ChIP-Seq data available for *A. thaliana* and no data for pABS in particular, leaving us with the assumption that TFs of the other Brassicaceae would bind to the conserved TFBSs in a similar manner. Among the shared 324 elements in PURRs of *ABS*-like genes s.s., 54 were shown by ChIP-Seq experiments to be bound by TFs (supplementary fig. S2, Supplementary Material online). Among the TFs with binding motifs conserved in position between all analyzed species are SEP3, RIF4, HB5, ANAC102, HSFA1A, MYB44, ABF4, and PIF1 in ECR1. In ECR2, binding motifs for SEP3, JAG, PIF4, RGA, ANAC102, PHYA, CRY2, MYB44, BZIP28, ABF4, HAT2.2, RD26, ZAT6, and BBM were found in all species (supplementary fig. S2, Supplementary Material online).

We were then interested to learn if the TFs with predicted binding sites in CREs of *ABS*-like genes s.s. were co-expressed with *ABS*. We thus analyzed the expression of putative regulators in developing ovules and seeds by digital gene expression analysis and clustering according to their relative expression in two different stages of ovule development, where *ABS* expression is stronger in early

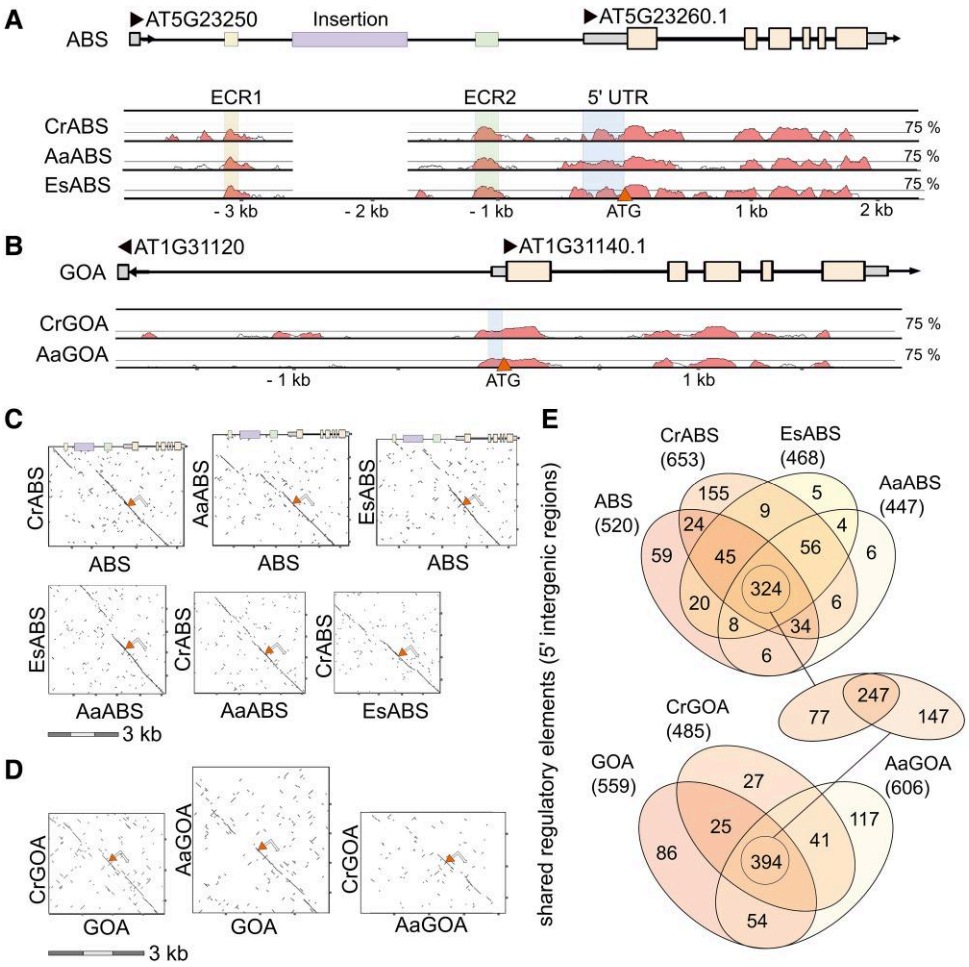


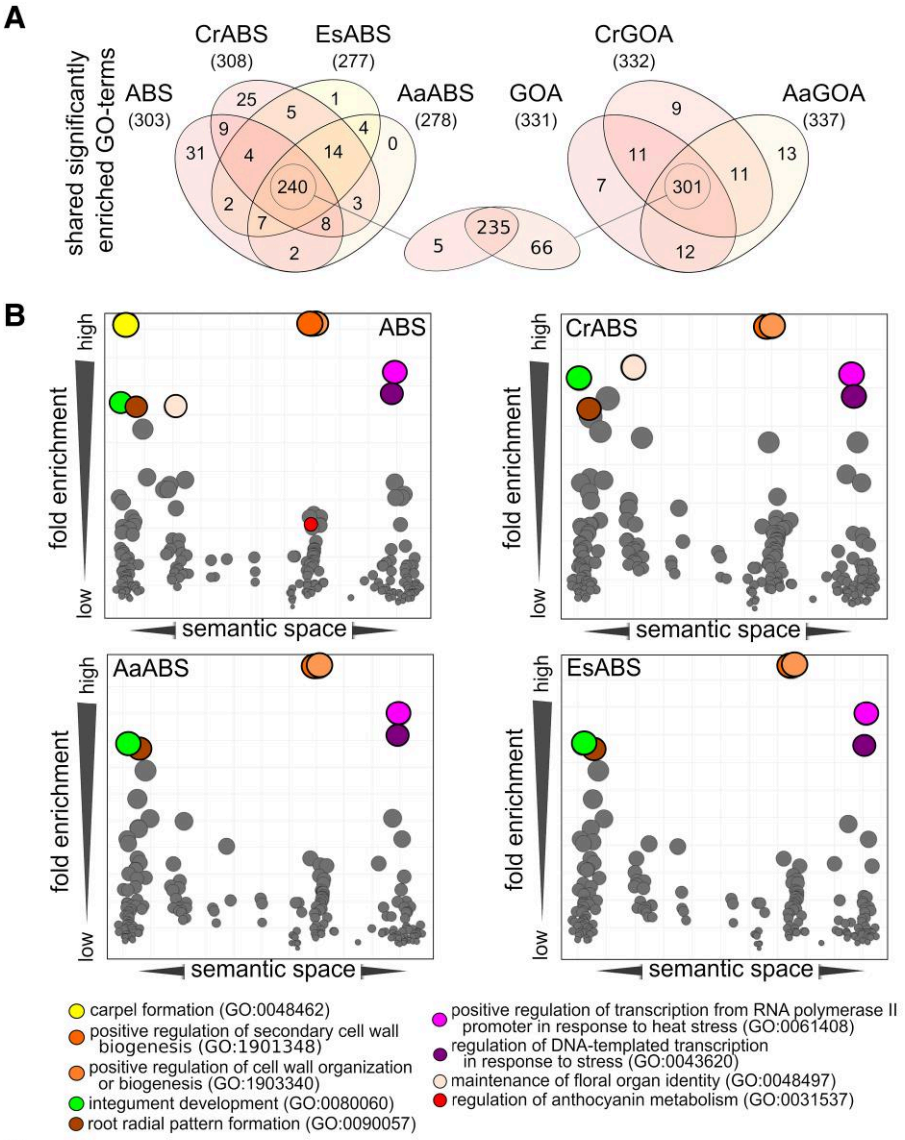
Fig. 1. Sequence conservation of Brassicaceae B_{sister} gene CREs. (*A*) Shadowing analysis of *ABS*-like s.s. PURRs and transcribed regions of *C. rubella*, *A. alpina*, and *E. salsugineum* in relation to the corresponding sequence of ABS. The position of protein coding exons and UTRs of ABS, the evolutionary conserved regions (ECR1 and ECR2), the TE insertion specific to *A. thaliana* are marked as boxes. Highlighted sequence segments show a conservation above 70% over 100 bp (cutoff at 50% nucleotide identity). (*B*) Shadowing analysis of *GOA*-like sequences for comparison. Dot blots visualizing pairwise alignments of ABS-like s.s. (*C*) and GOA-like (*D*) genomic sequences including the PURRs and the transcribed regions. Uninterrupted lines indicate sequence conservation, triangles mark start codons and the arrow indicates the direction of transcription. (*E*) Venn diagrams of in silico predicted shared and unique regulatory elements within the PURR based on *A. thaliana* CRE.

