Molecular Evidence for Functional Divergence and Decay of a Transcription Factor Derived from Whole-Genome Duplication in *Arabidopsis thaliana*[1][OPEN]

Melissa D. Lehti-Shiu*, Sahra Uygun, Gaurav D. Moghe, Nicholas Panchy, Liang Fang, David E. Hufnagel[2], Hannah L. Jasicki, Michael Feig, and Shin-Han Shiu*

Department of Plant Biology (M.D.L.-S., D.E.H., S.-H.S.), Genetics Program (S.U., N.P., S.-H.S.), Department of Energy Plant Research Laboratory (S.U.), Department of Biochemistry and Molecular Biology (G.D.M., L.F., M.F.), and Department of Chemistry (M.F.), Michigan State University, East Lansing, Michigan 48824; and LaPorte High School, LaPorte, Indiana 46350 (H.L.J.)

ORCID IDs: 0000-0003-1985-2687 (M.D.L.-S.); 0000-0003-0863-0384 (S.U.); 0000-0002-8761-064X (G.D.M.); 0000-0001-7062-0318 (H.L.J.); 0000-0001-6470-235X (S.-H.S.).

Functional divergence between duplicate transcription factors (TFs) has been linked to critical events in the evolution of land plants and can result from changes in patterns of expression, binding site divergence, and/or interactions with other proteins. Although plant TFs tend to be retained post polyploidization, many are lost within tens to hundreds of million years. Thus, it can be hypothesized that some TFs in plant genomes are in the process of becoming pseudogenes. Here, we use a pair of salt tolerance-conferring transcription factors, *DWARF AND DELAYED FLOWERING1* (*DDF1*) and *DDF2*, that duplicated through paleopolyploidy 50 to 65 million years ago, as examples to illustrate potential mechanisms leading to duplicate retention and loss. We found that the expression patterns of *Arabidopsis thaliana* (*At*)*DDF1* and *AtDDF2* have diverged in a highly asymmetric manner, and *AtDDF2* has lost most inferred ancestral stress responses. Consistent with promoter disablement, the *AtDDF2* promoter has fewer predicted cis-elements and a methylated repetitive element. Through comparisons of AtDDF1, AtDDF2, and their *Arabidopsis lyrata* orthologs, we identified significant differences in binding affinities and binding site preference. In particular, an AtDDF2-specific substitution within the DNA-binding domain significantly reduces binding affinity. Cross-species analyses indicate that both *AtDDF1* and *AtDDF2* are under selective constraint, but among *A. thaliana* accessions, *AtDDF2* has a higher level of nonsynonymous nucleotide diversity compared with *AtDDF1*. This may be the result of selection in different environments or may point toward the possibility of ongoing functional decay despite retention for millions of years after gene duplication.

Plant genomes have been shaped by several rounds of whole-genome duplication (WGD), which have had a significant impact on genome stability, molecular functions, physiology, and fitness (Adams and Wendel, 2005; De Smet and Van de Peer, 2012; Moghe and Shiu, 2014). Duplicate genes arising from these WGD events as well as tandem duplication, segmental duplication, and transposition are considered the raw material for evolutionary innovation (Ohno, 1970; Zhang, 2003). After duplication, one duplicate may carry out ancestral functions while the other duplicate is freed from selection and may acquire new functions (neofunctionalization; Ohno, 1970; Force et al., 1999) or optimize existing secondary functions (escape from adaptive conflict; Hittinger and Carroll, 2007; Des Marais and Rausher, 2008). Ancestral functions may also be partitioned between duplicates, resulting in the retention of both duplicates (subfunctionalization; Force et al., 1999). There is evidence that recent duplicates undergo functional divergence and selection soon after duplication (Moore and Purugganan, 2003; Wang et al., 2013); however, the most common fate after gene duplication is loss. Although the rates of gene duplication and initial duplicate retention are high, most duplicates are silenced within a few million years (Lynch and Conery, 2000). Interestingly, there is a significant bias in the types of gene duplicates that are retained and lost (Thomas et al., 2006; Freeling, 2009). For example, essential genes with housekeeping functions tend to be single-copy genes (De Smet et al., 2013), but duplicates involved in signaling and

[1] This work was supported by the National Science Foundation (grant no. MCB–1119778 to S.-H.S.).

[2] Present address: Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011.

* Address correspondence to lehtishi@msu.edu and shius@msu.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Shin-Han Shiu (shius@msu.edu).

M.D.L.-S. and S.-H.S. conceived and designed experiments; M.D.L.-S. and H.L.J. performed experiments; M.D.L.-S., S.U., G.D.M., N.P., and D.E.H. analyzed data; L.F. and M.F. performed modeling; M.D.L.-S. and S.-H.S. wrote the article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.15.00689

transcriptional regulation are preferentially retained (Conant and Wolfe, 2008). This bias may exist because the gain of novel functions in some genes may contribute to plant adaptation, such as those involved in the perception of and response to pathogens (Hanada et al., 2008; Lehti-Shiu et al., 2009). Alternatively, retention bias may reflect the need to maintain dosage balance after duplication, especially for proteins that form complexes (Birchler and Veitia, 2007; Freeling, 2009). Finally, duplicate gene retention may occur via predominantly neutral processes, such as genomic drift, and therefore may not necessarily be the result of selection (Nozawa et al., 2007; Nei et al., 2008).

The preferential retention of transcription factors (TFs) after gene duplication (Blanc and Wolfe, 2004; Seoighe and Gehring, 2004; Maere et al., 2005; Shiu et al., 2005) and the correlation of TF expansion with critical events in the evolution of land plants suggest that the expansion of TF families may provide an adaptive benefit (Lang et al., 2010). Consistent with this, the duplication of plant TFs with roles in development has led to the evolution of novel morphologies and life histories (Xiao et al., 2008; Blackman et al., 2010; Airoldi and Davies, 2012; for review, see Rensing, 2014). In addition to developmental innovation, the expansion and divergence in TF families involved in stress responses has potentially led to the diversification of responses to different abiotic stresses (Liu et al., 1998; Haake et al., 2002; Mizoi et al., 2012; Yang et al., 2014). Studying the mechanisms underlying duplicate retention and loss can give insights into how plants adapt to abiotic stress (Pires et al., 2013). However, how TF duplication and divergence contribute to the evolution of stress response pathways is not well documented.

One potential mechanism by which TF duplication can lead to the evolution of novelty is through binding site divergence. Although gene duplication is thought to be the major mechanism for generating paralogous TFs with novel functions, such as altered binding site preference (Hoekstra and Coyne, 2007), divergence in the binding site specificity of orthologous TFs is thought to be less likely due to possible pleiotropic effects (Prud'homme et al., 2007). Supporting this, studies in yeast and animals have indicated that, although orthologous TF-DNA interactions diverge rapidly between species (Borneman et al., 2007; for review, see Dowell, 2010), this divergence appears to be largely due to changes in binding site turnover rather than changes in TF binding preference (Schmidt et al., 2010; Paris et al., 2013). There is evidence that such binding site fluidity also underlies the diversification of transcriptional networks in plants (Moyroud et al., 2011; Rosas et al., 2014). However, the assumption that orthologous TFs do not readily undergo changes in binding site preference has been challenged by recent findings. For example, orthologous *Saccharomyces cerevisiae* and *Candida albicans* mating-type α1 TFs have highly divergent recognition sequences (Baker et al., 2011), and in plants, the TF *LEAFY* may have undergone changes in DNA-binding specificity during land plant evolution without the aid of duplication (Sayou et al., 2014). Thus, studies of DNA-binding preference between both paralogous and orthologous TFs are necessary to determine the contribution of binding site divergence to functional divergence.

The APETALA2 (AP2)/ETHYLENE RESPONSE FACTOR (ERF) gene family provides a good model for studying the mechanisms underlying TF duplicate retention and how TF duplication contributes to the evolution of transcriptional responses to stress. AP2/ERF genes are characterized by the presence of a conserved 60-amino acid AP2 DNA-binding domain (Okamuro et al., 1997) and are divided into four main groups: AP2, ERF, RELATED TO ABSCISIC ACID INSENSITIVE3/VIVIPAROUS1, and DEHYDRATION RESPONSE ELEMENT BINDING (DREB; for review, see Dietz et al., 2010). While both expression divergence and binding site divergence have been documented among some AP2/ERF family members, and the importance of these factors in the functional divergence in abiotic and biotic stress response has been noted (Haake et al., 2002; Sakuma et al., 2002), no study has addressed how divergence in expression and binding site preference may lead to the retention or loss of TF duplicates derived from WGD. In this study, we examine the functional divergence of WGD-derived transcription factor paralogs using *DWARF AND DELAYED FLOWERING1* (*DDF1*) and *DDF2* (Magome et al., 2004) as examples. *DDF1* and *DDF2* are members of the DREB A-1 subfamily and are derived from the most recent α-duplication event (Bowers et al., 2003) in the Brassicaceae lineage 50 to 65 million years ago (MYA; Beilstein et al., 2010). In *Arabidopsis thaliana*, *DDF1* and *DDF2* are induced by salt and confer tolerance to salt stress when ectopically expressed (Magome et al., 2004). Together with other related TFs, they regulate responses to low temperature, drought, and high salinity (Dietz et al., 2010). Through a combination of phylogenetic analysis, expression studies, and protein-binding microarray (PBM) experiments, we assessed the functional divergence of *AtDDF1* and *AtDDF2* and their orthologs in the closely related species *Arabidopsis lyrata* to identify factors contributing to retention and functional decay.

RESULTS AND DISCUSSION

Phylogeny and Sequence Evolution among Brassicaceae DDF1 and DDF2 Orthologs

To better understand how these two duplicated genes have evolved, we first determined whether *DDF1* and *DDF2* have been retained among sequenced Brassicaceae species, including *A. lyrata*, *Capsella rubella*, *Thellungiella halophila*, and *Brassica rapa*, which diverged approximately 43 MYA (Beilstein et al., 2010). A phylogenetic tree was constructed with the syntenic DDF1 and DDF2 protein sequences, the other

four *A. thaliana* DREB A-1 genes (*C-REPEAT/DEHYDRATION RESPONSE ELEMENT BINDING FACTOR1* [*CBF1*], *CBF2*, *CBF3*, and *CBF4*), and the most closely related genes in the outgroup species *Carica papaya* and *Populus trichocarpa*. Three genes from the DREB A-4 subfamily closely related to, but distinct from, the DREB A-1 group (*HARDY*, AT1G12630, and AT5G52020) and the more distantly related *AtERF1* (AT4G17500) were included as outgroups (Nakano et al., 2006). This phylogenetic tree confirms that a Brassicaceae-specific duplication event gave rise to *DDF1* and *DDF2* (Fig. 1). *DDF1* and *DDF2* are present in single copy in all species except *B. rapa*, which has three *DDF1* paralogs but only one *DDF2* gene (Lee et al., 2012; Fig. 1). Considering that there was a genome triplication event (α') in the lineage leading to *B. rapa* (Wang et al., 2011; Fig. 1B), this pattern of *DDF1/2* retention is consistent with the finding that transcriptional regulators tend to be retained after the α WGD (Blanc and Wolfe, 2004; Seoighe and Gehring, 2004; Maere et al., 2005; Shiu et al., 2005) as well as the α' WGD (Wang et al., 2011; Moghe et al., 2014).

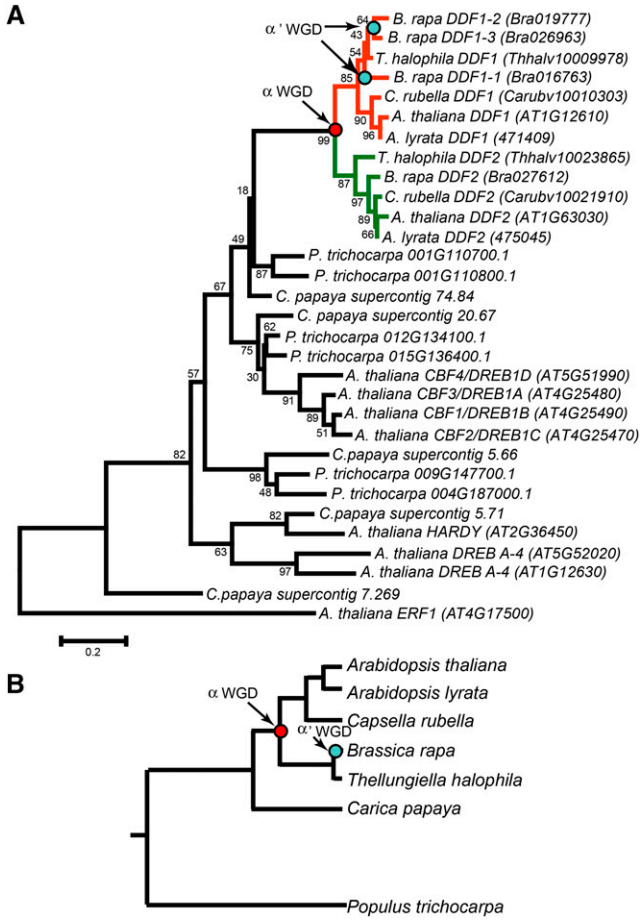

Figure 1. Relationships between *DDF1* and *DDF2* orthologous genes in the Brassicaceae. A, *DDF1/2* gene tree. Also included are *A. thaliana* DREB A-1 genes and related genes in *C. papaya* and *P. trichocarpa*, with *A. thaliana* DREB A-4 genes and *A. thaliana ERF1* as outgroups. The number at each node indicates the percentage bootstrap support based on 1,000 replicates. B, Species tree showing the relationships between the Brassicaceae species, *C. papaya*, and *P. trichocarpa*. Branch lengths do not represent the distance between species. The α (red) and α' (blue) WGD nodes are indicated.

