

Expression Pattern Shifts Following Duplication Indicative of Subfunctionalization and Neofunctionalization in Regulatory Genes of *Arabidopsis*

Jill M. Duarte,^{*1} Liying Cui,* P. Kerr Wall,* Qing Zhang,† Xiaohong Zhang,* Jim Leebens-Mack,* Hong Ma,* Naomi Altman,†‡ and Claude W. dePamphilis*

*Department of Biology, Institute of Molecular Evolutionary Genetics, and Huck Institutes of the Life Sciences, The Pennsylvania State University; †Bioinformatics Consulting Center, The Pennsylvania State University; and ‡Department of Statistics, The Pennsylvania State University

Gene duplication plays an important role in the evolution of diversity and novel function and is especially prevalent in the nuclear genomes of flowering plants. Duplicate genes may be maintained through subfunctionalization and neofunctionalization at the level of expression or coding sequence. In order to test the hypothesis that duplicated regulatory genes will be differentially expressed in a specific manner indicative of regulatory subfunctionalization and/or neofunctionalization, we examined expression pattern shifts in duplicated regulatory genes in *Arabidopsis*. A two-way analysis of variance was performed on expression data for 280 phylogenetically identified paralogous pairs. Expression data were extracted from global expression profiles for wild-type root, stem, leaf, developing inflorescence, nearly mature flower buds, and seedpod. Gene, organ, and gene by organ interaction ($G \times O$) effects were examined. Results indicate that 85% of the paralogous pairs exhibited a significant $G \times O$ effect indicative of regulatory subfunctionalization and/or neofunctionalization. A significant $G \times O$ effect was associated with complementary expression patterns in 45% of pairwise comparisons. No association was detected between a $G \times O$ effect and a relaxed evolutionary constraint as detected by the ratio of nonsynonymous to synonymous substitutions. Ancestral gene expression patterns inferred across a Type II MADS-box gene phylogeny suggest several cases of regulatory neofunctionalization and organ-specific nonfunctionalization. Complete linkage clustering of gene expression levels across organs suggests that regulatory modules for each organ are independent or ancestral genes had limited expression. We propose a new classification, regulatory hypofunctionalization, for an overall decrease in expression level in one member of a paralogous pair while still having a significant $G \times O$ effect. We conclude that expression divergence specifically indicative of subfunctionalization and/or neofunctionalization contributes to the maintenance of most if not all duplicated regulatory genes in *Arabidopsis* and hypothesize that this results in increasing expression diversity or specificity of regulatory genes after each round of duplication.

Introduction

Whole-genome duplication is especially common in flowering plant lineages relative to animal lineages, with between 50% and perhaps 70% or more of all angiosperms having at least one detectable genome duplication in their history (Wendel 2000; Blanc and Wolfe 2004b). Recent work indicates that whole-genome duplications are responsible for more than 90% of the expansion of regulatory genes in the angiosperm lineage over the last 350 Myr (Maere et al. 2005). Under the classical model of gene duplication (Ohno 1970), one duplicate maintains the original function, while the other evolves a new function (rare), is lost, or is silenced (common). However, newer models suggest additional outcomes for the evolutionary fate of duplicated genes (Force et al. 1999; Hughes 2002; Kondrashov 2002; Wagner 2002a). Under these revised models, duplicated genes (paralogs) may experience the following: (1) nonfunctionalization through silencing or null mutation, (2) neofunctionalization through gain of novel function, and (3) subfunctionalization through the partitioning of functional modules such that the complement of both copies represents the functional capability of the ancestral gene (Lynch and Conery 2000). The modification of regulatory modules through mutation or epigenetic effects can result in

specific expression pattern shifts between paralogs, resulting in regulatory subfunctionalization, neofunctionalization, or nonfunctionalization (fig. 1). Subfunctionalization and neofunctionalization contribute to the likelihood of maintenance of a paralogous set of genes, and nonfunctionalization contributes to the likelihood of loss of one member of a paralogous set of genes. A prediction of the duplication-degeneration-complementation (DDC) model is that subfunctionalization will most often result in a symmetric division of regulatory modules, whereas neofunctionalization will typically reflect the gain of a single regulatory module (Force et al. 1999). However, it is unclear how closely the biology of duplicate genes follows the DDC model. Evidence from studies in yeast supports asymmetric divergence between duplicate genes, both in terms of the protein-coding sequence and expression (Wagner 2002b; Gu 2004). Recent studies have also suggested that expression divergence tends toward rapid subfunctionalization followed by neofunctionalization (Kramer, Jaramillo, and Di Stilio 2004; He and Zhang 2005; Zahn et al. 2005a, 2005b).