stages if compared with later stages (Klepikova et al. 2016) (supplementary fig. S2, Supplementary Material online). Of the 54 TFs that were shown by ChIP-Seq experiments to bind to the PURRs of *ABS*-like genes s.s., nine are not expressed in ovules during the developmental stages in question leaving 45 factors. Further, we included *LEC2*, which was verified to bind to the p*ABS* PURR by yeast one-hybrid assay (not shown), and *GOA* for comparison. The expression analysis revealed a large cluster of genes coexpressed with *ABS* in ovules including *MYB3*, *FLM*, *ABF3*, *HB7*, *ANAC032*, *SEP3*, *ABI5*, *PIF5*, *PIF4*, *RGA*, *AP2*, *HB5*, *ANAC102*, *HB6*, *MYB3R3*, *HSFA1A*, *PHYA*, *CRY2*, *MYB44*, *TOC1*, *GBF3*, *BZIP28*, *ABF4*, *HAT22*, *FHY3*, *PIF1*, *GBF2*, and *RD26* with strong early and weaker late expression. Among these regulators are light-dependent factors like *PhyA*, *Cry2*, phytochrome interacting factors (*PIFs*), *GBF2*, and *FHY3*, suggesting that ABS expression may be regulated by light. Further, at least six factors are known to act in an abscisic acid (ABA)-dependent manner (*ABF3*, *ABF4*, *ABI5*, *HB5*, *HB6*, and *RD26*) linking *ABS* expression to seed development, maturation, and germination. Another set of genes including *LEC2*, *LEC1*, *ZAT6*, *IBH1*, *JAG*, and *BBM* showed a pattern complementary to the *ABS* expression, with low expression during early ovule development and stronger expression during later development. Both *LEC1* and *LEC2* play a critical role during embryo development, suggesting that TFs that positively regulate embryo development may negatively regulate *ABS* expression.

Go Enrichments Indicate a High Level of Conservation of a Core Set of Regulators of *ABS*-Like Genes

We were then interested to see if the regulators predicted by PlantPan 3.0 of *ABS*-like genes s.l. classify into distinct functional categories, prompting us to analyze their GO annotations (fig. 2*A*). We plotted shared GO terms of the TFs binding to TFBSs of the four *ABS*-like genes s.s. For each species-specific set of TFs with predicted binding sites, a GO enrichment analysis was carried out using unweighted lists of single representations (multiple binding sites were not considered) against the *A. thaliana* background to obtain an impression of regulatory conservation.

As expected, a comparison of GO terms with significant positive enrichments ($P \leq 0.05$) showed high similarity among the different *ABS-like* gene s.l. PURRs (fig. 2*A*). Of 337 different GO terms, roughly 70% (235 terms) were shared between all *ABS-like* gene s.l. PURRs. Among the



C

	factors asociated with anthocyanin acc. (GO based)	type	pABS ECR1	pABS Insert	pABS ECR2	pABS iP	pCrABS ECR1	pCrABS ECR2	pCrABS iP	pAaABS ECR1	pAaABS ECR2	pAaABS iP	pEsABS ECR1	pEsABS ECR2	pEsABS iP
predicted binding motif(s)	ABS	MADS-Box			x		x	x			x			x	
	TT8	bHLH		x		x									
	PIF3	G-Box / E-box binding		x		x									
	MYC3	bHLH		x											
	MYC4	bHLH Leucine Zipper		x											
	ANL2	homeodomain		x		x		x							x
	ARR(1, 10, 12)	type-B ARR		x		x			x					x	x

Fig. 2. Comparison of GO terms of regulatory factors binding to CRSs in PRGs of B_{sister} genes. (*A*) Venn Diagrams of significantly enriched GO terms for regulatory factors binding to CRSs of *ABS*-like genes s.s. (left), *GOA*-like genes (right), and CREs shared among all *ABS*-like genes s.l. (middle). The GO terms are derived from single representations of each factor; multiple binding sites within the PURRs were not considered. (*B*) Comparison of enriched GO terms between regulatory factors of *ABS*-like gene s.s. CREs using semantic spaces. Each circle represents a single term (see figure legend); size and *y*-axis indicate enrichment strength. The eight GO terms with the strongest enrichment among predicted regulatory factors of *A. thaliana* as well as the only significantly enriched term related to the anthocyanin biosynthesis (GO:00311537) in all *ABS*-like gene s.s. data sets are highlighted by color and used for comparison to other *ABS*-like gene s.s. PURRs (*y*-axis, relative fold enrichment; *x*-axis, semantic similarity). (*C*) Distribution of binding sites for TFs related to anthocyanin biosynthesis that were predicted to bind. The presence of a binding site for a factor (column) in a certain sequence segment (row) is marked with x.

significantly enriched and shared terms, we identified several terms related to the known functions of *ABS* and *GOA* like "plant ovule development (GO:0048481)" and "integument development (GO:0080060)."

Interestingly, four GO terms related to transcriptional activity in response to stress were enriched in all *ABS*-like data sets s.s. but not in all *GOA*-like data sets (fig. 2A). These GO terms were positive regulation of transcription from RNA polymerase II promoter in response to heat stress (GO:0061408), positive regulation of transcription from RNA polymerase II promoter in response to stress (GO:0036003), regulation of DNA-templated transcription in response to stress (GO:0043620), and regulation of transcription from RNA polymerase II promoter in response to stress (GO:0043618). Within the *ABS*-like gene s.s. data sets, pABS showed the highest number of uniquely enriched terms including the term "regulation of anthocyanin metabolic process (GO:0031537)," the only enriched term related to the anthocyanin biosynthesis in all *ABS*-like gene s.l. data sets.