The single-copy *B. rapa DDF2* gene is an apparent deviation from this generalization. In *Raphanus raphanistrum*, a close relative of *B. rapa* that also experienced the α' triplication, there are two *DDF1* orthologs (RrC621_p7 and RrC23381_p1), but there is no gene that is orthologous to *DDF2* (Moghe et al., 2014). There is one *Raphanus sativum* complementary DNA (cDNA) sequence that matches *DDF1* (FY449579), but none match *DDF2*. The *R. raphanistrum* genome assembly is not as complete as that of *B. rapa* and is more fragmented (Moghe et al., 2014). Therefore, we cannot rule out the possibility that *DDF2* sequences are not present in the assembly. However, given that two *DDF1* genes could be identified, it is likely that one or more *R. raphanistrum DDF2* sequences have been lost. Although the retention of both *DDF1* and *DDF2* in multiple Brassicaceae species is consistent with the hypothesis that TFs are retained due to a dosage balance requirement (Birchler et al., 2005; Freeling and Thomas, 2006; Birchler, 2012), the retention of only one *DDF2* gene in *B. rapa* and the probable loss in *R. raphanistrum* suggests that, in these species, the dosage requirement for *DDF2* was not as strong as for *DDF1*. Loss of *DDF1* and *DDF2* duplicates after triplication is also consistent with the finding that, even though there is an initial high rate of TF retention after a duplication event, in the long run, most duplicates are lost (Lynch and Conery, 2000; Hanada et al., 2008). Furthermore, gene duplicates that are lost after successive duplication events tend to be those that are lowly expressed, under less purifying selection, and theoretically contribute less to gene dosage (Schnable et al., 2012). This raises the question of whether *DDF2* fits the profile of a duplicate that is destined to be lost and whether it became less functional prior to or after the divergence of the *B. rapa* and *R. raphanistrum* lineages from the other Brassicaceae.

We next compared amino acid sequences to identify possible divergence and/or loss of function among paralogous and orthologous sequences. All Brassicaceae DDF1 and DDF2 protein orthologs contain AP2 DNA-binding domains flanked by sequences conserved among DREB A-1 proteins (Jaglo et al., 2001; Supplemental Fig. S1, blue boxes) and, with the exception of *T. halophila* DDF2, an LWNYS motif (Supplemental Fig. S1, orange box), which has transcriptional attenuation and transcriptional activation functions (Wang et al., 2005; Zhang et al., 2013). Despite the similarity among DREB A-1 proteins, all DDF2 proteins are missing a region of about 20 amino acids at the C terminus that in all the DDF1 proteins is Ser and Gly rich (Supplemental Fig. S1, green box). Because other DREB A-1 proteins have sequence in this region (Supplemental Fig. S1, green box), it is not

clear whether the Ser- and Gly-rich motif was acquired via the insertion or mutation of existing sequence and whether this occurred prior to the duplication of the ancestral *DDF* gene. The acquisition of the Ser- and Gly-rich motif in *DDF1*, or its loss in *DDF2*, may have led to altered posttranslational regulation and/or interactions with different protein partners, thereby contributing to functional divergence between DDF1 and DDF2. Thus, *DDF1* and *DDF2* may have been initially retained not only due to the requirement for dosage balance but also due to neofunctionalization, a possibility that needs to be tested experimentally. The loss of the LWNYS motif in *T. halophila* DDF2 may indicate loss of function, but even though there are several amino acid differences among orthologous sequences, including some in conserved amino acids (Supplemental Fig. S1), there is no clear evidence that DDF2 orthologs are less functional than DDF1 orthologs.

Expression Divergence between DDF1 and DDF2

Studies of *DDF1* and *DDF2* overexpression lines in *A. thaliana* have revealed that ectopic expression of both genes results in increased salt tolerance as well as a dwarf plant phenotype (Magome et al., 2004). The current model is that *DDF1* regulates the response to salt stress by decreasing plant growth through the up-regulation of *GIBBERELLIN2-OXIDASE7* (*GA2Ox7*) and related enzymes that convert GA_3 to inactive forms (Magome et al., 2008). The fold induction of *DDF2* in response to salinity stress is much lower than that of *DDF1*, despite the visually identical phenotypes and similar up-regulation of *GA2Ox7* in ectopic expression lines, and it is speculated that *DDF2* may have a minor role in salt stress response (Magome et al., 2004, 2008). In addition to salt stress, *DDF1* overexpression lines have increased tolerance to cold, drought, and heat (Kang et al., 2011), and *DDF2* has been implicated in the regulation of *HIGH-AFFINITY POTASSIUM TRANSPORTER5* in response to phosphate deficiency (Hong et al., 2013). However, these findings have yet to be corroborated with loss-of-function studies. Plants harboring a *ddf1* loss-of-function allele are available and are reported to have increased root growth under some salt concentrations (Magome et al., 2008), but no *ddf2* transfer DNA insertion lines are available.

In the absence of genetic data for *DDF2*, we turned to expression data to evaluate the extent of *DDF1* and *DDF2* functional divergence. We compared their expression profiles in the AtGenExpress development, light, and abiotic/biotic stress expression data sets (Fig. 2; Supplemental Table S1; Schmid et al., 2005; Kilian et al., 2007). For comparison, the extent of expression divergence for all WGD-derived AP2/ERF duplicates was also examined (Fig. 2A; Supplemental Table S1). The expression patterns of AP2/ERF WGD duplicates are significantly more correlated than random gene pairs ($P \leq 0.002$; Supplemental Fig. S2). This is also true for *DDF1/2* (Pearson's correlation coefficient [PCC] = 0.64, $P < 10^{-32}$), but stress-responsive expression is more highly correlated in the shoot (PCC = 0.88, $P < 10^{-9}$; Fig. 2B) than in the root (PCC = 0.45, $P = 0.012$; Fig. 2C). Based on a threshold PCC value greater than 95% of random gene pairs (PCC = 0.52 and 0.5 for the shoot and root, respectively), 15 duplicate pairs have correlated expression in both the root and the shoot, five have correlated expression only in the shoot, and 10 have correlated expression only in the root. This suggests that organ-specific responses to stress contribute to the functional divergence of some AP2/ERF duplicates, including *DDF1* and *DDF2*.

Despite the overall significant correlation in expression pattern, there are two significant differences between these two paralogs. First, *DDF1* is more highly expressed than *DDF2* in all of the AtGenExpress developmental series samples (Fig. 2D). Second, *DDF2* induction by salt, cold, and wounding is much lower compared with *DDF1* (Fig. 2, B and C). Note that the normalized array intensities for *DDF1* and *DDF2* expression in the shoot (7.72 and 2.3, respectively) and root (1.52 and 3.67, respectively) at time zero are similarly low, so the fold inductions reported in Figure 2 reflect a relative increase in expression from approximately the same baseline level of expression. The observed induction of *DDF2* by salt in the root (Fig. 2C; 1.4-fold at 1 h) is lower than that observed by Magome et al. (2004, 2008), but the overall reduced induction of *DDF2* by abiotic stress compared with *DDF1* is consistent with their studies. Taken together, the divergence pattern is largely asymmetric, with one duplicate, *DDF1*, expressed more broadly and at a higher level under both control and stressful environments. The asymmetric expression patterns of *DDF1* and *DDF2* are contrary to the subfunctionalization hypothesis (Lynch and Force, 2000), which stipulates that duplicates may be retained due to the partitioning of ancestral functions, but they are consistent with the finding that most duplicates have asymmetric expression patterns (Ganko et al., 2007; Zou et al., 2009).

Genome dominance, where one parental genome is preferentially expressed, provides one potential explanation for this asymmetry. Consistent with genome dominance, the α WGD block containing *DDF2*, A03-b, has experienced more fractionation (or gene loss) than the homologous block, A03-a, containing *DDF1* (Thomas et al., 2006). Our expectation was that the genes that are in the same block as *DDF1* might be expressed at a higher level than the genes in the homologous block. To test this hypothesis, we compared the expression levels of duplicate gene pairs in the α WGD block A03 across the AtGenExpress light, development, and stress data sets (109 gene pairs, 280 samples). We found that 48 out of 109 genes on the A03-a block had significantly higher expression levels than their duplicate pair on the A03-b block (Wilcoxon

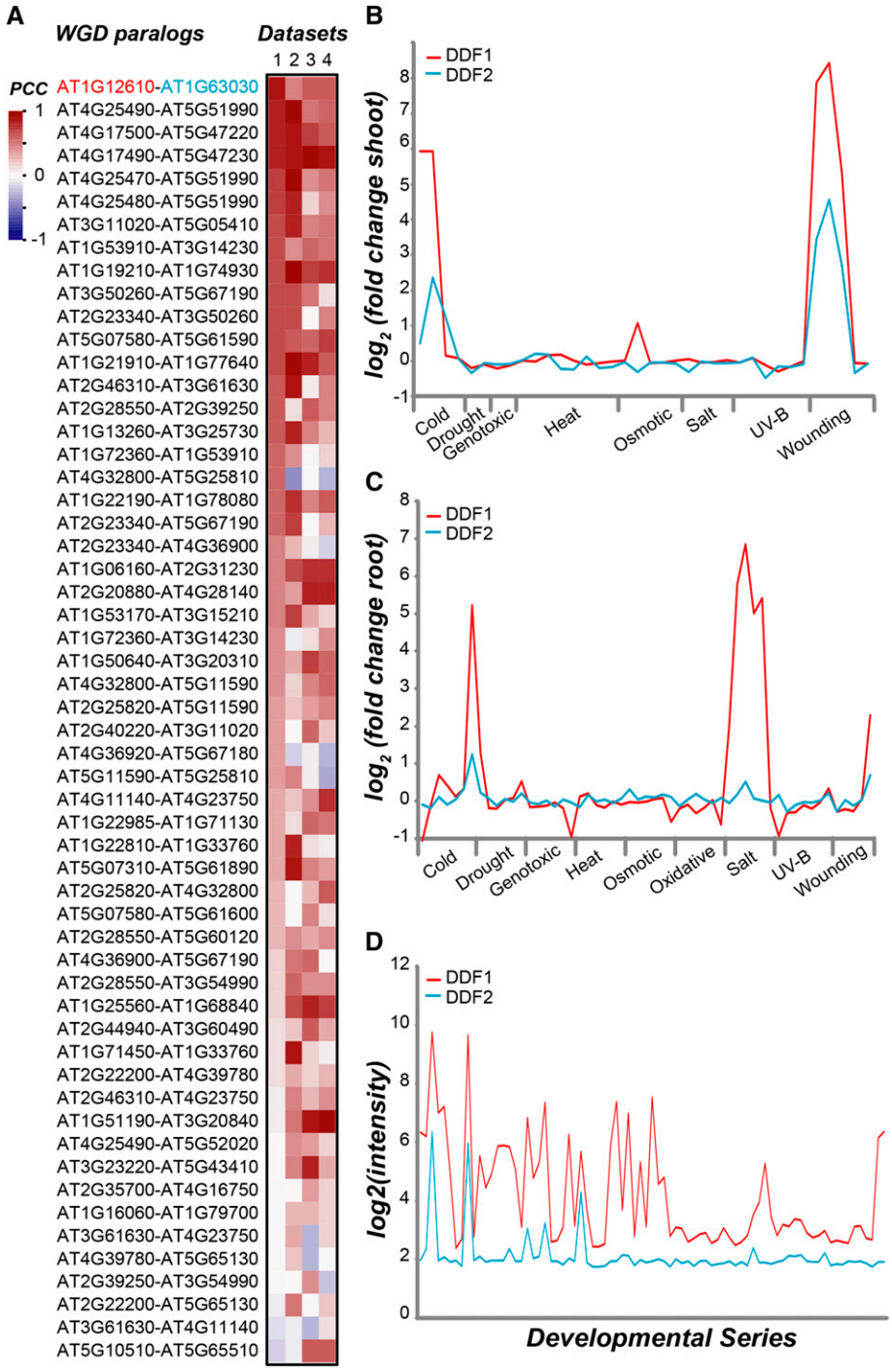

Figure 2. AP2/ERF WGD homolog expression correlation. A, Heat map of PCC values for correlated expression between AP2/ERF WGD pairs in abiotic stress-treated shoots (1), abiotic stress-treated roots (2), developmental series (3), and abiotic and biotic stress, development, and light combined (4). Dark-red and blue shading indicate high positive and negative expression correlation, respectively. White indicates low or no correlation. *DDF1* and *DDF2* are highlighted in red and blue, respectively. B, Expression profiles of *DDF1* (red line) and *DDF2* (blue line) under abiotic stress in the shoot. C, Expression profiles of *DDF1* and *DDF2* under abiotic stress in the root. D, Expression profiles of *DDF1* and *DDF2* at different developmental stages.

signed rank test, $P < 0.05$). Of the remaining 61 gene pairs, 55 show the opposite pattern, with the duplicate on the A03-b block being more highly expressed than its pair on the A03-a block ($P < 0.05$). Duplicates found on the A03-a block are not significantly more highly expressed than those on the A03-b block ($P = 0.841$, Fisher's exact test based on comparisons of duplicate pairs in all α WGD blocks), potentially due to erasure of the dominant expression signature over the past approximately 50 million years since the α event.