Arabidopsis is hypothesized to have undergone two, or possibly three, whole-genome duplications in its past (Vision, Brown, and Tanksley 2000; Simillion et al. 2002; Vandepoele, Simillion, and Van de Peer 2002; Blanc, Hokamp, and Wolfe 2003). In addition, studies based on a limited set of genes in polyploids show evidence for rapid expression pattern shifts after duplication (Adams et al. 2003; Osborn et al. 2003). Previous studies on individual sets of regulatory paralogs in flowering plants provide evidence for regulatory subfunctionalization in important

¹ Jill M. Duarte was previously known as Jill M. Ricker.

Key words: expression, gene duplication, ANOVA, microarray, regulatory genes.

E-mail: cwd3@psu.edu.

Mol. Biol. Evol. 23(2):469–478. 2006

doi:10.1093/molbev/msj051

Advance Access publication November 9, 2005

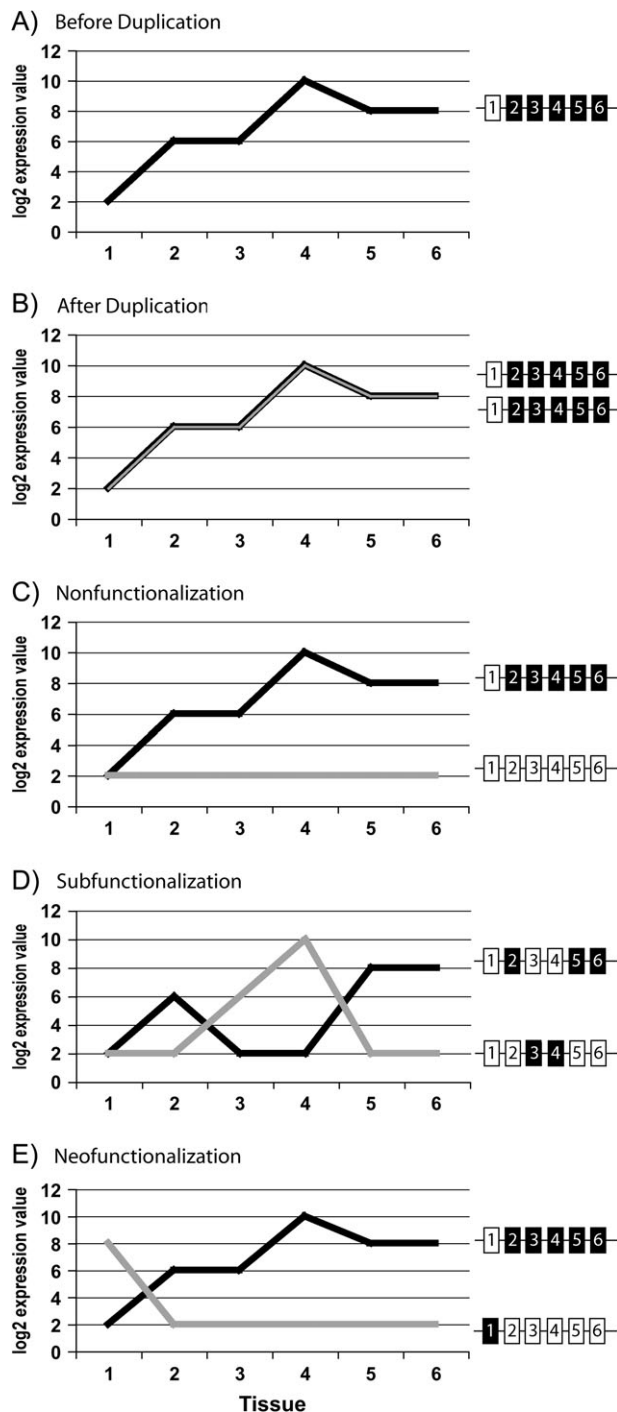


FIG. 1.—Fate of duplicated genes and the effect of each fate on expression patterns. Following the DDC model (Force et al. 1999), before duplication, a given hypothetical gene has a set expression pattern, depicted here as the quantitative expression level in six organs (A). Immediately after duplication, both copies of the gene have the same expression pattern (B). As time passes, the fate of the two copies can be described as (C) nonfunctionalization, where one copy loses expression in all organs; (D) subfunctionalization, where the expression patterns of both copies are complementary and combined are equal to the expression pattern before duplication; and (E) neofunctionalization, where one copy has a novel increase in expression in one or more organs. Values below $\log_2 50 = 5.6$ are assumed to be undetectable expression (Zhang et al. 2005).