We then plotted the enriched GO terms of *ABS*-like gene s.s. data sets in semantic spaces, clustering the GO

terms based on their distance within the GO hierarchy (e.g., the term "integument development" is a child term of "plant ovule development," which is a child term of "carpel development"; these three are more similar to each other than to root radial pattern formation; while all four are child terms of "anatomical structure development"). We highlighted the eight terms with the highest enrichment in each individual set plus the term "regulation of anthocyanin biosynthesis" (fig. 2*B*) although the latter was included because of its unexpected absence in *ABS*-like gene s.s. data sets other than pABS. The distribution of terms within the semantic spaces is highly similar but not identical between the regulators of *ABS*-like genes s.s. Among the eight terms with the highest enrichment in the data set of *A. thaliana*, six were shared by all *ABS*-like genes s.s. Here, terms related to the cell wall biogenesis showed the highest enrichment in all data sets (fig. 2*B*, orange circles). Further, two of the GO terms are related to the regulation of transcription in response to stress (fig. 2*B*, purple circles). The terms "integument development (GO:0080060)" (fig. 2*B*, green circles) and "root radial pattern formation (GO:0090057)" were strongly enriched in all data sets (fig. 2*B*, brown circles). The term "maintenance of floral organ identity (GO:0048497)" (fig. 2*B*, beige circles) was specific to the regulators of *ABS*-like genes s.s. of lineage I. Only *A. thaliana* showed an enrichment of the terms "carpel formation (GO:0048462)" (fig. 2*B*, yellow circles) and "regulation of anthocyanin metabolic process (GO:0031537)" (fig. 2*B*, red circle). The latter indicates that the genetic connection of *ABS* to the regulation of anthocyanin metabolism reported earlier (Nesi et al. 2002) is specific to *A. thaliana*.

To identify the reason for the enrichment of anthocyanin related terms, we investigated the distribution of predicted binding sites (fig. 2*C*). All regulatory sequences exhibit at least one CArG-box within ECR2, which might serve as binding site for ABS, but no other predicted binding site for an anthocyanin related factor is conserved between the PURRs. Further, TRANSPARENT TESTA 8 (TT8), PHYTOCHROME INTERACTING FACTOR 3 (PIF3), MYC 3, and MYC 4 as well as ANTHOCYANINLESS 2 (ANL2) and three different ARRs (*Arabidopsis* response regulators) were predicted to bind in varying PURRs. Interestingly, the *ABS*-specific TE insertion harbors predicted binding sites for all of these factors only lacking a predicted binding site for ABS.

Functionality of the *ABS* PURRs Is Partially Conserved in Late Floral Developmental Stages Throughout Brassicaceae

We were interested in the readout of the *cis*-regulatory system of the four *ABS*-like genes s.s. in the *trans*-acting system of *A. thaliana*. We thus transformed *A. thaliana* Col-0 plants with constructs in which the PURRs of *A. thaliana*, *C. rubella*, *A. alpina*, and *E. salsugineum ABS*-like genes s.s. were fused to GUS (supplementary table S2, Supplementary Material online). For each *ABS*-like gene, two reporter lines were introduced into *A. thaliana*, the PURR with and without the respective transcribed regions including all exons and introns. Between 7 and 17 independently transformed lines per construct were analyzed (table 1). We observed the activity of the PURRs using five different developmental stages of flower development (Smyth et al. 1990) beginning with the initiation of the carpel between stages 5 and 6, the elongation around stage 8, ovule initiation at stage 10, and ending before the opening of the bud at stage 12. GUS staining in ovules was observed around anthesis (stage 13). Here, we describe the GUS staining pattern of the transgenic lines (table 1 and fig. 3).

Table 1. ***Arabidopsis thaliana*** **GUS Reporter Lines**

Reporter Construct	Number of Independent Lines	Consistent Staining Pattern	Percentage of Consistent Staining
pABS:ABS:GUS	17	15	88%
pCrABS:CrABS:GUS	13	13	100%
pAaABS:AaABS:GUS	12	10	83%
pEsABS:ABS:GUS	8	8	100%
pABS:GUS	7	6	85%
pCrABS:GUS	13	13	100%
pAaABS:GUS	12	10	83%
pEsABS:GUS	7	6	85%
pABSΔI:GUS	12	9	75%
pABSΔECC1:GUS	10	4	40%
pABS_iP:GUS	10	6	60%

NOTE.—Number of independent and consistently stained lines. For pABS_iP:GUS, six lines showed staining while four lines did not (60%); none of the lines shows staining in ovules (100%).

In stages 5–8, no GUS staining was observed for the PURRs of *ABS* and *EsABS*, regardless of the presence of the transcribed sequences. The lines carrying pCrABS:CrABS:GUS (with transcribed sequences) and pCrABS:GUS (without transcribed sequences) constructs showed GUS expression in the entire carpel primordia while both pAaABS:GUS and pAaABS:AaABS:GUS constructs showed expression at the base of the developing carpels (fig. 3*A*–*D*). The expression in these tissues ceased before stage 8. Staining was observed in the developing locules of the anthers for both pAaABS constructs around stage 8. In this stage, the expression for both pCrABS constructs was limited to the base of the developing carpel. Lines carrying pCrABS:GUS showed additional expression at the base of the stamens.

Beginning with stage 10, the expression of all transgenic lines was limited to the developing ovules and continued until the opening of the buds at stage 13. Additional expression could be observed in the stigmatic papillae of pCrABS:GUS lines but not in pCrABS:CrABS:GUS lines.

The expression of GUS in the ovules (fig. 3*E*) varied between the different reporter lines. We observed expression in the inner integument layers and the chalazal end of pABS:ABS:GUS ovules in line with previous reports (Nesi et al. 2002; Ehlers et al. 2016). However, the pABS:GUS reporter lines showed even distribution of GUS staining