DDF1 and DDF2 Ancestral Expression States and Promoter Divergence

Previously, we found that duplicates that have lost stress responses are more likely to gain novel functions

(Zou et al., 2009). To determine if there are conditions where only *DDF2* is expressed, which would suggest neofunctionalization, we compared *DDF1* and *DDF2* expression profiles in the NASCArray data set, which has a greater number (4,995) of samples (Craigon et al., 2004; Supplemental Fig. S3). As in the AtGenExpress data sets, *DDF1* expression exceeds that of *DDF2* under most conditions (440 of 465 conditions where the intensity of either gene is greater than 30). In experiments done using RNA samples from *Lepidium sativum*, there is increased *DDF2* expression in the endosperm cap and radicle (Supplemental Fig. S3), but in *A. thaliana*, *DDF2* is not highly expressed (intensity $\leq$ 10.45) in the radicle (Dekkers et al., 2013) or micropylar endosperm (Le et al., 2010; Dekkers et al., 2013). The expression level of *DDF2* is 2-fold higher than that of *DDF1* in the chalazal endosperm (Le et al., 2010), and this could indicate a function for *DDF2* in the endosperm. However, it should be noted that endosperm expression may also represent general deregulation of expression due to demethylation (Gehring et al., 2009; Hsieh et al., 2009). In other cases where the magnitude and relative expression levels of *DDF2* are higher compared with *DDF1*, the pattern is not consistent between replicates. Thus, overall *DDF2* expression is reduced compared with *DDF1*, and although the possibility of a tissue-specific function for *DDF2* cannot be ruled out, there is no clear evidence for subfunctionalization or neofunctionalization at the expression level. In addition, compared with *DDF1*, *DDF2* is less responsive or is not responsive to most stress conditions. To further determine when this hypothesized reduction of stress responsiveness occurred, we looked at the stress-responsive expression of ancestral DREB A-1 genes.

Like *AtDDF1*, *B. rapa DDF1* orthologs are up-regulated by salt stress (Lee et al., 2012), indicating that either the ancestral *DDF* gene was salt responsive or that gain of salt responsiveness occurred in the *DDF1* lineage. Reconstruction of putative ancestral expression states allows the inference of gain and loss of expression among gene duplicates (Oakley et al., 2006; Zou et al., 2009; Liu et al., 2011). Therefore, to test the hypothesis that *DDF2* lost responsiveness or became less responsive to stress, we inferred the stress-response patterns of the ancestral *DDF1/2* and DREB A-1 subfamily genes in the roots and shoots (Fig. 3). Three genes from a closely related but separate subgroup, DREB A-4, were included as outgroups (Figs. 1A and 3). For this analysis, we used the AtGenExpress abiotic stress conditions from Figure 2, where DREB A-1 genes are induced under four conditions (cold, drought-root only, salt, and wounding), each with multiple time points (separated by vertical dashed lines in Fig. 3). For each condition, the expression state (up-regulated or not changed) of each ancestral node was inferred (see "Materials and Methods"). Here, responsiveness is defined as 2-fold or greater change in expression with treatment compared with the control, and loss can mean a reduction in responsiveness. Our results indicate that the *DDF1/2* ancestral gene was likely salt responsive in the root and that the salt response was lost or significantly reduced in the *DDF2* lineage, along with loss of the wounding response in the root (Fig. 3A). On the other hand, loss of the shoot salt response likely took place prior to the *DDF1* and *DDF2* divergence (Fig. 3B). Whereas *DDF1* and *DDF2* more closely resemble the ancestral gene in terms of stress response in the shoot (Fig. 3B), *DDF1* has retained more ancestral stress responses in the root (Fig. 3A). Although we cannot rule out the possibility that the full range of *DDF2* expression under abiotic stress is not represented in the AtGenExpress data set, overall, reconstruction of the

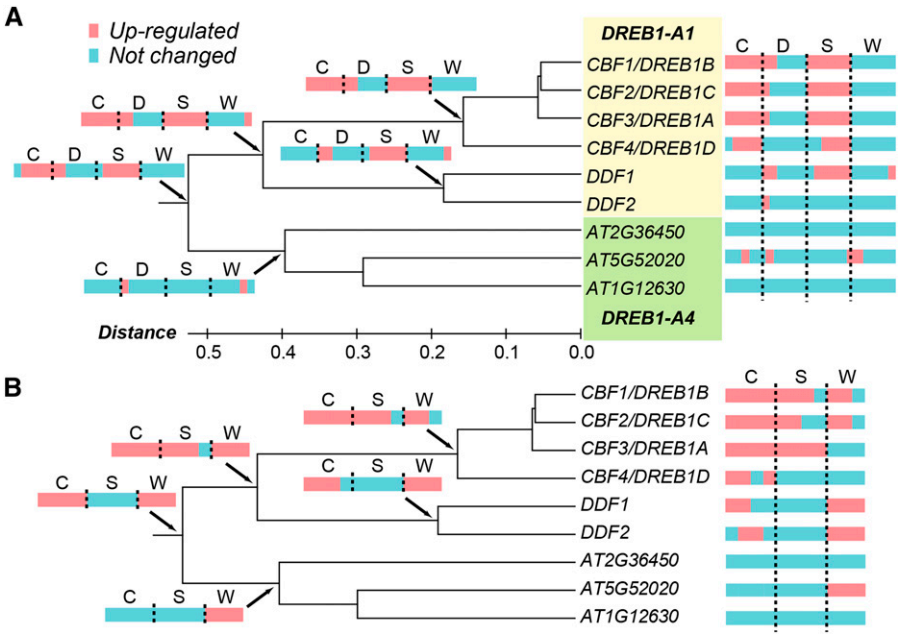

Figure 3. Ancestral stress responses in the DREB A-1 subfamily. The ancestral expression profiles of DREB A-1 subfamily genes in response to cold (C), drought (D), salt (S), and wounding (W) stress in the root (A) and shoot (B) were inferred using BayesTrait. Three DREB A-4 subfamily genes were included as outgroups. Dashed vertical lines outline the time points for each condition. Pink and blue indicate that gene expression is up-regulated and not changed, respectively.

history of the gain and loss of abiotic stress response supports the hypothesis that *DDF2* may have a less important role in abiotic stress compared with *DDF1* and, at the expression level, *DDF2* has experienced significant decay compared with the ancestral state.

If *DDF2* is experiencing a decay of expression, this may be apparent as a loss of cis-elements in the promoter region. *AtDDF1* does have more potential cis-elements from the PlantCARE database (Lescot et al., 2002) in the 1.5-kb putative promoter and 5′-untranslated regions compared with *AtDDF2* (48 and 36, respectively; Supplemental Table S2), but *AlDDF1* and *AlDDF2* have similar numbers of potential cis-elements (54 and 58, respectively). In the *AtDDF2* promoter, an annotated transposable element (TE; AT1TE76985) is located 438 bp upstream of the transcriptional start site (TSS), and the best match for this TE sequence in *A. lyrata* maps to 159 bp upstream of the predicted TSS of *AlDDF2* (68% identity). Because methylation of TEs can impact the expression of neighboring genes (Hollister and Gaut, 2009), we next looked for evidence for methylation of the *AtDDF2* promoter using published data (Stroud et al., 2013). We found that there is indeed CG methylation, not of AT1TE76985 but of a repetitive element approximately 300 bp upstream of the *AtDDF2* TSS (Maumus and Quesneville, 2014; Supplemental Fig. S4, A and B). Interestingly, 24-nucleotide small RNAs also map to this element (Lister et al., 2008; Supplemental Fig. S4B). Based on alignment with the *AtDDF2* promoter, this methylated repeat sequence is absent in the *AlDDF2* promoter. In addition, although there is a repeat 500 bp upstream of the *AtDDF1* TSS (Maumus and Quesneville, 2014; Supplemental Fig. S4D), there is no evidence of small RNA mapping or methylation (Supplemental Fig. S4, C and D). Of the other DREB A-1 genes, only *CBF3* has a methylated TE, and this is located more than 500 bp upstream of the TSS. Pseudogenes, including expression pseudogenes with functional protein sequences but low expression, are more likely to have TEs located within 500 bp (Yang et al., 2011). Thus, the presence of a TE and associated CG methylation is consistent with the idea that the promoter of *AtDDF2* may be decaying.

Assessing Paralogous and Orthologous DDF1 and DDF2 Binding Profiles

DNA-binding preference can vary widely within TF gene families, even among TFs with highly similar DNA-binding domains (Berger et al., 2008), indicating that binding site divergence can contribute to the evolution of regulatory networks. There have been broad surveys of TF-binding preference within plant TF families, including the AP2/ERF family (Sakuma et al., 2002; Godoy et al., 2011; Franco-Zorrilla et al., 2014), but no study has yet examined differences in binding site preference between paralogous and orthologous plant TFs. Here, we examine differences in binding site preference between DDF1 and DDF2 paralogs and orthologs as an example to evaluate the potential contribution of binding site divergence to functional divergence.

A universal PBM (Berger et al., 2006) was used to assay the affinities of AtDDF1 and AtDDF2 and their *A. lyrata* orthologs, AlDDF1 and AlDDF2, for all possible eight-nucleotide sequences. Affinity for each motif is reflected by an enrichment (E) score ranging from 0 to 0.5, where $E \geq 0.45$ indicates selective binding (Berger and Bulyk, 2009; for E scores and median intensities of top-scoring 8-mers, see Supplemental Table S3). All four TFs bound to the same 8-mer motif, GCCGACAT, with the highest affinity (Fig. 4; Supplemental Table S3). This motif is similar to A/GCCGAC, the drought response element (DRE)/C-repeat (Baker et al., 1994; Yamaguchi-Shinozaki and Shinozaki, 1994; Sakuma et al., 2002), and matches the DDF1 binding site identified in a recent PBM study of plant TFs (Franco-Zorrilla et al., 2014). The fact that there are 8-mers with high E scores ($E > 0.49$; Supplemental Table S3) indicates that all four proteins specifically bind DNA (Berger and Bulyk, 2009). However, although position weight matrices (PWM) generated for each protein based on the top-ranked motif are virtually indistinguishable (Fig. 4A; Supplemental Table S4), the number of high-affinity 8-mers ($E \geq 0.45$) differs substantially, even between orthologs, ranging from 70 for AtDDF2 to 2,511 for AtDDF1 (Fig. 4B). As expected, the top-ranking motifs resemble each other, because TFs can usually tolerate substitutions within the binding site (Berger and Bulyk, 2009). The difference in the number of significant 8-mers that we observed here, therefore, may reflect differences in affinity and/or tolerance to substitutions in the DNA sequence. Based on overall lower intensities and the relatively small number of high-affinity 8-mers, AtDDF2, but not AlDDF2, likely has an overall reduced affinity for DNA. The reduced affinity of AtDDF2 for GCCGACAT compared with AtDDF1 was further confirmed by electrophoretic mobility shift assay (EMSA) using the same GLUTATHIONE *S*-TRANSFERASE (GST)-AtDDF2 protein applied to the microarray (Supplemental Fig. S5) and using GST-AtDDF2 protein produced using an independent method (in vitro translation; Fig. 5) and, therefore, is not due to reduced protein concentration in our PBM study. The differences in the number of high-affinity 8-mers suggest that there are significant differences in DNA binding, not only between paralogous TFs that duplicated 50 to 65 MYA (i.e. AtDDF1 and AtDDF2) but also between orthologs that diverged more recently, approximately 13 MYA (AtDDF2 and AlDDF2; Beilstein et al., 2010).