genes for reproductive developmental pathways (Bomblies et al. 2003; Hileman and Baum 2003; Matsunaga et al. 2003; Kramer, Jaramillo, and Di Stilio 2004; Zahn et al. 2005a, 2005b). Given the overrepresentation of regulatory genes in the set of maintained duplicate genes in *Arabidopsis* (Blanc and Wolfe 2004a) and their important roles in developmental pathways, we have chosen to focus on shifts in expression patterns of regulatory genes after duplication, as inferred through global expression profiling. Selective constraints on regulatory proteins after duplication, such as dosage imbalance and protein-protein interactions, suggest that expression divergence may frequently play a role in the maintenance of both duplicates (Doebley and Lukens 1998). Therefore, in the case of duplicated regulatory genes, where there are constraints on protein functional divergence and where either the expression or protein function must diverge in order for both copies to survive, we hypothesize that regulatory subfunctionalization and/or neofunctionalization as described by the DDC model (Force et al. 1999) are common in maintained duplicated regulatory genes. In order to test this hypothesis, we have developed a two-way analysis of variance (ANOVA) for the analysis of microarray data.

Microarrays provide a rich resource for investigating the evolution of gene expression. Previous studies of expression divergence between paralogs using microarray data applied a correlation-based approach (Wagner 2000; Gu et al. 2002; Makova and Li 2003; Blanc and Wolfe 2004a). These studies showed that expression divergence does occur between paralogs and that this is correlated with d_s (a measure of silent substitutions) but not with d (a measure of protein divergence). A recent analysis of paralogs in *Arabidopsis* using correlation indicated that 57% of recent duplicates and 73% of older duplicates had divergent expression patterns (Blanc and Wolfe 2004a). Whereas correlation is a good approximation for a relative level of expression divergence that has an easily calculated distance measure, there are limitations in using correlation to analyze the role of specific effects and interaction between effects in expression divergence. In particular, correlation studies only address the frequency of overall divergence and therefore only provide general information about the evolution of expression after duplication. Correlation can result in false negatives if expression pattern shifts are limited to a small number of the data points (e.g., a spike in expression in one of the paralogs due to neofunctionalization in a single organ would be obscured) and false positives if hybridization strength is not uniform over all probes. Furthermore, differences in overall expression levels among paralogs as detected by correlation can be the result of technical differences in probe design and hybridization (Hekstra 2003).

We use ANOVA to more closely examine the relationship of expression patterns between genes in a paralogous pair. ANOVA is widely used in the analysis of microarray experiments (Jin et al. 2001) because of its power, flexibility, and robustness. These qualities also make ANOVA well suited to isolate the factors that contribute to gene expression divergence in duplicate genes. With ANOVA, one can examine the effect of differential and spatial expression and the interaction between these two effects on the overall

divergence of expression patterns between paralogs. With a focus on the consequence of gene duplication, this analysis can distinguish between expression shifts contributing to maintenance of both paralogs (regulatory subfunctionalization or neofunctionalization) and shifts that would lead to the silencing of one paralog (regulatory nonfunctionalization). Whereas correlation analyses only provide general trends in expression divergence since duplication, the ANOVA approach described here tests more directly the predictions of the DDC model (Force et al. 1999).

In order to test the hypothesis that retained duplicated regulatory genes will have evidence of expression pattern shifts that would contribute to maintenance, we applied a split-plot two-way ANOVA to expression data for paralogous pairs of *Arabidopsis* regulatory genes using gene expression profiles for six organs (root, leaf, stem, young inflorescences, later stage flower buds, and silique). Our findings indicate that 85% of the regulatory genes analyzed in this study have undergone significant expression shifts, likely contributing to regulatory subfunctionalization and/or neofunctionalization. In addition, we provide evidence that some paralogous pairs exhibit expression profiles that are not clearly identifiable as regulatory nonfunctionalization, neofunctionalization, or subfunctionalization. These findings have important implications for the evolution of regulatory networks in plants, suggesting that the expression of homologous developmental regulators is likely to vary across plant lineages with distinct histories of ancient polyploidy. We propose that the majority of expression shifts we have detected here contribute to maintenance of regulatory paralogs after duplication via regulatory subfunctionalization and/or neofunctionalization.