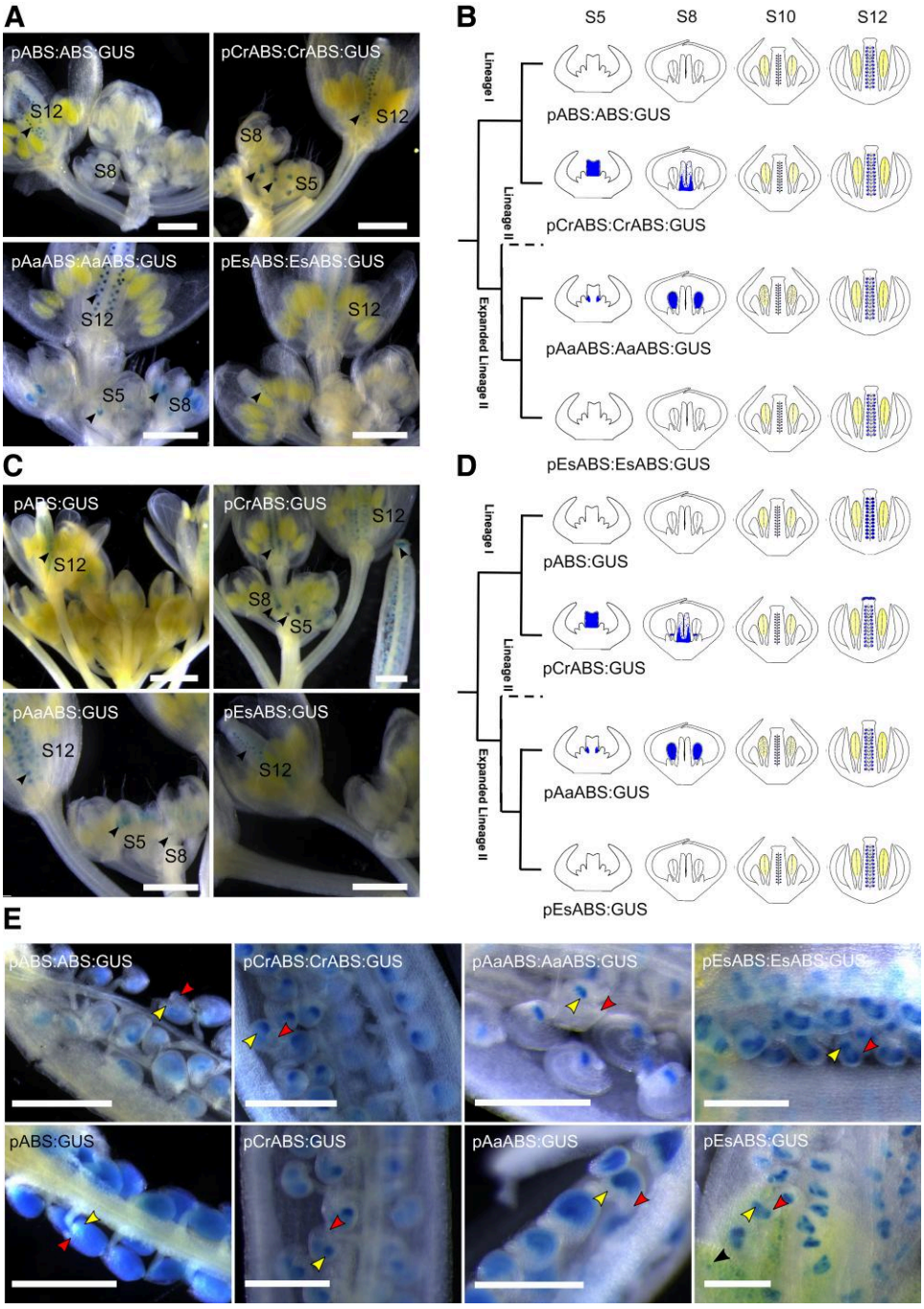


FIG. 3. Readout of *ABS*-like regulatory sequences by *A. thaliana trans*-acting factors visualized by GUS staining. Activity of PURRs and transcribed regions of Brassicaceae *ABS*-like genes s.s. in transgenic *A. thaliana* lines (*A*) GUS staining pattern of *A. thaliana* lines containing the PURR and transcribed regions of Brassicaceae *ABS-like* genes s.s. translationally fused to the *GUS* gene. (*B*) Schematic drawing of the relationship of species and GUS staining pattern in four floral stages of the *A. thaliana* plants carrying the constructs in *A.* (*C*) GUS staining pattern of *A. thaliana* lines with only the PURR of Brassicaceae *ABS*-like genes s.s. driving GUS expression. (*D*) Schematic drawing of the relationship of species and GUS staining pattern in four floral stages of the *A. thaliana* plants carrying the constructs in *C.* (*E*) GUS staining of ovules at anthesis (S13). Stages according to Smyth et al. (1990): S5, carpel initiation; S8, carpel elongation; S10, ovule initiation; and S12, ovule development completed. Scale bar is 200 μm; black arrowheads emphasize stained tissue especially mentioned in the text, yellow arrowheads point to the chalazal end, and red arrowheads point to the micropylar end.

throughout the ovule. Differences between lines including and those lacking the transcribed sequences were observed for all constructs, although the absence of the *ABS*-like s.s. transcribed regions led to expanded expression domains in *ABS* and *AaABS* CRS containing lines. The GUS staining in pCrABS:CrABS:GUS and pAaABS:ABS:GUS plants was limited to the chalazal end. In contrast, pCrABS:GUS lines showed additional GUS staining at the micropylar end of the ovules, while pAaABS:GUS lines showed also GUS staining in the integument. Both pEsABS constructs showed GUS staining in the chalazal and the micropylar region as well as in the inner integument layers, while the expression within the integument layers appeared patchy in the pEsABS:GUS lines.

Similar constructs for *GOA*, *CrGOA*, and *AaGOA* did not show observable staining in any of the lines corroborating the extremely weak expression of these genes already reported by Hoffmeier et al. (2018).

We were then interested in the contribution of ECR1, ECR2, and the transposon insertion of the PURR of *ABS* to the expression of *ABS*. We thus designed pABS:GUS fusion constructs either lacking ECR1 (ΔECR1) or the insertion (ΔI) and a third construct carrying only the iP including ~830-bp upstream of the transcription start site (fig. 4*A*) and analyzed them in stage 12 flowers. The

full PURR of *ABS* in pABS:GUS drives GUS expression in the entire ovule, with the strongest staining at the chalazal end (fig. 4*B*). Interestingly, removal of ECR1 led to a restriction of GUS staining to a narrow region at the chalaza (fig. 4*C*). In contrast, the removal of the MITE transposon leads to a substantial expansion of expression, with GUS staining present throughout the ovule, also including a larger portion of the integuments, and in speckles in the pedicels (fig. 4*D*). Further, GUS staining is almost abolished when its expression is driven from the iP only, indicating an important role for ECR2 in activating *ABS* expression (fig. 4*E*). Four iP plants showed staining of stigmatic tissue and/or the pedicel junction as of stage 13. Two other iP plants exhibited stained anthers. These six plants taken together were regarded as consistently stained, because of their lack of GUS staining in ovules (table 1, 60%), and four iP plants showed no staining. The ABS PURR analysis shows that ECR1 activates expression in the endothelium and ECR2 activates expression at the chalazal end. Conversely, the MITE insertion represses expression of ABS in the nucellus and the integuments in ovules and in the pedicel (fig. 4*F*).

In summary, we showed that GUS staining is quite similar both across independent transgenic lines and across the reporters representing different species. Further, we show consistent GUS staining in ovules from stages 10 to 12 in all transgenic lines, with construct-dependent variation of GUS staining in earlier floral stages. GUS staining in ovules at anthesis varies substantially between the different transgenic lines, indicating species-specific fine-tuning of ABS expression in ovules before fertilization. Interestingly, the *A. thaliana*-specific MITE insertion restricts GUS staining to the ovules.

Discussion

In this study, we combine in planta analysis of PURRs with in silico TFBS detection of *ABS*-like genes s.l. across Brassicaceae. Our work provides evidence that conserved sequence blocks drive highly conserved patterns of gene expression. Our data show that some aspects of *ABS*-like gene s.s. expression in tissues and developmental stages are not phylogenetically conserved. Conversely, we find that the PURRs of all *ABS*-like genes in the species under investigation drive expression in *A. thaliana* ovules. However, the sequence conservation of their PURRs is not excessively high. We thus suggest an evolutionary dynamic landscape with a capacity for robust expression in tissues and developmental stages relevant to the functions of *ABS*-like genes s.s.