Differences in Binding Preference

We next focused on differences in DNA-binding site preference between paralogs by examining the effect of substitutions at each site of the top-ranked motif. Consistent with previous studies examining the

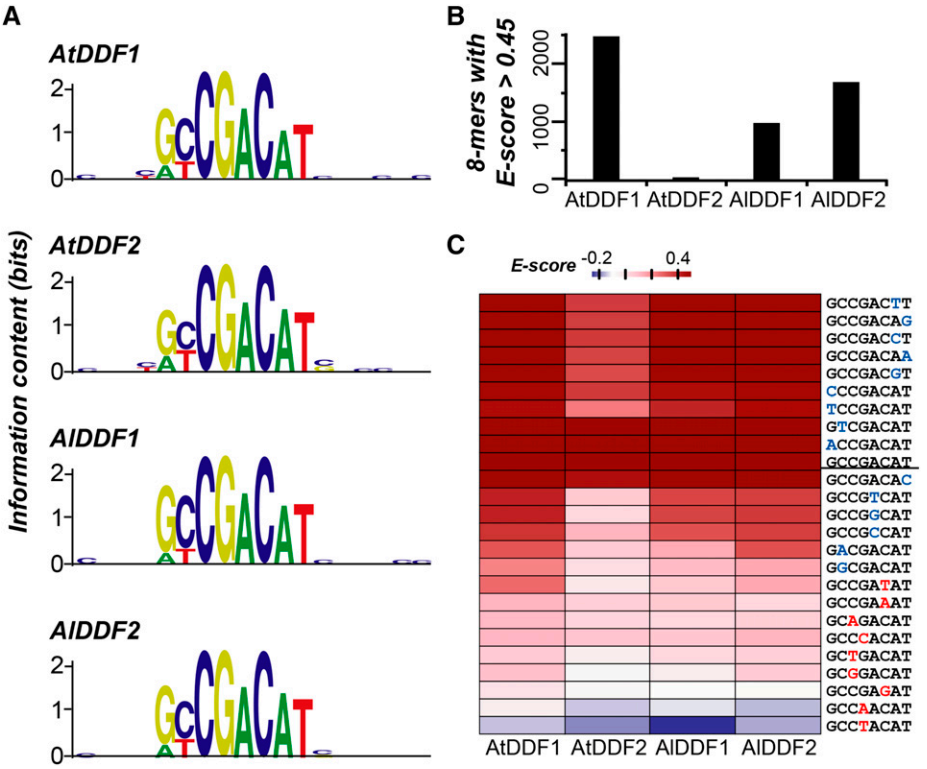

Figure 4. DDF1 and DDF2 bind to variants of the DRE. A, Sequence logo representations of binding site preferences (Workman et al., 2005). B, Number of 8-mers (E ≥ 0.45) bound by each protein. C, Effects of nucleotide substitutions on DDF1 and DDF2 binding affinity. Shading represents E scores for each 8-mer variant; darker red indicates a high E score and darker blue indicates a low E score. Nucleotides that differ from the top-ranked motif (GCCGACAT) are colored: red, substitutions at positions previously shown to be critical for binding; and blue, substitutions at positions not previously shown to be critical for binding.

binding preference of DREB A-1 and A-2 subfamily members (Sakuma et al., 2002), 8-mer motifs with substitutions that are known to be critical for DRE binding have low E scores (Fig. 4C). Although AlDDF1, AlDDF2, and AtDDF1 have similar affinities for most variants of the DRE, binding of AtDDF2 is dramatically affected even by substitutions in nucleotides not critical for binding (Fig. 4C, nucleotides highlighted in blue). Furthermore, there are no AtDDF2-specific motifs (i.e. motifs with E ≥ 0.45 that are only bound by AtDDF2; Supplemental Fig. S6) that would indicate that the binding preference of AtDDF2 has diverged from AtDDF1. To identify more subtle differences in binding site preference between DDF1 and DDF2, the affinities for the top-10 motifs for each TF were compared. Because the E scores for these top-ranked motifs are very similar, normalized hybridization intensity (ranging from 0 to 1) was calculated as a proxy for the relative affinity for each motif (Supplemental Table S3). For example, the E scores for GCCGACAT and GTCGACAT (0.4989 and 0.4968, respectively) indicate that AtDDF1 binds similarly to these two sequences, but normalized median intensities differ by 20% (1 and 0.802, respectively), consistent with a similar difference in affinity observed with EMSA (35% less binding to GTCGACAT than to GCCGACAT; $P < 0.05$; Fig. 5A). Based on the normalized median intensity measure, AtDDF1 and AtDDF2 have similar relative affinities for only five of the top ranked 8-mers, and AtDDF1 binds more motifs with high relative affinity (Fig. 6A; Supplemental Fig. S7A). In contrast, AlDDF2 binds more motifs with higher affinity than AlDDF1 (Fig. 6A; Supplemental Fig. S7B). AlDDF1 has a much higher relative affinity for GCCGACAT compared with other motifs, indicating a narrower sequence preference than the other TFs (Fig. 6A; Supplemental Fig. S7B).

We expected that DDF1 and DDF2 orthologs would have more similar binding preferences than their respective paralogs, given that the divergence time between the two species (approximately 13 MYA) is more recent compared with the α WGD event (50–65 MYA). Consistent with our expectation, the correlation between E scores is higher for DDF1 orthologs compared with their respective DDF2 paralogs (Table I). In addition, AtDDF1 and AlDDF1 both have higher affinity for motifs containing GCCGAC, whereas AtDDF2 and AlDDF2 have higher affinity for motifs containing GTCGAC (Fig. 6B; Supplemental Fig. S7). This suggests that binding site divergence and differential recognition of target genes constitute a possible mechanism for functional divergence between DDF1 and DDF2. However, despite these similarities between orthologs, when relative binding affinities are compared, AlDDF1 and AlDDF2 are more highly correlated to each other than to AtDDF1 and AtDDF2, respectively (Supplemental Table S5). In addition, the overlap in motifs bound is greater for AtDDF1 and AlDDF2 than for AtDDF1 and AlDDF1 (Supplemental Fig. S6, B and C). The higher similarity in binding between AtDDF1 and AlDDF2 is likely due to the fact that AtDDF1 and AlDDF2

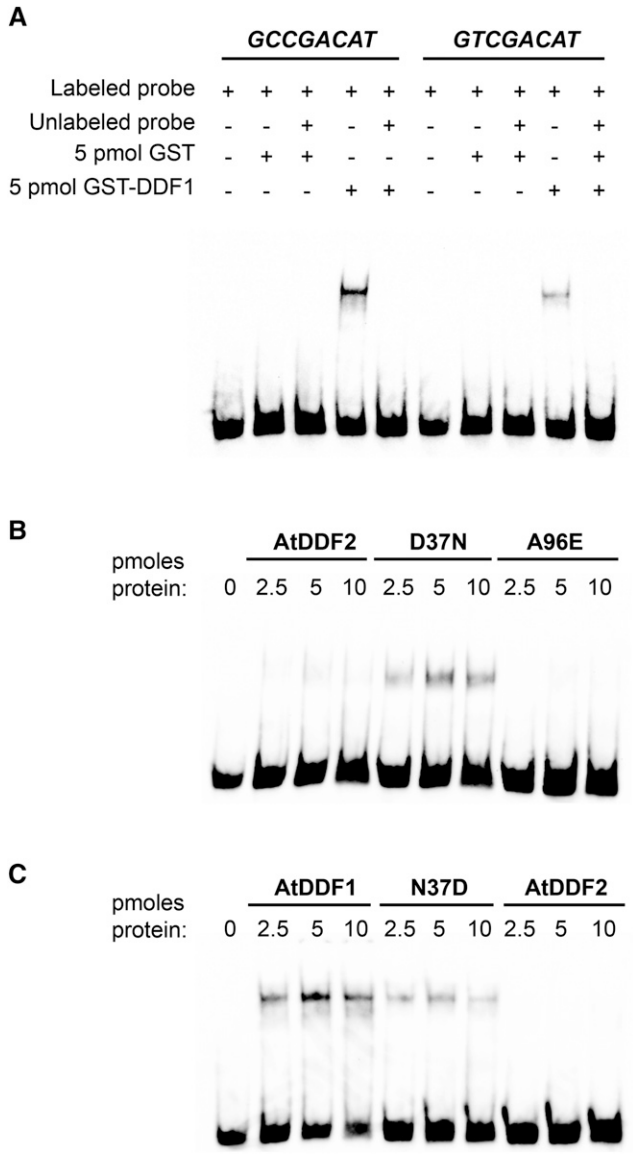

Figure 5. AtDDF1 has reduced affinity for GTCGACAT compared with GCCGACAT, and amino acid Asn-37 is important for binding site affinity. A, Five picomoles of in vitro-translated GST-AtDDF1 or GST was incubated with 40 pmol of biotin-labeled GCCGACAT or GTCGACAT probe plus or minus 8 pmol of an unlabeled competitor probe. B, Increasing amounts of GST-AtDDF2, GST-AtDDF2 D37N, and GST-AtDDF2 A96E in vitro-translated protein were incubated with 40 fmol of biotin-labeled probe containing the GCCGACAT 8-mer motif. C, Increasing amounts of GST-AtDDF1, GST-AtDDF1 N37D, and GST-AtDDF2 in vitro-translated protein were incubated with 40 pmol of biotin-labeled GCCGACAT probe. For all EMSAs, one representative blot from at least three replicates is shown.

bind a much higher number of 8-mers compared with AlDDF1 and AtDDF2.

Based on the observation that AlDDF2 has higher binding affinity than AtDDF2, we speculated that it might have retained responsiveness to salt stress. To test this, we compared *DDF1* and *DDF2* levels in 14-d-old *A. thaliana* and *A. lyrata* seedlings after 3 h of 150 mM NaCl or control treatment with levels at time zero. Consistent with previous studies, *AtDDF1* is more highly induced by salt than *AtDDF2* (588- and 56-fold higher expression relative to the water control, respectively; Supplemental Fig. S8A). Similarly, *AlDDF1* is more highly expressed under salt treatment compared with *AlDDF2* (14- and 6-fold higher expression relative to the water control, respectively; Supplemental Fig. S8B). Therefore, like *AtDDF2*, *AlDDF2* has reduced responsiveness to salinity stress, suggesting that reduced salt responsiveness predates the divergence of *A. thaliana* and *A. lyrata*. On the other hand, the significantly lower overall affinity is specific

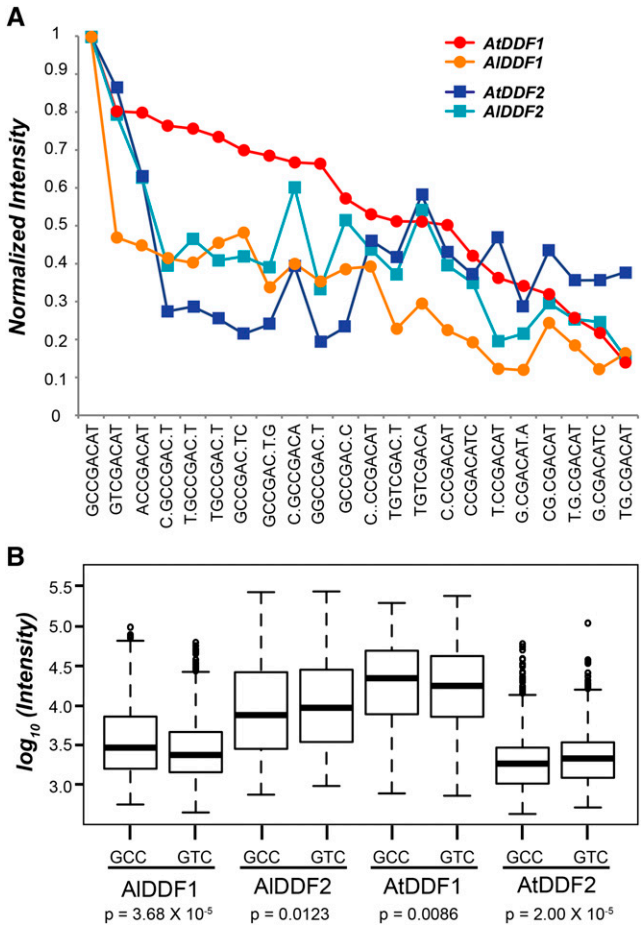

Figure 6. Divergence in binding site preference between DDF1 and DDF2 paralogs and orthologs. A, Relative intensities for motifs that are ranked in the top 10 (based on E score rankings) for at least one protein: AtDDF1 (red circles), AtDDF2 (dark-blue squares), AlDDF1 (orange circles), and AlDDF2 (light-blue squares). B, Box plots showing the distribution of $\log_{10}$ binding intensities for probes containing the 6-mer sequences GCCGAC (GCC) and GTCGAC (GTC). For each protein, Wilcoxon rank-sum tests were used to test for significant differences between median binding intensities for probes containing GCCGAC versus GTCGAC. *P* values are shown below the box plots.

Table I. *PCCs reflecting correlation between 8-mer E scores (contiguous and gapped, E ≥ 0.45 for at least one protein)*

All *P* values are 2×10^{-16} or less.

Protein	AtDDF1	AtDDF2	AlDDF1	AlDDF2
AtDDF1	1	0.6526	0.8776	0.8366
AtDDF2		1	0.6983	0.7540
AlDDF1			1	0.8201
AlDDF2				1

to AtDDF2, suggesting that the reduction in DDF2 affinity took place in the *A. thaliana* lineage only.

Overall, DDF1 and DDF2 PBM data support the idea that orthologous as well as paralogous TFs can have divergent binding site preferences. For DDF2 orthologs, there is a difference in binding affinity, and AlDDF1 exhibits higher sequence selectivity compared with AtDDF1 (Fig. 6A). This suggests that, despite the prediction of negative pleiotropic effects (Prud'homme et al., 2007), differences in binding site preference between orthologs may be common. Orthologs are often assumed to have more similar functions than paralogs that arose from duplication events prior to speciation, because they tend to have higher sequence conservation. However, this assumption is not always valid, and functional equivalence needs to be determined on a case-by-case basis (Gabaldón and Koonin, 2013). It is interesting that orthologous *DDF1* and *DDF2* genes clearly have divergent binding profiles, despite having a high degree of sequence similarity (96% and 98% for DDF1 and DDF2 orthologs, respectively).