Methods

Microarray Experiments

Microarray experiments were performed on the Affymetrix ATH1 array, which is based on the *Arabidopsis thaliana* var. Columbia whole-genome sequence. Two biological replicates were used for six structures (root, leaf, stem, young inflorescences, Stage-12 flower, and siliques), which we refer to as organs, from *A. thaliana* var. Landsberg erecta (Zhang et al. 2005). The microarray data were normalized using the robust multiarray average method (Irizarry et al. 2003) in Bioconductor (Gentleman et al. 2004). The normalized data are expressed in logarithmic units (base 2). Based on Zhang et al. (2005), we considered genes with expression level below 5.6 to be below reliable detection.

Phylogenetic Identification of *Arabidopsis* Paralogous Pairs

In order to identify duplicate pairs of regulatory genes, we first identified putative regulatory gene families in the PlantTribes database, an objectively defined database of putative protein families developed through Markov clustering (Enright, Van Dongen, and Ouzounis 2002; Enright, Kunin, and Ouzounis 2003) of the rice and *Arabidopsis* proteomes (<http://floralgenome.org/cgi-bin/tribedb/tribe.cgi>).

The PlantTribes gene families were inferred using predicted protein sequences from the *A. thaliana* var. Columbia and *Oryza sativa japonica* genomes downloaded from the Institute for Genome Research (<http://www.tigr.org/tdb/euk/>). An all-against-all BlastP (Altschul et al. 1990) analysis was performed. TribeMCL (Enright, Van Dongen, and Ouzounis 2002; Enright, Kunin, and Ouzounis 2003) was used to cluster proteins into putative gene families (<http://floralgenome.org/cgi-bin/tribedb/tribe.cgi>). A multiple amino acid sequence alignment for each regulatory gene family was produced with POA (Lee, Grasso, and Sharlow 2002) followed by RASCAL (Thompson, Thierry, and Poch 2003) to improve poor regions of the alignment. This two-step alignment strategy was used to balance the need for speedy alignments (POA is very fast; Lassmann and Sonnhammer 2002) and accuracy (RASCAL polishing improves alignment accuracy under diverse alignment conditions; Thompson, Thierry, and Poch 2003; K. Beckman, J. Leebens-Mack, and C. W. dePamphilis, unpublished data). Maximum parsimony (MP) phylogenetic analysis was performed for each regulatory tribe amino acid alignment using PAUP* 4.0b (Swofford 2001). Tree searching was performed using 10 random sequence additions and Tree Bisection-Reconnection branch swapping for alignments with fewer than 50 taxa, subtree pruning reconnection branch swapping for alignments with greater than 50 taxa, and fast jackknife for alignments with greater than 100 taxa. MP bootstrapping was performed (1,000 replicates) with heuristic searches using random sequence additions as above. A paralogous pair was identified as a distinct monophyletic clade of two *Arabidopsis* genes with greater than 50% bootstrap support in a given tribe phylogeny. Additional comparisons were made using an 80% cutoff for paralog identification to test the effect of bootstrap support on results. Pairs with missing expression data for one or both genes were eliminated from further analyses.

Statistical Analyses

In order to identify which components contribute to expression pattern divergence within each duplicate pair, a split-plot two-way ANOVA was used to partition the gene (G) effect (the subplot treatment), organ (O) effect (the whole-plot treatment), and gene by organ interaction ($G \times O$) effect. In a preliminary analysis, a term was also included for biological replication, but this was not statistically significant for any gene pair, possibly because biological variability is small compared to technical variability in this inbred species. Hence, we used a standard split-plot ANOVA, using microarrays as the whole plots, which allows for correlation among paralogs measured on the same array, but independence between arrays. Analysis was done using SAS PROC MIXED (Littell et al. 1996). The resulting analysis allows us to assess whether there are differences in the mean expression for the two genes averaged over all organs (G effect), differences in the mean expression for each of the six organs averaged over the two genes (O effect), or whether the within-organ mean expression varies by gene ($G \times O$ effect). A duplicate pair with expression levels A and B was said to have a complementary expression pattern in organ X (organs X and Y)

if the $G \times O$ effect was statistically significant, and $A > B$ in some organs (both X and Y) while $A < B$ in the other organs, or vice versa (e.g., fig. 1E).

For genes with expression lower than the threshold ($2^{5.6}$) in all organs, a one-way ANOVA to test for significant O effects was used to distinguish putative silencing events from lowly expressed genes. Genes lower than the threshold in all organs and without a significant O effect were identified as putative “silent” loci. Putative silent loci were further examined for evidence of expression in other conditions using all Affymetrix whole-genome microarray data stored at the Arabidopsis information resource (TAIR) (<http://www.Arabidopsis.org>).