The Value of In Silico Predicted TFBSs

Our analysis of ECRs in the PURRs of Brassicaceae B_{sister}-like genes assumes that they serve as regulatory sequences for the downstream positioned B_{sister}-like genes and, in addition, assumes that the Brassicaceae TFs have not changed their DNA-binding motifs. Further, although we find conserved sequence blocks in the promoters of *ABS*-like genes s.s. (ECR1 and 2), their function remains unclear. PURRs of *ABS*-like genes s.s. are almost double the size than those of the *GOA*-like genes (fig. 1), and they show better conservation. Previously published work showed that *GOA*-like genes are expressed at a substantially lower level than *ABS*-like genes s.s. (Hoffmeier et al. 2018), and the GUS reporter gene assays carried out here corroborate these findings, because GUS staining could not be observed when the GUS expression was driven by the PURRs of *GOA*-like genes.

Interestingly, the number of different in silico predicted TFBS is roughly similar between the *ABS*- (s.s.) and the *GOA*-like genes, and 247 TFBSs are shared between *ABS*- and *GOA*-like genes. The extremely low level of expression of *GOA*-like genes suggests that those 247 are nonfunctional or may act as binding sites for repressors of expression.

CRE Evolution Allows Precise Expression Boundaries and Nonconserved Gene Expression Patterns

A typical question addressed in evolutionary studies is "which changes in regulatory machinery explain differences in gene expression levels or phenotypes?". In contrast, our work elucidates how regulatory mechanisms remain stable in evolutionary times despite substantial changes to the regulatory sequences and hence provides experimental evidence of selection acting to stabilize gene expression as hypothesized by Hodgins-Davis et al. (2015). Ovule and seed development and fertilization are essential developmental processes, ensuring reproductive success of species. It is reasonable to assume that selection acts strongly against changes in protein coding and expression of genes involved in the formation of these organs. However, mutants occur frequently and randomly in evolution, and negative selection removes unfavorable mutations from populations. In coding sequences, strong negative selection can be detected easily, but detecting selective signatures in regulatory sequences is more difficult to achieve. Previous work has suggested that small changes in regulatory state can cause massive phenotypic divergence, even though divergence of gene expression may be subtle (Romero et al. 2012). Our data provide evidence that genes encoding essential TFs like *ABS* orthologs maintain highly conserved expression pattern correlating to the tissue in which they exert their function. In addition, the CRE readout seems to allow some additional expression in tissues in which *ABS*-like genes s.s. are likely to be nonfunctional. This extra expression is species specific (fig. 3*B* and *D* and table 1) but may not be harmful and is thus not selected against.

A Transposon Insertion Contributes to Spatial Precision of *ABS* Expression in *A. thaliana*

It was described previously that *ABS* in *A. thaliana* contributes to proper ovule and embryo formation (Ehlers et al. 2016; Xu et al. 2016) and its expression at the chalazal end of the ovule (Ehlers et al. 2016) is in line with these

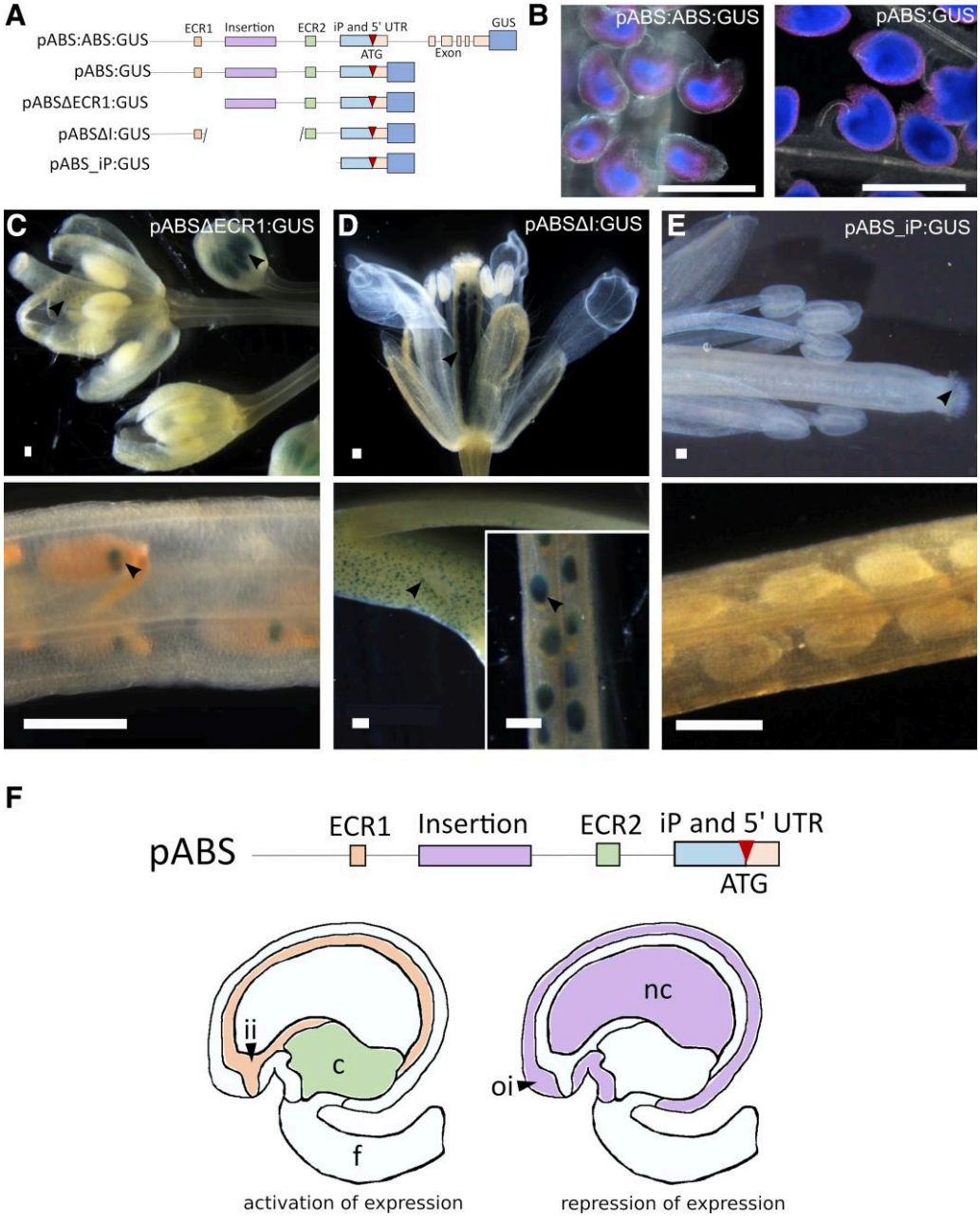


FIG. 4. GUS expression driven by different deletion constructs of the *A. thaliana ABS* PURR. (*A*) Schematic representation of the GUS reporter constructs. (*B*) Ovules at anthesis of plants containing pABS:ABS:GUS and pABS:GUS. (C, D) GUS reporter staining is shown in overview (top) and close up (bottom) of ovules during early embryo development. GUS reporter lines carrying the PURR of *ABS* lacking (*C*) ECR1 or (*D*) the transposon insertion. (*E*) GUS expression driven by the iP lacking ECR1, ECR2, and the transposon insertion. (*F*) Schematic model of transcriptional activation and repression by ECR1 (orange), ECR2 (green), and the insertion (purple). Black arrows indicate staining; the scale bar is 200 μm (ii, inner integument; oi, outer integument; c, chalazal region; nc, nucellus; f, funinculus).

functions. The contribution of B_{sister} genes to ovule development is also found in, for example, the monocot species barley and Phalaenopsis orchid (Shen et al. 2021; Shoesmith et al. 2021), suggesting that also gene expression in ovules is conserved at least since the separation of the monocot from the eudicot lineage.