Determinant of Reduced AtDDF2 Binding Affinity Relative to AtDDF1

The DNA-binding domains of AtDDF2 and AlDDF2 are 97% identical, but they have very different binding affinities (Fig. 4). To determine the mechanistic basis for the reduced affinity of AtDDF2, we compared its amino acid sequence with those of other Brassicaceae DDF1 and DDF2 proteins (Supplemental Fig. S1). There are two amino acid substitutions that are found in AtDDF2 but not in the other DDF1 and DDF2 orthologs: (1) an Asp at position 37, which is located within the AP2 DNA-binding domain; and (2) an Ala at position 96, which is C terminal to the DSAWR CBF signature motif. To determine if one or both of these amino acid differences are responsible for reduced AtDDF2 binding affinity, we assayed two site-directed mutants, AtDDF2 D37N and A96E, which resemble the ancestral DDF sequence, for their in vitro binding affinity to the top-ranking GCCGACAT motif. AtDDF2 D37N has increased affinity for a probe containing GCCGACAT compared with wild-type AtDDF2; however, an increase in affinity was not observed for AtDDF2 A96E (Fig. 5B). Consistent with these results, the AtDDF1 N37D site-directed mutant has reduced affinity for GCCGACAT (Fig. 5C) compared with wild-type AtDDF1 (Fig. 5C; 1.8-fold less binding; $P < 0.005$), indicating that position 37 is important for binding affinity.

To determine how the Asp-37 substitution affects binding affinity, we took advantage of the crystal structure available for AtERF1 bound to the GCC box (TAGCCGCCA; Allen et al., 1998) to model the binding of AtDDF1 and AtDDF2 to DNA via homology modeling followed by short molecular dynamics simulations. Representative snapshots from the simulations are shown in Figure 7. Based on these models, AtDDF1 residue Asn-37 interacts via a water molecule with the A nucleotide located one base upstream of the GCC box (Fig. 7A, A^2), and the neighboring amino

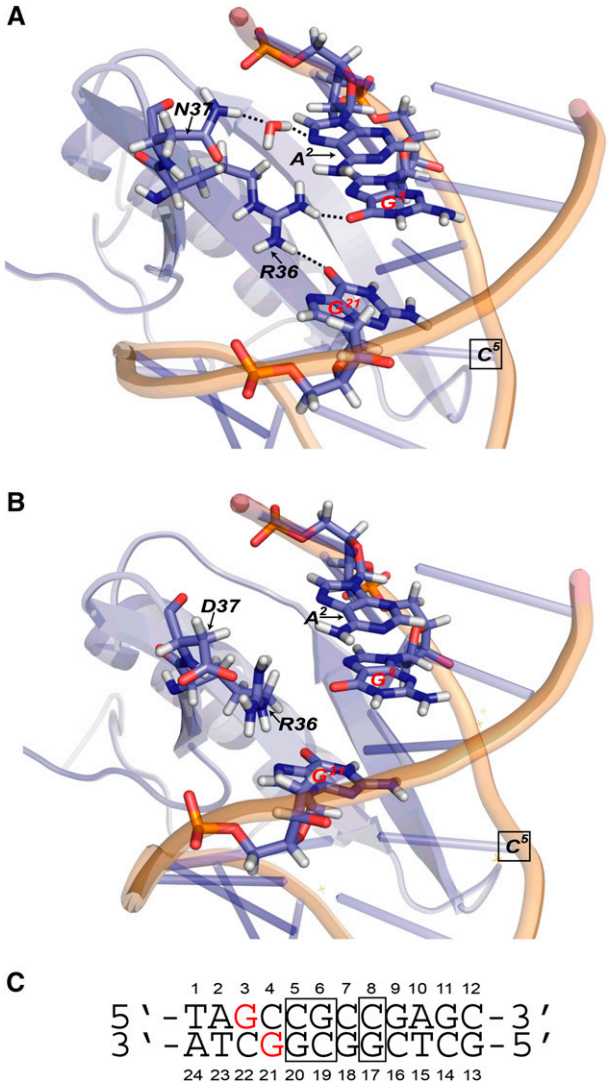

Figure 7. In AtDDF2, an intramolecular interaction between Arg-36 and Asp-37 disrupts protein-DNA interactions. A and B, Models of AtDDF1 (A) and AtDDF2 (B) bound to the GCC box. C, GCC box sequence and numbering of DNA nucleotides. Nucleotides that interact with Arg-36 are highlighted in red, and core nucleotides previously shown to be critical for TF-GCC box binding are outlined in boxes.

Table II. *Genetic diversity of DREB A-1 genes in A. thaliana*

Gene	Length[a]	K_a/K_s	π[b]	π_{syn}[c]	π_{nonsyn}[d]	H[e]	S[f]	H_n[g]
DDF1/AT1G12610	209	0.126	0.0117	0.0443	0.0016	14	30	0.3786
DDF2/AT1G63030	181	0.115	0.0072	0.0206	0.0033	16	19	0.1337
CBF2/AT4G25470	213	0.159	0.0044	0.0128	0.0022	15	24	−1.6620[h]
CBF3/AT4G25480	216	0.169	0.0039	0.0064	0.0034	22	27	−0.8396
CBF1/AT4G25490	216	0.088	0.0050	0.0162	0.0020	16	16	−0.6957
CBF4/AT5G51990	224	0.213	0.0033	0.0164	0.0001	5	7	−0.7265

[a]Number of amino acids. [b]Nucleotide diversity at all sites. [c]Nucleotide diversity at synonymous sites. [d]Nucleotide diversity at nonsynonymous sites. [e]Number of haplotypes. [f]Number of segregating sites. [g]Normalized H (Fay and Wu, 2000; Zeng et al., 2006). [h]$0.05 < P < 0.1$.

acid residue Arg-36 interacts with nucleotides G-3 and G-21 within the GCC box (Fig. 7A). In AtDDF2, both of these interactions are disrupted due an intramolecular interaction between Arg-36, which is positively charged, and the Asp-37 residue, which is negatively charged (Fig. 7B). Neither Asn-37 nor Arg-36 contacts nucleotides that are critical for binding to both the GCC box and DRE (Allen et al., 1998; Sakuma et al., 2002; Fig. 7, nucleotides in boxes). This explains why the Asp-37 AtDDF2 substitution reduces DNA-binding affinity but does not abolish binding. This model also indicates that the reduced affinity of AtDDF2 for DNA is not due to general protein misfolding. Interestingly, what is not clear from this model is how AtDDF2 and AlDDF2 share a similar preference for GTCGAC, given that the interaction of Asp-37 with, in this case, A-21 (the complement of T-4) is also presumably disrupted. Thus, it is not clear which amino acids contribute to differences in sequence specificity.

Selective Constraints on DDF1 and DDF2

Although TFs that bind to sites with low affinity play important roles in transcriptional regulation (Tanay, 2006; Ramos and Barolo, 2013), the fact that *AtDDF2* has significantly lower binding affinity, more reduced and narrower expression, and greater loss of ancestral stress responsiveness compared with *AtDDF1* suggests that *AtDDF2* may be under relaxed selection (i.e. still subjected to purifying selection but not as strongly) and/or in the process of becoming a pseudogene. To determine whether there is evidence of strong purifying selection (selection against mutations) on *DDF1* and *DDF2*, we looked at the ratio of nonsynonymous to synonymous substitution rates (K_a/K_s) between orthologs. The pairwise K_a/K_s between *A. thaliana* and *A. lyrata* DDF1 (0.126) and DDF2 (0.115) sequences are less than 29% of all *A. thaliana* genes and less than 18% of AP2/ERF genes (Table II; Supplemental Fig. S9A). This suggests that, contrary to our expectation, both *AtDDF1* and *AtDDF2* are under a stronger than average selective constraint.

The *A. thaliana* and *A. lyrata* lineages diverged from one another approximately 13 MYA (Beilstein et al., 2010), and it is possible that nonfunctionalization of *DDF2* occurred in the *A. thaliana* lineage due to the relaxation of selection pressure. In this case, lack of functional constraint may be reflected by an elevated nucleotide diversity (π) similar to what has been observed for pseudogenes (Waters et al., 2008; Yang et al., 2011; Wang et al., 2012a). Therefore, we estimated nucleotide diversity at all (π), synonymous (π_s), and nonsynonymous (π_n) sites among 80 sequenced *A. thaliana* accessions (Cao et al., 2011). Surprisingly, we found high values of π for both *DDF1* and *DDF2*, at the 95th and 88th percentiles of all *A. thaliana* genes, respectively, and higher than the other DREB A-1 genes (Table II; Supplemental Fig. S9B). However, the high π observed for *DDF1* is due to π_s, and *DDF1* π_n is 2-fold lower than π_n observed for *DDF2* (Table II). This is consistent with the hypothesis that *DDF2* is under lower functional constraint, but it is important to note that the *CBF1*, *CBF2*, and *CBF3* genes also have elevated π_n compared with *DDF1*. Previously, it was shown that high nucleotide diversity among *CBF1*, *CBF2*, and *CBF3* is due in part to relaxed purifying selection in southern populations (Zhen and Ungerer, 2008). Similarly, *DDF2* may be under relaxed purifying selection only in specific environments.

If *DDF2* is under relaxed functional constraint, the amino acid substitutions among *A. thaliana* accessions may be more likely to impact protein function than those in *DDF1*. However, with the exception of a

Table III. *MK test of neutrality*

Gene	P_n[a]	P_s[b]	P_n/P_s	D_n[c]	D_s[d]	D_n/D_s	P[e]
DDF1/AT1G12610	6	23	0.261	9	22	0.409	0.556
DDF2/AT1G63030	9	8	1.125	7	18	0.389	0.125
CBF2/AT4G25470	15	9	1.667	17	29	0.586	0.048
CBF3/AT4G25480	18	6	3	18	30	0.600	0.005[f]
CBF1/AT4G25490	8	8	1	12	34	0.353	0.120
CBF4/AT5G51990	1	6	0.167	12	18	0.667	0.383

[a]Number of nonsynonymous polymorphisms within *A. thaliana*. [b]Number of synonymous polymorphisms within *A. thaliana*. [c]Number of nonsynonymous fixed differences between *A. thaliana* and *A. lyrata*. [d]Number of synonymous fixed differences between *A. thaliana* and *A. lyrata*. [e]*P* value from the MK test of neutrality (McDonald and Kreitman, 1991). [f]Significant after Bonferroni correction.

frame-shift mutation in *DDF2* in a single accession (Supplemental Fig. S10), none of the amino acid changes in DDF1 and DDF2 (six and nine changes, respectively) are predicted to disrupt protein function (Supplemental Fig. S10). Notably, all *DDF2* alleles code for Asp-37, suggesting that the mutation occurred before the divergence of different *A. thaliana* accessions more than 10,000 years ago (Cao et al., 2011). It is possible that the fixation of Asp-37 occurred due to genetic drift and was not selected against, because reduced DNA-binding affinity did not impact fitness. It is also possible that fixation occurred because it was advantageous, for example, by alleviating paralog interference (Baker et al., 2013). Previous studies have shown evidence for a selective sweep of *CBF2* (Lin et al., 2008) and the *CBF1*, *CBF2*, and *CBF3* cluster in certain *A. thaliana* populations (Barboza et al., 2013) based on Fay and Wu's H statistic (2000). This statistic is calculated based on polymorphism data within species and divergence between species (Fay and Wu, 2000). A significantly negative H indicates an excess of high-frequency-derived alleles (i.e. alleles not found in the outgroup species) and is consistent with a selective sweep. A significantly positive H indicates a deficit of intermediate and high-frequency-derived alleles and is suggestive of balancing selection. Finally, an H near zero is consistent with neutrality. To determine if there is also evidence for recent selection of *DDF1* and *DDF2*, we calculated normalized H (H_n) using the *A. lyrata* orthologs as outgroups and including the DREB A-1 genes for comparison (Table II). Consistent with previous studies, the *CBF* genes have negative H_n values (Lin et al., 2008; Barboza et al., 2013), but here, H_n is only marginally significant for *CBF2* ($P = 0.056$ before Bonferroni correction). In contrast to the DREB A-1 genes, *DDF1* and *DDF2* have positive H_n estimates that are not significantly different from zero ($P = 0.519$ and $P = 0.374$, respectively).

We also performed the McDonald-Kreitman (MK) test of selection (McDonald and Kreitman, 1991) to detect possible negative as well as positive selection (Zhai et al., 2009). The basis for the MK test is that, under the neutral expectation, the ratio of nonsynonymous to synonymous substitutions fixed between species, in this case *A. thaliana* and *A. lyrata*, should equal the ratio of nonsynonymous to synonymous substitutions polymorphic among *A. thaliana* accessions. A significant difference in ratios determined by Fisher's exact test provides evidence for selection. The MK test is significant for *CBF3* ($P = 0.005$) and for *CBF2* ($P = 0.048$), but the latter is not significant after Bonferroni correction. In both cases, there is an excess of replacement polymorphisms consistent with negative selection (Table III). There is no evidence for the selection of either *DDF1* or *DDF2* (Table III), although *DDF2* differs from *DDF1* in having a higher ratio of nonsynonymous to synonymous polymorphisms (1.125 and 0.261, respectively).