Complete linkage cluster analysis using correlation distance (Anderbert 1973; Spath 1980) was used to find relatively homogeneous clusters of organs using transcript abundance. A total of 1,186 probe sets representing regulatory genes were identified from the 22,810 probe sets on the array. Using jackknife samples of 1,000 genes, 200 cluster trees were built using complete linkage clustering with correlation distance. PHYLIP (Felsenstein 1989) was used to construct a consensus tree for organ similarity using the extended majority rule. Complementary expression patterns in paralogous pairs as previously described were mapped onto the resulting tree of organs to assess independence of regulatory modules and extent of expression pattern shifts.

Tests for Adaptive Protein Evolution

Codon-based alignments for each gene pair were obtained by translating the DNA sequences, aligning in MUSCLE (Edgar 2004), and forcing the DNA sequences back onto the amino acid alignments. Each alignment was also inspected visually. In order to analyze the changes in selective constraint for each paralogous pair, maximum-likelihood, codon-based analyses of nonsynonymous to synonymous nucleotide substitution ratios ($\omega = d_N/d_S$) were performed for each pair using the codeml program in PAML version 3.13 (Yang 1997) using program default values. To investigate whether paralogous pairs with expression pattern shifts are more or less likely to exhibit altered constraint at the protein level, Mann-Whitney U tests were used to compare the d_N , d_S , and d_N/d_S distributions between (1) paralogous pairs with a significant $G \times O$ effect and paralogous pairs without a significant $G \times O$ effect; (2) paralogous pairs containing at least one putative nonfunctionalization event and all other paralogous pairs; and (3) paralogous pairs with a significant $G \times O$ effect and paralogous pairs without a significant $G \times O$ effect, with cases of putative nonfunctionalization events removed from the data set.

Inferring Ancestral Expression Patterns in Type II MADS-Box Genes

The mean expression value for each organ for each Type II MADS-box gene in *Arabidopsis* was obtained and sorted into seven discrete categories based on twofold expression differences starting with the threshold for detectable expression of $2^{5.6}$. For each organ, the character states for each gene (0–6) were mapped onto a modified phylogeny of Type II MADS-box *Arabidopsis* genes (Martinez-Castilla and Alvarez-Buylla 2003) as unordered states with ancestral

expression inferred under Fitch optimization in MacClade 4.06 (D. R. Maddison and W. P. Maddison 2001). Equivocal ancestral expression levels were resolved using ACCTRAN. The results from each organ tree were compiled in order to infer expression patterns over all six organs for each ancestral gene prior to duplication. The expression patterns from all extant Type II MADS-box paralogous pairs were then compared to the expression pattern for the ancestral gene in order to detect regulatory subfunctionalization, neofunctionalization, and nonfunctionalization.

Results

By using phylogenies to identify paralogous pairs of genes, we are able to focus our study on paralogs resulting from duplications since the separation between the rice and *Arabidopsis* lineages. A total of 91 gene clusters (Tribes) containing 1,464 genes in the PlantTribes database (<http://www.floralgenome.org/cgi-bin/tribedb/tribe.cgi>) were identified as containing at least one gene functionally characterized as a regulator of floral development. Using MP phylogenies for 91 regulatory tribes in *Arabidopsis* and rice, we identified 354 pairs of paralogs (monophyletic *Arabidopsis* sister genes) maintained in the *Arabidopsis* genome. A total of 280 pairs remained after paralogous pairs with missing microarray data were excluded (Supplementary Table 1, Supplementary Material online). Accordingly, at least 63% of the paralogous pairs in this study appear to have resulted from polyploidy events in *Arabidopsis*, with at least 53% associated with the most recent whole-genome duplication (Blanc, Hokamp, and Wolfe 2003).

The two-way ANOVA results for representative paralogous pairs are reported in figure 2, with graphs of expression levels in all six organs for one example from each possible result from the two-way ANOVA. ANOVA results for all paralogous pairs are reported in Supplementary Table 1 (Supplementary Material online). Two pairs showed no significant effects (e.g., fig. 2A). Thirteen pairs showed only a G effect (e.g., fig. 2B). Three pairs showed only an O effect (e.g., fig. 2C). Twenty-four pairs showed significant G and O effects but no significant $G \times O$ effect (e.g., fig. 2D). A significant $G \times O$ effect without significant G or O effects was observed in two paralogous pairs (e.g., fig. 2E). Twelve pairs showed a G effect coupled with a $G \times O$ effect (e.g., fig. 2F), and a significant O effect coupled with a $G \times O$ effect was observed in 22 paralogous pairs (e.g., fig. 2G). Finally, 202 pairs have a significant G, O, and $G \times O$ effect (e.g., fig. 2H). A total of 85% of paralogous pairs have a significant $G \times O$ effect at $\alpha = 0.05$, which is representative of regulatory subfunctionalization and/or neofunctionalization. This result is not significantly altered by reducing the data set to include only the most strongly supported duplicate gene pairs (bootstrap values above 80%) or considering paralogous pairs with a $G \times O$ effect significant at a more stringent level ($\alpha = 0.01$) (Supplementary Table 1, Supplementary Material online).