Sequence analysis of *ABS*-like gene s.l. PURRs revealed a lineage-specific insertion of a MITE into the PURR of *ABS* (fig. 1). MITEs are DNA transposons and are known to facilitate the distribution of preexisting CREs throughout genomes. For example, concerted and independent evolution of C_4 from C_3 photosynthesis in more than 60 plant lineages (Sage et al. 2011) is associated with many regulatory motifs that are located in TEs, especially in MITEs (Cao et al. 2016; Qiu and Köhler 2020). Moreover, TE-localized regulatory sequences are also shown to synchronize metabolic pathways by providing binding sites for key regulatory TFs (Barco et al. 2019).

MITEs are also known to repress expression of nearby genes, and this seems to be correlated with the presence

of a large number of MITE-derived small RNAs. Transcripts of genes close to MITE insertion may become potential targets of MITE-derived small RNAs. However, they may also face downregulation through siRNA pathways triggered by MITE methylation that has been shown to account for expression differences between closely related species (Hollister and Gaut 2009; Hollister et al. 2011; Lu et al. 2012). Hence, the MITE insertion may contribute to the spatial restriction of *ABS* expression, and consequently, the lack of this MITE may lead to a broader GUS staining, as observed as speckles on the inflorescence stem (fig. 4*D*).

Arabidopsis thaliana is the only Brassicaceae studied in this work whose *ABS* regulators show an enrichment in the GO term "regulation of anthocyanin metabolism" (fig. 2*B* and *C*). Zooming into the regulatory regions specifically enriched in this term, the transposon insertion was identified as the major source of binding sites for TFs involved in anthocyanin metabolism regulation (fig. 2*B* and *C*). Among the TFs with predicted binding sites in the transposon are PIF3 and TT8, which both have been reported to directly bind to the PRGs of anthocyanin biosynthesis genes and/or regulators (Baudry et al. 2004; Shin et al. 2007). Endothelium and chalaza are tissues that regularly express anthocyanin-related genes, such as TT8 (Xu et al. 2014). This suggests that *ABS* expression might be influenced by regulators of the anthocyanin biosynthesis pathway whose binding sites are absent from the PURRs of the other Brassicaceae. Interestingly, anthocyanin biosynthesis is regulated by a genetic network including also LEC1, a protein shown to interact with pABS by yeast 2-hybrid (Y2H) screen and ABI5, which is known to regulated seedling response to ABA (Parcy et al. 1997; Brocard-Gifford et al. 2003). TFBSs recognized by these proteins are distributed along all ABS-like genes s.s. PURRs, suggesting genetic regulation of a seed development/maturation module including ABI5 and LEC1/2 for all Brassicacea under analysis, with MITE-derived TT8 regulation being specific for ABS from *A. thaliana*.

Possibly, the insertion of the TE into one site of the PURR of *ABS* restricting expression to the ovule has allowed other regions of the PURR to evolve under relaxed selection after this insertion event. The anthocyanin-related CREs in the TE may contribute to the tight spatial control to the ABS expression pattern, which, in the other Brassicaceae PURRs, may be achieved by nonanthocyanin related regulators. Thus, the MITE insertion in *A. thaliana* may balance expression in a complex interplay between *trans*-acting factors, CREs in ECR1, ECR2, and the exons and introns of *ABS*.

Our comparative study has focused on *cis*-regulatory mechanisms driving transcriptional gene regulation, but many other molecular processes regulate gene expression posttranscriptionally and posttranslationally. Further, our data show that the entire genomic locus including all exons and introns is involved in restricting expression patterns to specific parts of the ovule. These factors also deserve to be addressed in the light of our question regarding the maintenance of stable gene expression through evolutionary time. However, while genome, pangenome, and transcriptome sequencing is well established throughout land plants (Bayer et al. 2020), blind spots in our insights result from a lack of data from methylomes, even from different tissues of *A. thaliana*, or chromatin states of divergent plant species and tissue types. Further, we cannot rule out that *trans*-acting factors have changed their binding preferences since the species analyzed here diverged ~23 Ma (Hohmann et al. 2015). However, a large number of *trans*-acting factors seem conserved between species and using a single transgenic host species for CRE analysis of relatively closely related species may thus be a good start (Wray 2007). Moreover, it seems that, as species diverge in time, *cis*-regulatory differences become the dominant type of regulatory differences between eukaryotic species and when *cis*- and *trans*-regulatory differences accumulate on a gene they may act in a compensatory way to stabilize gene expression (Metzger et al. 2017). The maintenance of *ABS*-like gene s.s. expression in the chalazal region of Brassicaceae ovules might thus be a paradigm for stabilizing selection on temporally and spatially precise gene expression. Furthermore, the MITE insertion introducing anthocyanin-related CREs into the PURR or *A. thaliana ABS* may serve as an example how TE insertion provides a novel combination of CREs involved in spatial restriction of expression.

Materials and Methods

Sequence Analysis

Putative TFBS were identified using the PRG Analysis and ChIP-Seq Search tools of PlantPan3.0 (Chow et al. 2019) and PlantPan2.0 (Chow et al. 2016) respectively using the PURR sequences (starting from the translation start site to the 3′ end of the upstream gene) of *ABS*-like genes s.l. as input. The output of PlantPan consists of a list linking the identified motif, the motif position within the input sequence, the TF binding to the motif, and the TF's protein family. Sequence information of *A. thaliana* was retrieved from tair (Berardini et al. 2015). Phylogenetic shadowings were calculated using MULTILAGAN (Brudno et al. 2003) with a cutoff of 50% (supplementary fig. S1, Supplementary Material online) or 60% (fig. 1). Sequences were aligned using MAFFT (Katoh and Standley 2013) with standard settings for DNA sequences. GO enrichment analysis was carried out using PANTHER GENEONTOLOGY 15.0 (overrepresentation test released 20200407) (Mi et al. 2019) using the *A. thaliana* background against unrated lists of TFs consisting of the respective PlantPan output data ($P < 0.05$, binominal test, and Bonferroni correction), Annotation Version Gene Onthology 2018-12-01 (Carbon and Mungall 2023), and GO biological process complete. Dot plots were generated using Yass (Noé and Kucherov 2005). Absolute and relative expression data of ChIP-Seq verified factors was obtained from the eFP Browser (Winter et al. 2007). Venn diagrams of absent/present TFBS were generated using

InteractiVenn (Heberle et al. 2015). Coexpression clusters of identified TFs were calculated using UPGMA method and Spearman correlation distances of the Cluster function of Babelomics 5 with relative expression data (Alonso et al. 2015). Semantic spaces using single representation lists of TFs were generated with ReviGO (Supek et al. 2011) using SimRel to measure semantic similarity against the complete *A. thaliana* background.