Although cross-species comparisons indicate strong selective constraints, based on comparisons between *A. thaliana* accessions, there is no evidence for the recent selection of *DDF1* and *DDF2*. Although this is consistent with the hypothesis that these genes are under relaxed functional constraint, failure to reject the hypothesis of neutrality for a sequence does not mean that the sequence is nonfunctional but that the observed polymorphism does not affect fitness. Furthermore, tests for selection lack power to detect positive selection, especially at single codons and in *A. thaliana*, where inefficient selection against slightly deleterious mutations results in an excess of nonsynonymous polymorphisms (Bustamante et al., 2002; Hughes, 2007). Another issue is that, because these population genetic statistics are a summary of all analyzed accessions, there is less power to detect genes that are under selection in some environments but not in others. Such differential selection among populations is expected among genes that are relevant to environmental adaptation (Hancock et al., 2011; Lee and Mitchell-Olds, 2012), and this has already been shown to be the case for the *CBF1*, *CBF2*, and *CBF3* genes (Alonso-Blanco et al., 2005; Zhen and Ungerer, 2008).

CONCLUSION

Through phylogenetic, expression, and PBM analyses, we show evidence for the functional divergence of *DDF1* and *DDF2* derived from WGD. Divergent protein structures and binding preferences between DDF1 and DDF2 orthologs indicate that functional divergence may have occurred soon after duplication and argue against dosage balance as the mechanism for retention. Despite this, the loss of ancestral stress responses, loss of binding affinity, and elevated π_n indicating recent relaxed selection all point toward the possibility that *AtDDF2* is undergoing functional decay. Yang et al. (2011) proposed that promoter disablement may be a largely unrecognized first step in pseudogenization, and the consequent relaxation of selection may be followed by the disablement of coding regions. Consistent with this hypothesis, among class IV homeodomain-Zip TF duplicate genes derived from the α WGD, one duplicate often has a higher rate of evolution and a more limited expression pattern than its paralog, suggesting possible pseudogenization through the disablement of regulatory regions (Zalewski et al., 2013). *AtDDF2* also shows evidence of promoter disablement, such as reduced expression, fewer predicted cis-elements compared with *DDF1*, and a methylated repetitive element found in close proximity to the TSS. The timing of reduced *DDF2* responsiveness to stress relative to the Asp-37 substitution that affects binding affinity is unclear, and more extensive expression profiling would need to be done to determine whether *AlDDF2*, like *AtDDF2*, is lowly expressed under all conditions. There is no strong evidence for the accelerated evolution of *AtDDF2* compared with *AtDDF1*, and *AtDDF2* is a pseudogene in only one *A. thaliana* accession. Future studies to

determine the effect of *AtDDF2* loss of function on plant fitness are necessary to determine if *AtDDF2* is truly nonfunctional. Nevertheless, the apparent degradation of *AtDDF2* function highlights the fact that, while 25% of the TF duplicates generated during the α WGD still remain in the *A. thaliana* genome (Blanc and Wolfe, 2004), gene loss is an ongoing process (Schnable et al., 2012), and many TF duplicates derived from paleopolyploidy may be in various stages of functional decay. Future studies that evaluate the impact of duplicate gene loss on plant fitness will be important to assess whether more functional TF duplicates contribute more to plant fitness than those with evidence of functional decay.

MATERIALS AND METHODS

Sequences and Phylogenetic Analysis

The *Arabidopsis thaliana* DREB A-1 genes *DDF1* (AT1G12610), *DDF2* (AT1G63030), *CBF1* (AT5G25490), *CBF2* (AT5G25470), *CBF3* (AT5G25480), and *CBF4* (AT5G51990), DREB A-4 genes *HARDY* (AT2G36450), AT5G52020, and AT1G12630, and *ERF1* (AT4G17500) DNA and protein sequences were obtained from The Arabidopsis Information Resource (www.arabidopsis.org), and the *Arabidopsis lyrata DDF1* (GI: 9328759) and *DDF2* (GI: 9324047) sequences were obtained from Phytozome (http://phytozome.jgi.doe.gov/pz/portal.html). The N terminus of the annotated AlDDF1 protein sequence lacks 10 amino acids that are present in the predicted transcript and align to AtDDF1; therefore, we included these amino acids in this study. To identify syntenic orthologs in *Capsella rubella*, *Thellungiella halophila*, and *Brassica rapa*, the protein-coding sequences of 20 genes upstream and 20 genes downstream of *AtDDF1* and *AtDDF2* were obtained from Phytozome (version 9) and used as a query in a BLAST (Altschul et al., 1997) search against Phytozome proteins for each species. BLAST hits were used in MCScanX version 4.4.5 (Wang et al., 2012b) to identify syntenic blocks. *C. rubella DDF1* (Carubv10010303) and *DDF2* (Carubv10021910), *T. halophila DDF1* (Thhalv10009978) and *DDF2* (Thhalv10023865), and *B. rapa DDF1-1* (Bra016763), *DDF1-2* (Bra019777), *DDF1-3* (Bra026963), and *DDF2* (Bra027612) were identified as syntenic orthologs. *Populus trichocarpa* sequences (Potri.004G187000.1, Potri.009G147700.1, Potri.015G136400.1, Potri.012G134100.1, Potri.001G110700.1, and Potri.001G110800.1) were identified in previous analyses as being closely related to the DREB A-1 subfamily (Benedict et al., 2006; Chen et al., 2013). *Carica papaya* sequences (Cpapaya_supercontig_74_84, Cpapaya_supercontig_20_67, Cpapaya_supercontig_5_66, Cpapaya_supercontig_5_71, and Cpapaya_supercontig_7_269) had $E \leq 10^{-24}$ in a BLAST search with the *AtDDF1* protein. Protein sequences were aligned with MUSCLE (Edgar, 2004), and a visual representation was generated with Boxshade (http://www.ch.embnet.org/software/BOX_form.html). A phylogenetic tree was generated using the maximum likelihood method with 1,000 bootstrap replicates in MEGA 5.2.2 and rooted with *AtERF1* (AT4G17500; Tamura et al., 2011).

Expression Analysis of AP2/ERF Duplicate Genes

Developmental series, light, and stress treatment expression data were obtained from AtGenExpress (https://www.arabidopsis.org/portals/expression/microarray/ATGenExpress.jsp; Schmid et al., 2005; Kilian et al., 2007), root cell type salt stress-responsive expression data were obtained from Dinneny et al. (2008), and the ATH1 microarray compendium was downloaded from the NASCArrays Web site (http://affymetrix.arabidopsis.info/link_to_iplant.shtml). Processing and differential expression calculations (fold changes) for the AtGenExpress and root cell type data were performed as described previously (Zou et al., 2009). NASCArray data were used in the processed format (MAS normalized) as it was downloaded from the database. PCCs were calculated between NASCArray ATH1 chip data sets, and only one representative data set was kept for data sets with PCC > 0.98. The SciPy library (http://www.scipy.org/) was used to calculate the pairwise PCCs of AP2/ERF duplicate genes and to calculate the two-sample Kolmogorov-Smirnov test comparing the PCC distributions of AP2/ERF duplicate genes with the random PCC distribution. To visualize the relationships between pairwise PCCs of AP2/ERF duplicates, a level plot was generated with the lattice package implemented in R (http://stat.ethz.ch/R-manual/R-devel/library/lattice/html/levelplot.html; http://www.r-project.org/). To determine the significance of these relationships, expression PCCs for pairs of random genes were calculated. The PCCs of AP2 duplicates and the same number of randomly selected genes were plotted as histograms using the ggplot2 package in R (http://cran.r-project.org/web/packages/ggplot2/index.html). To test the genome dominance hypothesis, the expression intensities of genes in the same α WGD block as *DDF1* (A03-a) were retrieved and compared with the expression of their duplicate pair in the same block as *DDF2* (A03-b). Specifically, the significance of differences in intensity distributions across the AtGenExpress light, stress, and development data sets for each duplicate pair was tested using the Wilcoxon rank-sum test. Fisher's exact test was used to test the hypothesis that duplicate pairs in block A03-a are not more likely to be more highly or more lowly expressed than genes in block A03-b. In this contingency table, the expected numbers of genes with higher expression in block a compared with block b and vice versa were the numbers of α WGD duplicate pairs on all other blocks. For this analysis, the α WGD block a and b designations were those defined by Bowers et al. (2003).

Promoter Analysis

Putative cis-elements in the 1.5-kb predicted promoter sequences and 5′-untranslated regions of the *DDF* genes were identified via searches of the PlantCARE database (Lescot et al., 2002; http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Methylation was determined using the data from Stroud et al. (2013). Repetitive elements were defined based on the annotations by Maumus and Quesneville (2014).

Ancestral Expression Pattern Inference

AtGenExpress abiotic stress expression data (Kilian et al., 2007) were used to determine the responsiveness of DREB A-1 and A-4 genes to cold, drought, salt, and wounding stress in roots and shoots. Time points in which none of the DREB A-1 genes were responsive were dropped from further analysis. A gene was called responsive if it had a greater than 2-fold change in expression compared with controls under a given condition. Protein sequences were aligned with MUSCLE (Edgar, 2004), and a phylogeny of the DREB A-1 genes and three DREB A-4 genes as outgroups was inferred from amino acid sequences using the maximum likelihood method in MEGA 5.2.2 (Tamura et al., 2011). This information, in conjunction with the response states of the extant DREB A-1 and A-4 genes, was used to infer the ancestral stress response states with BayesTrait (Pagel et al., 2004) using a maximum likelihood model of trait evolution. Ancestral state probabilities for nodes labeled in Supplemental Figure S11 are shown in Supplemental Table S6.

Real-Time PCR

A. thaliana Columbia (CS70000) and *A. lyrata* (CS22696) seed stocks were obtained from the Arabidopsis Biological Resource Center (www.arabidopsis.org). Seeds were surface sterilized with bleach, sown on vertical plates containing 0.5× Murashige and Skoog basal salts (MP Biomedicals) and 7 g L^{-1} bacto agar, and stratified for 3 d at 4°C. After 14 d of growth under constant light, seedlings were frozen in liquid nitrogen (time zero) or transferred to filter paper containing either water or 150 mM NaCl for 3 h before freezing in liquid nitrogen. Total RNA was isolated with the RNeasy Plant Mini Kit (Qiagen). Eluted RNA was treated with RQ1 RNase-free DNase (Promega) and then further purified by phenol:chloroform:isoamyl alcohol (25:24:1) extraction followed by sodium acetate and ethanol precipitation or by spin column purification (Qiagen). First-strand cDNA was synthesized from 0.5 μg of total RNA with the iScript cDNA Synthesis Kit (Bio-Rad) scaled to a 10-μL reaction volume. Real-time PCR was done using Power SYBR Green PCR Master Mix (Life Technologies) on an Eppendorf Master Cycler Realplex^2. Reactions were scaled to 15 μL of total volume and included 1.5 μL of 1:5 diluted cDNA as template and 2.5 pmol of each primer (Supplemental Table S7). Samples were run in triplicate, and comparative threshold (C_T) values were obtained using Eppendorf RealPlex software. ΔC_T values were calculated by subtracting the average comparative threshold values of stably expressed reference genes: protein phosphatase 2A subunit (AT1G13320) and the *A. lyrata* ortholog of AT2G28390, a SAND family gene (GI: 9315166; Czechowski

et al., 2005), for *A. thaliana* and *A. lyrata*, respectively. Relative expression of *DDF1* and *DDF2* ($\Delta\Delta C_T$) was calculated for three biological replicates.

GST Fusion Protein Production

AtDDF1, *AtDDF2*, *AlDDF1*, and *AlDDF2* coding sequences were cloned into the pGEX5-1 vector (GE Healthcare), and proteins were expressed in protease-deficient *Escherichia coli* BL21 cells. After growth overnight at 28°C in Luria-Bertani medium containing 100 μg mL^{-1} ampicillin, cultures were diluted 1:100 and grown until an optical density of 0.6 to 1 was reached. Fusion protein expression was induced with 0.1 mM isopropyl β-D thiogalactoside, and cultures were grown for an additional 1 h at 28°C. Pelleted cells were resuspended in 1× phosphate-buffered saline with a protease inhibitor cocktail (Sigma). Resuspended pellets were sonicated on ice with four passes, 10 s each, at 5 to 10 W. Triton X-100 was added to a final concentration of 1% (v/v), and the sample was mixed gently at 4°C on a rotator for 30 min. After pelleting cell debris, the supernatant was filtered with a 0.45-μm Millex-HA syringe filter before column purification with 1-mL GSTrap FF columns according to the manufacturer's instructions. Purification for all samples was done at 4°C with the exception of AlDDF1, where purification at room temperature significantly increased yield. Proteins were eluted with 50 mM Tris-HCl and 10 mM reduced glutathione, pH 8, and concentrated with Microcon YM-30 columns. Glycerol was added to a final concentration of 30% (v/v), and samples were stored at −80°C. Molarities of fusion proteins were determined by comparing with a standard curve composed of a dilution series of purified GST (Genscript) on a western blot probed with rabbit anti-GST primary antibody (Invitrogen) and horseradish peroxidase-conjugated goat anti-rabbit secondary antibody (Bio-Rad). Chemiluminescence was detected and quantified using the Bio-Rad VersaDoc Imager and Quantity One software or the Bio-Rad ChemiDoc MP imaging system and Image Lab software.