Because a large number of repeated tests were performed, we explored the possibility of false detection or false nondetection of significant effects in our study. A Bonferroni correction (Storey and Tibshirani 2003) assumes a priori that none of the gene pairs has a significant

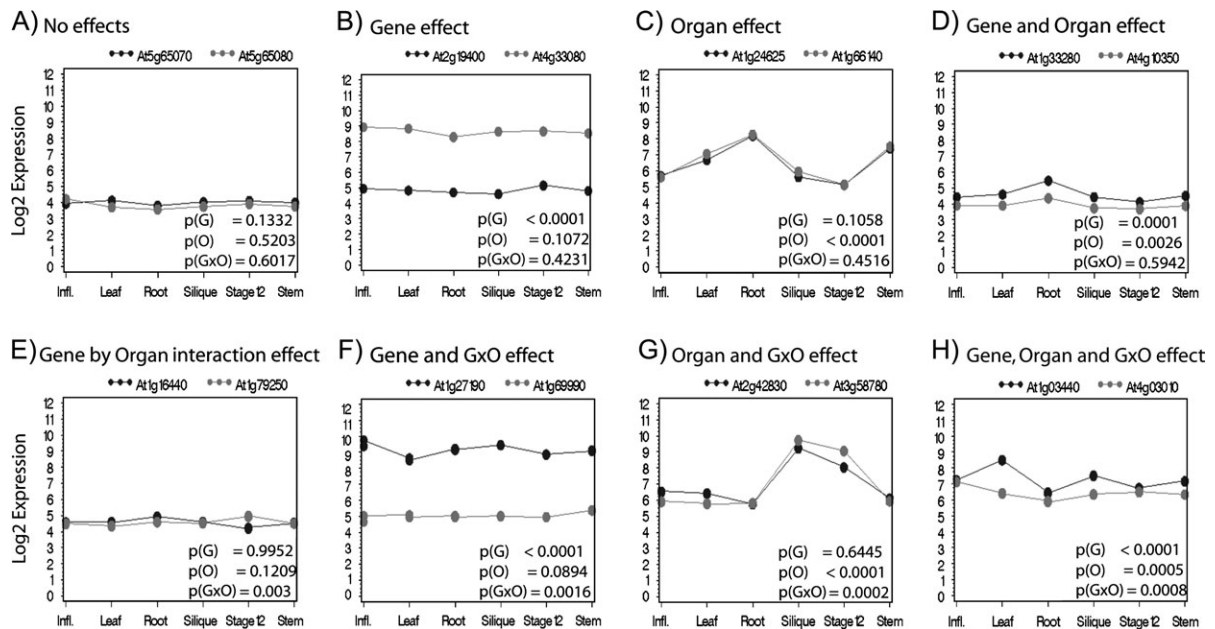


FIG. 2.—Range of ANOVA results. The depicted paralogous pairs are examples of all possible results, with the number of pairs exhibiting that result following in parentheses: (A) no significant effects (2 gene pairs); (B) G effect only (13 pairs); (C) O effect only (3 gene pairs); (D) G and O effects only (24 pairs); (E) $G \times O$ effect only (2 pairs); (F) G and $G \times O$ effects (12 pairs); (G) O and $G \times O$ effects (22 pairs); and (H) G, O, and $G \times O$ effects (202 pairs).

interaction and then very conservatively adjusts the target P value to $0.05/282$ (0.00017). With the Bonferroni correction, only 31.6% of the gene pairs have a significant $G \times O$ effect. However, there is good reason to think that this adjustment is unnecessary and introduces a considerable false nondetection bias because multiple comparisons adjustments unduly inflate the false-negative rate when the true differential expression rate is high (DeLongchamp et al. 2004). The positive false discovery rate (FDR) method improves on the Bonferroni method by estimating metaexperiment-wise “q values” in place of P values. This procedure also estimates the true percentage of significant effects for the set of tests considered (Storey and Tibshirani 2003). In this case, the estimated percentage of gene pairs in our experiments with a $G \times O$ effect is 98%. At $\alpha = 0.05$, the estimated FDR is 0.12% (0.0012), and using the Bonferroni-corrected α , the estimated FDR is 0.001%. This suggests that false nondetection is a greater problem than false detection for these data and that testing at $\alpha = 0.05$ provides better control of the overall error rate than adjusting for the familywise error rate, which assumes a priori that none of the interactions are truly significant and that the false detection rate is very tiny when tested at the uncorrected value $\alpha = 0.05$.