Plant Material and Growth Parameters

Arabidopsis thaliana ecotype Col-0, *C. rubella* ecotype Monte Gargano (MTE, NASC:N9609), *A. alpina* ecotype Dorfertal (Dor) (Albani et al. 2012), and *E. salsugineum* plants, as well as the various GUS reporter lines and deletion lines generated in this study (table 1 and supplementary table S2, Supplementary Material online), were grown in a peat-based potting soil (ED73) with perlite (H. Nitsch & Sohn GmbH & Co.KG, Kreuztal-Eichen, Germany). The plants were grown in a greenhouse under long day growth conditions (16-h light and 8-h dark) with light varying between 120 and 180 μmol m^{-2} s^{-1} photons. The temperature in the green house was 22 °C (day) and 18 °C (night) with a relative humidity of ~60%.

Vector Cloning and GUS Staining

Genomic DNA was extracted from Brassicaceae leaf or floral tissue using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. PURRs of *ABS*, *CrABS*, *AaABS*, and *EsABS* were amplified from respective genomic DNA using Phusion high fidelity polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany) and cloned into the pDONR221 plasmid; GUS was then assembled in frame with part of the first exon in the pGWB633 plasmid. Another set of constructs was prepared where the PURRs and transcribed sequences of *ABS*, *CrABS*, *AaABS*, and *EsABS* were amplified and GUS was fused in frame with part of the last exon in the pGWB633 plasmid (Nakamura et al. 2010) with phosphinothricin acetyltransferase resistance marker for selecting transformed plants. The cloning was performed using Gateway recombination system (Life technologies GmbH, Darmstadt, Germany) according to the manufacturer's instructions.

Sequential PURR deletion constructs were generated for ABS, lacking the conserved CREs in ECR1 (pABSΔECR1:GUS) or the insertion (pABSΔI:GUS). A third construct was generated using overlapping PCR and the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs GmbH, Frankfurt, Germany) to create a construct lacking both conserved Blocks and the TE (pABS_iP:GUS). All three constructs were cloned into the pGWB633 plasmid (Nakamura et al. 2010), and GUS was fused in frame with 24 nt of the ABS MADS-box. Sequences, constructs and primers are listed in supplementary tables 2–4, Supplementary Material online.

Arabidopsis thaliana Col-0 plants were transformed by *Agrobacterium tumefaciens*-assisted floral dipping (Davis et al. 2009), with the addition of 100 μM acetosyringone and 25 g/L of sucrose to the inoculation medium.

GUS staining was done according to (Weigel and Glazebrook 2002). GUS stained tissue was cleared according to (Yadegari et al. 1994), observed using Leica M165C stereoscope and Leica DM5500 B microscope using dark field technique and photographed with Leica DFC450 color camera (Leica Microsystems, Wetzlar, Germany). The yeast 1-hybrid assay was carried out as described previously (Gross and Becker 2021) using ECR2 of the ABS PURR from *A. thaliana* that was cloned into the yeast 1-hybrid vector pAbAi (Takara Clontech, Saint-Germain-en-Laye, France) via HindIII and XhoI restriction sites.

Author Contributions

G.T. and A.B. conceived the project; C.R. analyzed the data; A.S.B. provided and analyzed the data; A.H. provided and analyzed the data; J.S. provided the data; T.G. provided the data; L.G. analyzed the data; C.R. and A.B. wrote the manuscript; G.T. and L.G. edited the manuscript; and all authors read and approved the final version of the manuscript.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Dietmar Haffer for his help in growing plants and Klaus Mummenhoff for helpful discussions throughout the project. This work was supported by the German Research Foundation (DFG) (BE2547/9-1 to A.B. and TH 417/9-1 to G.T.) within the SPP 1529 "Evolutionary plant solutions to ecological challenges: Molecular mechanisms underlying adaptive traits in the Brassicaceae s.l. (Adaptomics)".

References

Albani MC, Castaings L, Wötzel S, Mateos JL, Wunder J, Wang R, Reymond M, Coupland G. 2012. PEP1 of *Arabis alpina* is encoded by two overlapping genes that contribute to natural genetic variation in perennial flowering. *PLoS Genet.* **8**(12):e1003130.

Alonso R, Salavert F, Garcia-Garcia F, Carbonell-Caballero J, Bleda M, Garcia-Alonso L, Sanchis-Juan A, Perez-Gil D, Marin-Garcia P, Sanchez R *et al.* 2015. Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic Acids Res.* **43**(W1):W117–W121.

Barba-Montoya J, Dos Reis M, Schneider H, Donoghue PCJ, Yang Z. 2018. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous terrestrial revolution. *New Phytol.* **218**(2):819–834.

Barco B, Kim Y, Clay NK. 2019. Expansion of a core regulon by transposable elements promotes *Arabidopsis* chemical diversity and pathogen defense. *Nat Commun.* **10**(1):3444.

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J Cell Mol Biol.* **39**(3):366–380.