Protein-Binding Microarray

Universal protein-binding microarray experiments were carried out as described by Berger and Bulyk (2009), except that anti-GST antibody (Invitrogen; A5800) labeled with AlexaFluor 647 (Molecular Probes; A20173) was used in place of AlexaFluor 488 conjugate. Cy3 and AlexaFluor 647 scans were obtained with an Agilent Technologies Scanner (G2505B) using the Cy3 (532-nm excitation) and Cy5 (633-nm excitation) lasers, respectively. Two replicates for each fusion protein were performed using two independent array designs (AMADID nos. 015681 and 016060), which are constructed with different de Bruijn sequences of order 10 (Berger and Bulyk, 2009). The efficiency of array double stranding was determined by measuring Cy3-labeled dUTP incorporation. Scans to detect AlexaFluor 647 fluorescence were performed at multiple photomultiplier values to ensure that all spots were detected below saturation, and intensities were combined with Masliner software (Dudley et al., 2002) as described by Berger and Bulyk (2009). Identification of 8-mers, their median intensities, and E scores was done using PBM analysis software (http://the_brain.bwh.harvard.edu/software.html), and PWM were identified using the Seed and Wobble program (Supplemental Table S4; Berger et al., 2006) using the highest ranked motif as a seed. PBM data have been deposited in the UniPROBE database (Hume et al., 2015; http://the_brain.bwh.harvard.edu/uniprobe/) with accession numbers UP00555 to UP00558.

To compare intensities between arrays, normalized intensities were calculated separately for each protein and replicate:

$$\text{Normalized Intensity of } i = (i - \min)/(\max - \min)$$

where i is the ith intensity, min is the minimum intensity across probes, and max is the maximum intensity across probes.

Generation of Site-Directed Mutant Proteins

Overlap extension PCR (Ho et al., 1989) was used to generate site-directed mutants. *AtDDF2* in pGEX was used as a template in two reactions containing Pfu Turbo DNA polymerase (Agilent Technologies) and *Bam*HI-AtDDF2 F primer plus D37N R primer or AtDDF2-*Xho*I R primer plus D37N F primer (for primer sequences, see Supplemental Table S7). The products from these two reactions were used as a template in a third reaction containing Pfu Turbo DNA polymerase and *Bam*HI-AtDDF2 F primer and AtDDF2-*Xho*I R primer. PCR products were digested with *Bam*HI and *Xho*I and ligated into a similarly cut pGEX vector. The AtDDF2 D37N pGEX construct was used as a template in PCR with Pfu and *Nde*I-GST F primer and AtDDF2-*Xho*I R primer. The purified PCR product was subcloned into the pGEM T-easy vector (Promega). After sequence verification, the insert was released with *Nde*I and *Xho*I digestion and ligated into a similarly cut in vitro transcription/translation plasmid provided with the PurExpress kit (New England BioLabs). The same protocol was used to generate the GST-AtDDF2 A96E and GST-AtDDF1 N37D fusion protein constructs (for primer sequences, see Supplemental Table S7). Wild-type GST-AtDDF1, GST-AtDDF2, and GST in vitro transcription/translation constructs were also generated.

EMSA

EMSAs were performed with the LightShift Chemiluminescent EMSA kit (Thermo Scientific). Binding reactions consisted of 1× binding buffer, 50 ng μL^{-1} poly(dI•dC), 10% (v/v) glycerol, 50 mM EDTA, 0 to 5 pmol of GST column-purified protein or in vitro-translated protein (PurExpress; New England BioLabs), 40 fmol of biotin-labeled probe, plus or minus 8 pmol of unlabeled probe. Double-stranded labeled and unlabeled probes were made and quantified as described by Berger et al. (2006). Briefly, a 57-nucleotide single-stranded oligonucleotide consisting of the 8-mer sequence flanked at the 5′ end by 14 random bases (N) and at the 3′ end by 15 N, and a constant sequence for primer annealing, was made double stranded in a primer extension reaction with *Bacillus stearothermophilus* polymerase and either a 5′ biotin-labeled or unlabeled 20-bp complementary primer (IDT; for oligonucleotide sequences, see Supplemental Table S7). Chemiluminescence was detected using the Bio-Rad VersaDoc Imager or the Bio-Rad ChemiDoc MP imaging system. Intensities of bound and unbound probe were determined using ImageJ (Schneider et al., 2012).

Structural Modeling

The structures for DDF1 and DDF2 were built by homology modeling based on the GCC box-binding domain structure as a template (Protein Data Bank Identifier 1GCC; Allen et al., 1998), identified by using a PSI-BLAST search (Altschul et al., 1997). Multiple models were constructed using MODELER version 9v7 (Sali and Blundell, 1993), and the best model was selected using MODELER's DOPE assessment method (Shen and Sali, 2006) aided by visual inspection. The resulting models were completed by adding hydrogen atoms. His ionization states (protonation on Nδ or Nε) were predicted using PROPKA3.1 (Bas et al., 2008) and confirmed visually based on the local protein environment. The homology models were relaxed via a 200-step steepest descent energy minimization followed by a 2,000-step conjugate gradient minimization, using a force constant of 2 kcal mol^{-1} Å^{-2} on C_α and C_β atoms of the protein. The two minimized models for DDF1 and DDF2 were then merged with the DNA structure from the 1GCC crystal structure. We further performed additional energy minimization for the two protein-DNA complexes with 500-step steepest descent steps followed by 5,000-step conjugate gradient steps, again using a force constant of 2 kcal mol^{-1} Å^{-2} on C_α atoms of protein and the heavy atoms of DNA. The CHARMM22/CMAP force field (MacKerell et al., 1998, 2004) was used for the protein and the DNA, and a distance-dependent dielectric function was employed as a crude solvation model during the minimization. The two resulting systems were subsequently solvated in a cubic box using a TIP3P water model (Jorgensen, 1981) with a cutoff of 8 Å. The total dimension of each system was approximately 65 Å × 65 Å × 65 Å. Fifteen counter ions (Na^+) were added by randomly replacing 15 water molecules to neutralize the system. Each system contained approximately 26,850 atoms. Each solvated system was again energy minimized over 5,000 steps, followed by molecular dynamics with increasing temperature from 0 to 298 K in six 4-ps steps, while maintaining restraints of 2 kcal mol^{-1} Å^{-2} on the C_α atoms of the protein and the heavy atoms of the DNA. All restraints were then released, and the simulations were continued for another 10 ps in an ensemble with a constant number of particles, pressure, and temperature, with a temperature of 298 K and a pressure of 1 bar that were maintained using a Langevin thermostat and Langevin piston barostat. Solvated systems were simulated with the particle-mesh Ewald method (Darden et al., 1993) to calculate long-range electrostatic interactions. The direct electrostatic sum and the Lennard-Jones potential were truncated at 10 Å with a switching function effective at 8.5 Å. The SHAKE algorithm (Ryckaert et al., 1977) was applied to constrain the lengths of all bonds involving hydrogen atoms. The computation and analysis were carried out using CHARMM (Brooks et al., 2009), version c36a5, in conjunction with the MMTSB Tool Set (Feig et al., 2004). PyMOL (Schrödinger) was used for visualization.

K_a/K_s and Nucleotide Diversity

Genome-wide K_a/K_s calculations were estimated based on pairwise comparisons of *A. thaliana* genes and their *A. lyrata* orthologs. Sequences were aligned using PRANK version 11130 for codon-based alignment (Löytynoja and Goldman, 2005), and K_a/K_s were calculated using codeml (PAML version 4.4.5) with default parameters (Yang, 2007). For individual gene analysis, MUSCLE codon alignments between *A. thaliana* and *A. lyrata* orthologous pairs and calculations of K_a/K_s using the Nei-Gojobori method (Jukes-Cantor) were performed in MEGA 5.2.2 (Tamura et al., 2011). In *A. lyrata*, the *CBF1* to *CBF3* genes are misannotated as a single gene; therefore, the tandem *AlCBF1*, *AlCBF2*, and *AlCBF3* coding sequences were manually annotated. *AlCBF3* contains a 4-bp insertion that leads to a frame shift after amino acid 141, and codon alignments were manually adjusted to maintain the coding frame.

For conservation analyses within species, we used polymorphism data in the form of a genome matrix file from 80 different *A. thaliana* accessions (Cao et al., 2011). Genome-wide π for annotated features was calculated using Variscan (Vilella et al., 2005) with the following parameters (RefPos = 1, Outgroup = none, RunMode = 12, UseMuts = 0, CompleteDeletion = 0, FixNum = 1, and NumNuc = 60). Individual π, π_s, and π_n values were estimated using DnaSP 5.10 (Librado and Rozas, 2009). Accessions were excluded from analysis if there were 45 or more ambiguous bases. The DH software from Kai Zeng's laboratory (http://zeng-lab.group.shef.ac.uk/wordpress/?page_id=28; Zeng et al., 2006, 2007) was used to calculate H_n using the *A. lyrata* orthologs as outgroups. Significance levels were determined based on distributions generated by coalescent simulations (10,000 replicates). DnaSP 5.10 was used to calculate the number of fixed and polymorphic synonymous and nonsynonymous sites and to perform the MK test.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. MUSCLE alignment of *A. thaliana* DREB A-1, DREB A-4, and ERF1 (AT4G17500) proteins, DDF1 and DDF2 orthologous proteins in the Brassicaceae, and DREB A-1-related genes in *C. papaya* and *P. trichocarpa*.

Supplemental Figure S2. Expression patterns of AP2/ERF WGD genes are more highly correlated than randomly paired genes.

Supplemental Figure S3. Expression levels of *AtDDF1* and *AtDDF2* in NASCarray experiments.

Supplemental Figure S4. Methylation and predicted repetitive elements in the *AtDDF1* and *AtDDF2* promoters.

Supplemental Figure S5. AtDDF2 has reduced affinity for GCCGACAT compared with AtDDF1.

Supplemental Figure S6. Overlap in 8-mers bound by AtDDF1, AtDDF2, AlDDF1, and AlDDF2.

Supplemental Figure S7. Differences in binding preference between paralogs.

Supplemental Figure S8. Quantitative real-time PCR evaluation of *AtDDF1*, *AtDDF2*, *AlDDF1*, and *AlDDF2* transcription in 14-d-old seedlings treated for 3 h with water or 150 mM NaCl.

Supplemental Figure S9. Evidence for recent relaxation of selection on *AtDDF1* and *AtDDF2*.

Supplemental Figure S10. Location of polymorphic amino acids in AtDDF1 and AtDDF2 based on a survey of 80 *A. thaliana* accessions.

Supplemental Figure S11. Phylogenetic tree used for ancestral expression state prediction.

Supplemental Table S1. PCCs for AP2/ERF duplicate pairs in AtGenExpress expression data sets.

Supplemental Table S2. Numbers and types of PlantCARE elements found in the 1.5-kb promoter regions of *AtDDF1*, *AtDDF2*, *AlDDF1*, and *AlDDF2*.

Supplemental Table S3. E scores and median intensities for the top-scoring 8-mers based on combined analysis of two replicates for each protein.

Supplemental Table S4. PWM for AtDDF1, AtDDF2, AlDDF1, and AlDDF2 generated by the Seed and Wobble program (Berger et al., 2006) using the top-ranked motif, GCCGACAT, as a seed.

Supplemental Table S5. PCCs reflecting correlation between 8-mer relative affinities (contiguous and gapped, $E > 0.45$ for at least one protein).

Supplemental Table S6. Ancestral response state probabilities inferred with BayesTrait.

Supplemental Table S7. Primer sequences.

ACKNOWLEDGMENTS

We thank Martha Bulyk for helpful advice regarding the protein-binding microarray experimental setup and analysis, Jeff Landgraff for help in performing the PBM experiments and analysis, Ning Jiang for helpful discussions about transposable elements, Cheng Zou and Kelian Sun for experimental results not included here, and Yani Chen and Shan Yin for preparing fusion protein expression constructs.

Received May 20, 2015; accepted June 3, 2015; published June 23, 2015.