The number of pairs with a significant G effect (251 pairs) is very similar to the number with a $G \times O$ effect (238 pairs). However, it should be noted that not all pairs with a significant G effect have a significant $G \times O$ effect (37 pairs have a significant G effect without a significant $G \times O$ effect). This indicates that the type of expression divergence detected by correlation-based analyses is not always indicative of expression pattern shifts that would contribute to maintenance. In addition, there are 24 pairs with a significant $G \times O$ effect that do not have a significant G effect and therefore would not be identified as having significant expression divergence in a correlation-based analysis. This

illustrates the improvement in sensitivity by using a two-way ANOVA analysis versus a correlation-based analysis.

In addition, our statistical analyses identified genes that did not have a significant $G \times O$ effect and were likely candidates for regulatory nonfunctionalization. We identified 38 paralogous pairs where at least one paralog was silenced in our data set according to the following criteria: (1) low to no detectable expression and (2) no significant change in expression as measured by one-way ANOVA over all six organs. Analysis of expression data provided at the TAIR Microarray Expression database (<http://www.Arabidopsis.org>) indicated that of these 38 pairs, 20 pairs contain at least one paralog that is lowly expressed or not expressed (below 5.6). All other putative silent genes had medium to strong expression in at least one organ or treatment (<http://www.Arabidopsis.org>). The median d_N/d_S ratio for the 259 actively expressed paralogous pairs (0.160) versus the 20 paralogous pairs containing at least one silent paralog (0.255) differs significantly according to a Mann-Whitney U test ($P = 0.0007$). However, the median d_S and d_N values for actively expressed paralogous pairs and paralogous pairs containing at least one silent paralog are not significantly different according to a Mann-Whitney U test ($P = 0.094$ and $P = 0.093$, respectively).

We also identified a set of paralogous pairs that could not be easily classified according to the DDC model. There were 22 pairs in which one member of the pair was consistently expressed two- to threefold lower than the other member of the paralogous pair. Examining additional expression data from TAIR (<http://www.Arabidopsis.org>), we find this pattern to be relatively consistent in these pairs. Additionally, all these pairs do have a significant $G \times O$ effect. There are no significant differences in d_S , d_N , and d_N/d_S between these pairs and all other pairs according to a Mann-Whitney U test at $\alpha = 0.05$.