Bayer PE, Golicz AA, Scheben A, Batley J, Edwards D. 2020. Plant pan-genomes are the new reference. *Nat Plants*. **6**(8):914–920.
Becker A, Kaufmann K, Freialdenhoven A, Vincent C, Li M-A, Saedler H, Theissen G. 2002. A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol Genet Genom*. **266**(6):942–950.
Becker A, Saedler H, Theissen G. 2003. Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. *Dev Genes Evol*. **213**(11):567–572.
Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. 2015. The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis (New York, N.Y. : 2000)* **53**(8):474–485.
Brocard-Gifford IM, Lynch TJ, Finkelstein RR. 2003. Regulatory networks in seeds integrating developmental, abscisic acid, sugar, and light signaling. *Plant Physiol*. **131**(1):78–92.
Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. 2003. LAGAN and multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res*. **13**(4):721–731.
Cao C, Xu J, Zheng G, Zhu X-G. 2016. Evidence for the role of transposons in the recruitment of *cis*-regulatory motifs during the evolution of C4 photosynthesis. *BMC Genomics* **17**:201.
Carbon S, Mungall C. 2023. Gene ontology data archive (2023-05- 10) [2018-12-01]. *Zenodo*. https://doi.org/10.5281/zenodo.7942786
Chen J, Hu Q, Zhang Y, Lu C, Kuang H. 2014. P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res*. **42**(Database issue):D1176–D1181.
Chow C-N, Lee T-Y, Hung Y-C, Li G-Z, Tseng K-C, Liu Y-H, Kuo P-L, Zheng H-Q, Chang W-C. 2019. PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Res*. **47**(D1): D1155–D1163.
Chow C-N, Zheng H-Q, Wu N-Y, Chien C-H, Huang H-D, Lee T-Y, Chiang-Hsieh Y-F, Hou P-F, Yang T-Y, Chang W-C. 2016. PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res*. **44**(D1):D1154–D1160.
Clark JW, Donoghue PCJ. 2018. Whole-genome duplication and plant macroevolution. *Trends Plant Sci*. **23**(10):933–945.
Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**(6339):31–37.
Davis AM, Hall A, Millar AJ, Darrah C, Davis SJ. 2009. Protocol: streamlined sub-protocols for floral-dip transformation and selection of transformants in *Arabidopsis thaliana*. *Plant Methods* **5**:3.
Edger PP, Pires JC. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res*. **17**(5):699–717.
Ehlers K, Bhide AS, Tekleyohans DG, Wittkop B, Snowdon RJ, Becker A. 2016. The MADS box genes ABS, SHP1, and SHP2 are essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in *Arabidopsis thaliana*. *PloS One* **11**(10):e0165075.
Erdmann R, Gramzow L, Melzer R, Theissen G, Becker A. 2010. GORDITA (AGL63) is a young paralog of the *Arabidopsis thaliana* B_{sister} MADS box gene ABS (TT16) that has undergone neofunctionalization. *Plant J Cell Mol Biol*. **63**(6):914–924.
Gross T, Becker A. 2021. Transcription factor action orchestrates the complex expression pattern of CRABS CLAW in *Arabidopsis*. *Genes (Basel)* **12**:1663.
Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D, Ma T, Hu Q, Al-Shehbaz IA *et al.* 2017. Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* **18**(1):176.
Hajheidari M, Huang S-SC. 2022. Elucidating the biology of transcription factor-DNA interaction for accurate identification of *cis*-regulatory elements. *Curr Opin Plant Biol*. **68**:102232.
Heberle H, Meirelles GV, Da Silva FR, Telles GP, Minghim R. 2015. Interactivenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**:169.
Hodgins-Davis A, Rice DP, Townsend JP. 2015. Gene expression evolves under a house-of-cards model of stabilizing selection. *Mol Biol Evol*. **32**(8):2130–2140.
Hoffmeier A, Gramzow L, Bhide AS, Kottenhagen N, Greifenstein A, Schubert O, Mummenhoff K, Becker A, Theißen G. 2018. A dead gene walking: convergent degeneration of a clade of MADS-box genes in crucifers. *Mol Biol Evol*. **35**(11):2618–2638.
Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell*. **27**(10):2770–2784.
Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res*. **19**(8): 1419–1428.
Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A*. **108**(6):2322–2327.
Jiang P, Rausher M. 2018. Two genetic changes in *cis*-regulatory elements caused evolution of petal spot position in Clarkia. *Nat Plants*. **4**(1):14–22.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. **30**(4):772–780.
Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J Cell Mol Biol*. **88**(6):1058–1070.
Kusters E, Della Pina S, Castel R, Souer E, Koes R. 2015. Changes in *cis*-regulatory elements of a key floral regulator are associated with divergence of inflorescence architectures. *Development (Cambridge, England)* **142**(16):2822–2831.
Lovisetto A, Guzzo F, Busatto N, Casadoro G. 2013. Gymnosperm B-sister genes may be involved in ovule/seed development and, in some species, in the growth of fleshy fruit-like structures. *Ann Bot*. **112**(3):535–544.
Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. 2012. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol Biol Evol*. **29**(3):1005–1017.
Metzger BPH, Wittkopp PJ, Coolon JD. 2017. Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol Evol*. **9**(4):843–854.
Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, Thomas PD. 2019. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc*. **14**(3):703–721.
Nakamura S, Mano S, Tanaka Y, Ohnishi M, Nakamori C, Araki M, Niwa T, Nishimura M, Kaminaka H, Nakagawa T *et al.* 2010. Gateway binary vectors with the bialaphos resistance gene, bar, as a selection marker for plant transformation. *Biosci Biotechnol Biochem*. **74**(6):1315–1319.
Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L. 2002. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* **14**(10):2463–2479.
Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Bailey CD, Tsiantis M. 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol*. **222**(3):1638–1651.
Noé L, Kucherov G. 2005. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res*. **33**(Web Server issue): W540–W543.
Ohno. 1970. Evolution by gene duplication. 1970.
Panchy N, Lehti-Shiu M, Shiu S-H. 2016. Evolution of gene duplication in plants. *Plant Physiol*. **171**(4):2294–2316.

Parcy F, Valon C, Kohara A, Miséra S, Giraudat J. 1997. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of *Arabidopsis* seed development. *Plant Cell* **9**(8):1265–1277.

Prasad K, Zhang X, Tobón E, Ambrose BA. 2010. The *Arabidopsis* B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. *Plant J Cell Mol Biol*. **62**(2):203–214.

Qiu Y, Köhler C. 2020. Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. *Biochem Soc Trans*. **48**(3):1005–1017.

Rensing SA. 2014. Gene duplication as a driver of plant morphogenetic evolution. *Curr Opin Plant Biol*. **17**:43–48.

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet*. **13**(7):505–516.

Sage RF, Christin P-A, Edwards EJ. 2011. The C(4) plant lineages of planet Earth. *J Exp Bot*. **62**(9):3155–3169.

Shen C-Y, Chen Y-Y, Liu K-W, Lu H-C, Chang S-B, Hsiao Y-Y, Yang F, Zhu G, Zou S-Q, Huang L-Q *et al*. 2021. Orchid B_{sister} gene PeMADS28 displays conserved function in ovule integument development. *Sci Rep*. **11**(1):1205.

Shin J, Park E, Choi G. 2007. PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in *Arabidopsis*. *Plant J Cell Mol Biol*. **49**(6):981–994.

Shoesmith JR, Solomon CU, Yang X, Wilkinson LG, Sheldrick S, van Eijden E, Couwenberg S, Pugh LM, Eskan M, Stephens J *et al*. 2021. APETALA2 functions as a temporal factor together with BLADE-ON-PETIOLE2 and MADS29 to control flower and grain development in barley. *Development (Cambridge, England)* **148**(5):dev194894.

Smyth DR, Bowman JL, Meyerowitz EM. 1990. Early flower development in *Arabidopsis*. *Plant Cell* **2**(8):755–767.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**(7):e21800.

Theißen G, Becker A. 2004. Gymnosperm orthologues of class B floral homeotic genes and their impact on understanding flower origin. *CRC Crit Rev Plant Sci*. **23**(2):129–148.

Walden N, German DA, Wolf EM, Kiefer M, Rigault P, Huang X-C, Kiefer C, Schmickl R, Franzke A, Neuffer B *et al*. 2020. Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. *Nat Commun*. **11**(1):3795.

Weigel D, Glazebrook J. 2002. *Arabidopsis: A Laboratory Manual*. New York (NY): Cold Spring Harbor Laboratory Press.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PloS One* **2**(8):e718.

Winter KU, Becker A, Münster T, Kim JT, Saedler H, Theissen G. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc Natl Acad Sci U S A*. **96**(13):7342–7347.

Wray GA. 2007. The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet*. **8**(3):206–216.

Xu W, Bobet S, Le Gourrierec J, Grain D, De Vos D, Berger A, Salsac F, Kelemen Z, Boucherez J, Rolland A *et al*. 2017. TRANSPARENT TESTA 16 and 15 act through different mechanisms to control proanthocyanidin accumulation in *Arabidopsis* testa. *J Exp Bot*. **68**(11):2859–2870.

Xu W, Fiume E, Coen O, Pechoux C, Lepiniec L, Magnani E. 2016. Endosperm and nucellus develop antagonistically in *Arabidopsis* seeds. *Plant Cell* **28**(6):1343–1360.

Xu W, Grain D, Bobet S, Le Gourrierec J, Thévenin J, Kelemen Z, Lepiniec L, Dubos C. 2014. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol*. **202**(1):132–144.

Yadegari R, Paiva G, Laux T, Koltunow AM, Apuya N, Zimmerman JL, Fischer RL, Harada JJ, Goldberg RB. 1994. Cell differentiation and morphogenesis are uncoupled in *Arabidopsis* raspberry embryos. *Plant Cell* **6**(12):1713–1729.