LITERATURE CITED

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Curr Opin Plant Biol **8:** 135–141

Airoldi CA, Davies B (2012) Gene duplication and the evolution of plant MADS-box transcription factors. J Genet Genomics **39:** 157–165

Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M (1998) A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. EMBO J **17:** 5484–5496

Alonso-Blanco C, Gomez-Mena C, Llorente F, Koornneef M, Salinas J, Martínez-Zapater JM (2005) Genetic and molecular analyses of natural variation indicate *CBF2* as a candidate gene for underlying a freezing tolerance quantitative trait locus in Arabidopsis. Plant Physiol **139:** 1304–1312

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Baker CR, Hanson-Smith V, Johnson AD (2013) Following gene duplication, paralog interference constrains transcriptional circuit evolution. Science **342:** 104–108

Baker CR, Tuch BB, Johnson AD (2011) Extensive DNA-binding specificity divergence of a conserved transcription regulator. Proc Natl Acad Sci USA **108:** 7493–7498

Baker SS, Wilhelm KS, Thomashow MF (1994) The 5′-region of Arabidopsis thaliana cor15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. Plant Mol Biol **24:** 701–713

Barboza L, Effgen S, Alonso-Blanco C, Kooke R, Keurentjes JJ, Koornneef M, Alcázar R (2013) Arabidopsis semidwarfs evolved from independent mutations in GA20ox1, ortholog to green revolution dwarf alleles in rice and barley. Proc Natl Acad Sci USA **110:** 15818–15823

Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. Proteins **73:** 765–783

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc Natl Acad Sci USA **107:** 18724–18728

Benedict C, Skinner JS, Meng R, Chang Y, Bhalerao R, Huner NP, Finn CE, Chen TH, Hurry V (2006) The CBF1-dependent low temperature signalling pathway, regulon and increase in freeze tolerance are conserved in Populus spp. Plant Cell Environ **29:** 1259–1272

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, et al (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell **133:** 1266–1276

Berger MF, Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat Protoc **4:** 393–411

Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW III, Bulyk ML (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat Biotechnol **24:** 1429–1435

Birchler JA (2012) Insights from paleogenomic and population studies into the consequences of dosage sensitive gene expression in plants. Curr Opin Plant Biol **15:** 544–548

Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. Trends Genet **21:** 219–226

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell **19:** 395–402

Blackman BK, Strasburg JL, Raduski AR, Michaels SD, Rieseberg LH (2010) The role of recently derived FT paralogs in sunflower domestication. Curr Biol **20:** 629–635

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16:** 1679–1691

Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, Wang LY, Gerstein M, Snyder M (2007) Divergence of transcription factor binding sites across related yeast species. Science **317:** 815–819

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al (2009) CHARMM: the biomolecular simulation program. J Comput Chem **30:** 1545–1614

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL (2002) The cost of inbreeding in Arabidopsis. Nature **416:** 531–534

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, et al (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet **43:** 956–963

Chen Y, Yang J, Wang Z, Zhang H, Mao X, Li C (2013) Gene structures, classification, and expression models of the DREB transcription factor subfamily in Populus trichocarpa. ScientificWorldJournal **2013:** 954640

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet **9:** 938–950

Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. Nucleic Acids Res **32:** D575–D577

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol **139:** 5–17

Darden T, York D, Pedersen L (1993) Particle Mesh Ewald: an N.Log(N) method for Ewald sums in large systems. J Chem Phys **98:** 10089–10092

Dekkers BJ, Pearce S, van Bolderen-Veldkamp RP, Marshall A, Widera P, Gilbert J, Drost HG, Bassel GW, Müller K, King JR, et al (2013) Transcriptional dynamics of two seed compartments with opposing roles in Arabidopsis seed germination. Plant Physiol **163:** 205–215

Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature **454:** 762–765

De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci USA **110:** 2898–2903

De Smet R, Van de Peer Y (2012) Redundancy and rewiring of genetic networks following genome-wide duplication events. Curr Opin Plant Biol **15:** 168–176

Dietz KJ, Vogel MO, Viehhauser A (2010) AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. Protoplasma **245:** 3–14

Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN (2008) Cell identity mediates the response of Arabidopsis roots to abiotic stress. Science **320:** 942–945

Dowell RD (2010) Transcription factor binding variation in the evolution of gene regulation. Trends Genet **26:** 468–475

Dudley AM, Aach J, Steffen MA, Church GM (2002) Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. Proc Natl Acad Sci USA **99:** 7554–7559

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:** 1792–1797

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413

Feig M, Karanicolas J, Brooks CL III (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model **22:** 377–395

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc Natl Acad Sci USA **111:** 2367–2372

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol **60:** 433–453

Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res **16:** 805–814

Gabaldón T, Koonin EV (2013) Functional and evolutionary implications of gene orthology. Nat Rev Genet **14:** 360–366

Ganko EW, Meyers BC, Vision TJ (2007) Divergence in expression between duplicated genes in Arabidopsis. Mol Biol Evol **24:** 2298–2309

Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science **324:** 1447–1451

Godoy M, Franco-Zorrilla JM, Pérez-Pérez J, Oliveros JC, Lorenzo O, Solano R (2011) Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. Plant J **66:** 700–711

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. Plant Physiol **130:** 639–648

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol **148:** 993–1003

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J (2011) Adaptation to climate across the Arabidopsis thaliana genome. Science **334:** 83–86

Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. Nature **449:** 677–681

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77:** 51–59

Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. Evolution **61:** 995–1016

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:** 1419–1428

Hong JP, Takeshi Y, Kondou Y, Schachtman DP, Matsui M, Shin R (2013) Identification and characterization of transcription factors regulating Arabidopsis HAK5. Plant Cell Physiol **54:** 1478–1490

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D (2009) Genome-wide demethylation of Arabidopsis endosperm. Science **324:** 1451–1454

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity (Edinb) **99:** 364–373

Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML (2015) UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res **43:** D117–D122

Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF (2001) Components of the Arabidopsis C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. Plant Physiol **127:** 910–917

Jorgensen WL (1981) Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers: application to liquid water. J Am Chem Soc **103:** 335–340

Kang HG, Kim J, Kim B, Jeong H, Choi SH, Kim EK, Lee HY, Lim PO (2011) Overexpression of FTL1/DDF1, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in Arabidopsis thaliana. Plant Sci **180:** 634–641

Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J **50:** 347–363

Lang D, Weiche B, Timmerhaus G, Richardt S, Riaño-Pachón DM, Corrêa LG, Reski R, Mueller-Roeber B, Rensing SA (2010) Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biol Evol **2:** 488–503

Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, et al (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci USA **107:** 8063–8070

Lee CR, Mitchell-Olds T (2012) Environmental adaptation contributes to gene polymorphism across the Arabidopsis thaliana genome. Mol Biol Evol **29:** 3721–3728

Lee SC, Lim MH, Yu JG, Park BS, Yang TJ (2012) Genome-wide characterization of the CBF/DREB1 gene family in Brassica rapa. Plant Physiol Biochem **61:** 142–152

Lehti-Shiu MD, Zou C, Hanada K, Shiu SH (2009) Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. Plant Physiol **150:** 12–26

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res **30:** 325–327

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:** 1451–1452

Lin YH, Hwang SY, Hsu PY, Chiang YC, Huang CL, Wang CN, Lin TP (2008) Molecular population genetics and gene expression analysis of duplicated CBF genes of Arabidopsis thaliana. BMC Plant Biol **8:** 111

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133:** 523–536

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell **10:** 1391–1406

Liu SL, Baute GJ, Adams KL (2011) Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in Arabidopsis thaliana. Genome Biol Evol **3:** 1419–1436

Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci USA **102:** 10557–10562

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459–473

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B **102:** 3586–3616

MacKerell AD Jr, Feig M, Brooks CL III (2004) Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc **126:** 698–699

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA **102:** 5454–5459

Magome H, Yamaguchi S, Hanada A, Kamiya Y, Oda K (2004) dwarf and delayed-flowering 1, a novel Arabidopsis mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. Plant J **37:** 720–729

Magome H, Yamaguchi S, Hanada A, Kamiya Y, Oda K (2008) The DDF1 transcriptional activator upregulates expression of a gibberellin-deactivating gene, GA2ox7, under high-salinity stress in Arabidopsis. Plant J **56:** 613–626

Maumus F, Quesneville H (2014) Deep investigation of Arabidopsis thaliana junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS ONE **9:** e94101

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature **351:** 652–654

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. Biochim Biophys Acta **1819:** 86–96

Moghe GD, Hufnagel DE, Tang H, Xiao Y, Dworkin I, Town CD, Conner JK, Shiu SH (2014) Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. Plant Cell **26:** 1925–1937

Moghe GD, Shiu SH (2014) The causes and molecular consequences of polyploidy in flowering plants. Ann N Y Acad Sci **1320:** 16–34

Moore RC, Purugganan MD (2003) The early stages of duplicate gene evolution. Proc Natl Acad Sci USA **100:** 15682–15687

Moyroud E, Minguet EG, Ott F, Yant L, Posé D, Monniaux M, Blanchet S, Bastien O, Thévenon E, Weigel D, et al (2011) Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* LEAFY transcription factor. Plant Cell **23:** 1293–1306

Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol **140:** 411–432

Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet **9:** 951–963

Nozawa M, Kawahara Y, Nei M (2007) Genomic drift and copy number variation of sensory receptor genes in humans. Proc Natl Acad Sci USA **104:** 20421–20426

Oakley TH, Ostman B, Wilson AC (2006) Repression and loss of gene expression outpaces activation and gain in recently duplicated fly genes. Proc Natl Acad Sci USA **103:** 11637–11641

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, Berlin

Okamuro JK, Caster B, Villarroel R, Van Montagu M, Jofuku KD (1997) The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. Proc Natl Acad Sci USA **94:** 7076–7081

Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. Syst Biol **53:** 673–684

Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB (2013) Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. PLoS Genet **9:** e1003748

Pires IS, Negrão S, Pentony MM, Abreu IA, Oliveira MM, Purugganan MD (2013) Different evolutionary histories of two cation/proton exchanger gene families in plants. BMC Plant Biol **13:** 97

Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. Proc Natl Acad Sci USA (Suppl 1) **104:** 8605–8612

Ramos AI, Barolo S (2013) Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. Philos Trans R Soc Lond B Biol Sci **368:** 20130018

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

Rosas U, Mei Y, Xie Q, Banta JA, Zhou RW, Seufferheld G, Gerard S, Chou L, Bhambhra N, Parks JD, et al (2014) Variation in Arabidopsis flowering time associated with cis-regulatory variation in CONSTANS. Nat Commun **5:** 3651

Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of Cartesian equations of motion of a system with constraints: molecular dynamics of N-alkanes. J Comput Phys **23:** 327–341

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun **290:** 998–1009

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234:** 779–815

Sayou C, Monniaux M, Nanao MH, Moyroud E, Brockington SF, Thévenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y, et al (2014) A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. Science **343:** 645–648

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet **37:** 501–506

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al (2010)

Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science **328:** 1036–1040

Schnable JC, Wang X, Pires JC, Freeling M (2012) Escape from preferential retention following repeated whole genome duplications in plants. Front Plant Sci **3:** 94

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods **9:** 671–675

Seoighe C, Gehring C (2004) Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends Genet **20:** 461–464

Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci **15:** 2507–2524

Shiu SH, Shih MC, Li WH (2005) Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiol **139:** 18–26

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell **152:** 352–364

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:** 2731–2739

Tanay A (2006) Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res **16:** 962–972

Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res **16:** 934–946

Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J (2005) VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics **21:** 2791–2793

Wang J, Marowsky NC, Fan C (2013) Divergent evolutionary and expression patterns between lineage specific new duplicate genes and their parental paralogs in Arabidopsis thaliana. PLoS ONE **8:** e72362

Wang L, Si W, Yao Y, Tian D, Araki H, Yang S (2012a) Genome-wide survey of pseudogenes in 80 fully re-sequenced Arabidopsis thaliana accessions. PLoS ONE **7:** e51769

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, et al (2011) The genome of the mesopolyploid crop species Brassica rapa. Nat Genet **43:** 1035–1039

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al (2012b) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res **40:** e49

Wang Z, Triezenberg SJ, Thomashow MF, Stockinger EJ (2005) Multiple hydrophobic motifs in Arabidopsis CBF1 COOH-terminus provide functional redundancy in trans-activation. Plant Mol Biol **58:** 543–559

Waters ER, Nguyen SL, Eskandar R, Behan J, Sanders-Reed Z (2008) The recent evolution of a pseudogene: diversity and divergence of a mitochondria-localized small heat shock protein in Arabidopsis thaliana. Genome **51:** 177–186

Workman CT, Yin Y, Corcoran DL, Ideker T, Stormo GD, Benos PV (2005) enoLOGOS: a versatile web tool for energy normalized sequence logos. Nucleic Acids Res **33:** W389–W392

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science **319:** 1527–1530

Yamaguchi-Shinozaki K, Shinozaki K (1994) A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell **6:** 251–264

Yang L, Takuno S, Waters ER, Gaut BS (2011) Lowly expressed genes in Arabidopsis thaliana bear the signature of possible pseudogenization by promoter degradation. Mol Biol Evol **28:** 1193–1203

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24:** 1586–1591

Yang Z, Wang Y, Gao Y, Zhou Y, Zhang E, Hu Y, Yuan Y, Liang G, Xu C (2014) Adaptive evolution and divergent expression of heat stress transcription factors in grasses. BMC Evol Biol **14:** 147

Zalewski CS, Floyd SK, Furumizu C, Sakakibara K, Stevenson DW, Bowman JL (2013) Evolution of the class IV HD-zip gene family in streptophytes. Mol Biol Evol **30:** 2347–2365

Zeng K, Fu YX, Shi S, Wu CI (2006) Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics **174:** 1431–1439

Zeng K, Shi S, Wu CI (2007) Compound tests for the detection of hitchhiking under positive selection. Mol Biol Evol **24:** 1898–1908

Zhai W, Nielsen R, Slatkin M (2009) An investigation of the statistical power of neutrality tests based on comparative and population genetic data. Mol Biol Evol **26:** 273–283

Zhang JZ (2003) Evolution by gene duplication: an update. Trends Ecol Evol **18:** 292–298

Zhang L, Li Z, Li J, Wang A (2013) Ectopic overexpression of SsCBF1, a CRT/DRE-binding factor from the nightshade plant Solanum lycopersicoides, confers freezing and salt tolerance in transgenic Arabidopsis. PLoS ONE **8:** e61810

Zhen Y, Ungerer MC (2008) Relaxed selection on the CBF/DREB1 regulatory genes and reduced freezing tolerance in the southern range of Arabidopsis thaliana. Mol Biol Evol **25:** 2547–2555

Zou C, Lehti-Shiu MD, Thomashow M, Shiu SH (2009) Evolution of stress-regulated gene expression in duplicate genes of Arabidopsis thaliana. PLoS Genet **5:** e1000581