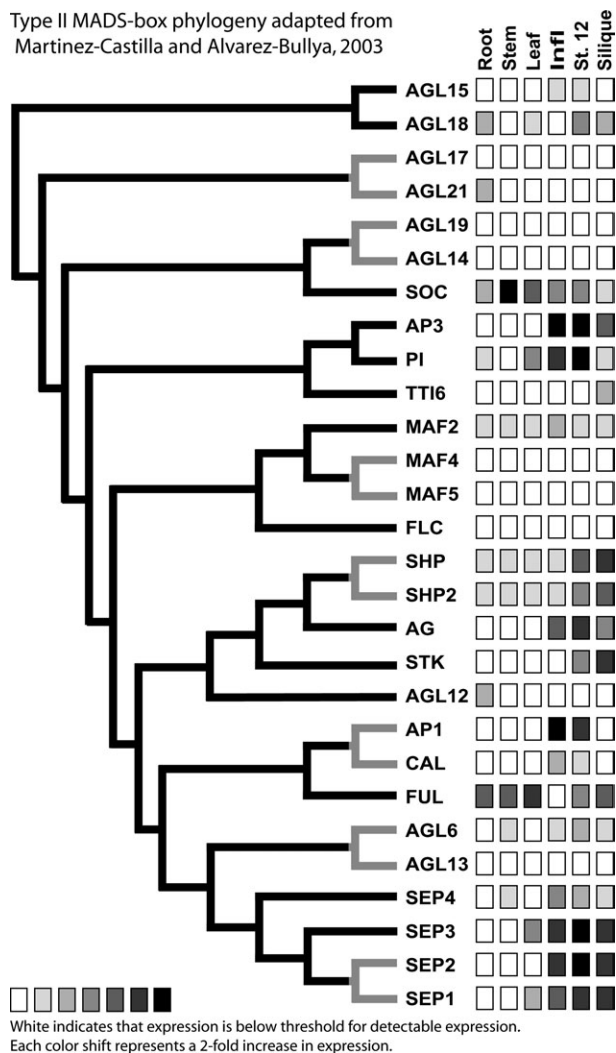


FIG. 3.—Inferring ancestral expression patterns in Type II MADS-box gene family in *Arabidopsis*. All gray pairs in the Type II MADS-box gene phylogeny (Martinez-Castilla and Alvarez-Buylla 2003) were detected in our MP phylogeny. All paralogous pairs in the Type II MADS-box gene family have a significant $G \times O$ effect, except *MAF4/MAF5*. This approach has allowed us to infer putative instances of regulatory neofunctionalization and nonfunctionalization by examining fold differences in expression between members of a paralogous pair and other closely related genes in a phylogenetic framework.

Identifying an expression pattern shift as subfunctionalization or neofunctionalization requires an approximation of the ancestral expression pattern before duplication (Force, Lynch, and Postlethwait 1999; Force et al. 1999; Lynch and Force 2000). We were able to make a detailed character reconstruction analysis for the Type II MADS-box gene family using MP (fig. 3). Six out of nine paralogous pairs were identified as having complementary expression patterns, and all pairs except one had a significant $G \times O$ effect, which is consistent with regulatory subfunctionalization and/or neofunctionalization. Eight of the nine paralogous pairs in this gene family can be mapped to either the most recent paleopolyploidy event or are in a tandem repeat. Pairs in tandem repeats have higher P values for the $G \times O$ effect (Supplementary Table 1, Supplementary Material online). Although complementary expression pat-

terns associated with pairs with a significant $G \times O$ effect are common in the Type II MADS-box gene family, there are no clear cases of regulatory subfunctionalization via the qualitative pathway of the DDC model, in which there is complementary loss of expression. However, using inference of the ancestral gene expression, regulatory neofunctionalization can be inferred for *PI*, *SHP1*, and *SHP2*. Our analysis uses successively more distant paralogs as outgroups to estimate ancestral expression patterns. An alternative and sometimes more sensitive approach would be to use corresponding expression data for one or more related species that diverged prior to the duplication event to serve as an outgroup to the paralogous pair of interest (Gu, Zhang, and Huang 2005). Unfortunately, corresponding genome-scale expression data for one or more related species are not yet available. Loss of leaf-specific expression can be inferred for *SEP2* because expression patterns for *SEP1* and *SEP3* suggest that expression in the leaf is the ancestral state. Reduction of expression is more common than complete loss of expression. Subtle neofunctionalization through increase of expression level is observed in multiple paralogs, such as *AP3* and *PI*. In one-third of paralogous pairs, the sister paralog is generally expressed two- to three-fold less than the other member of the paralogous pair, such as *API1/CAL*, *AGL6/AGL13*, and *SHP1/SHP2*.

Another important factor in the maintenance of duplicated genes is the evolution of the protein-coding region and the relationship between expression divergence and relaxation of constraint on the protein-coding region of the duplicated genes. Divergence in expression pattern was not associated with positive selection on the protein-coding portion of the genes, as detected by $d_N/d_S > 1.0$, except in a single pair that included a putative silencing event. There is a statistically significant difference in median d_N/d_S ($P = 0.014$) and d_S ($P = 0.009$) between the 237 paralogous pairs with a significant $G \times O$ effect and the 42 paralogous pairs without a significant $G \times O$ effect using a Mann-Whitney U test. However, the significant difference in median d_N/d_S may have been the result of relaxed constraint following loss of expression for one of the paralogs in some pairs. When the 20 paralogous pairs in the data set exhibiting silencing across all organs and treatments in one of the duplicates were removed from the analysis, the median d_S and d_N/d_S ratios between paralogous pairs with and without a $G \times O$ effect were not significant ($P = 0.059$ and $P = 0.279$, respectively). A series of 19 paralogous pairs with at least a twofold decrease in expression for one gene throughout all organs have no significant difference in median d_S , d_N , and d_N/d_S compared to pairs with a putative silencing event ($P = 0.768$, $P = 0.279$, and $P = 0.1189$, respectively); pairs in which both members are actively expressed ($P = 0.213$, $P = 0.921$, and $P = 0.120$, respectively); and all other pairs ($P = 0.262$, $P = 0.984$, and $P = 0.203$, respectively).

Given the mechanism for regulatory subfunctionalization in the DDC model, understanding the number and nature of the regulatory modules in the duplicate genes is necessary for distinguishing subfunctionalization and neofunctionalization. By mapping complementary expression patterns onto an organ similarity tree based on expression profiles, we were able to ascertain the relative independence

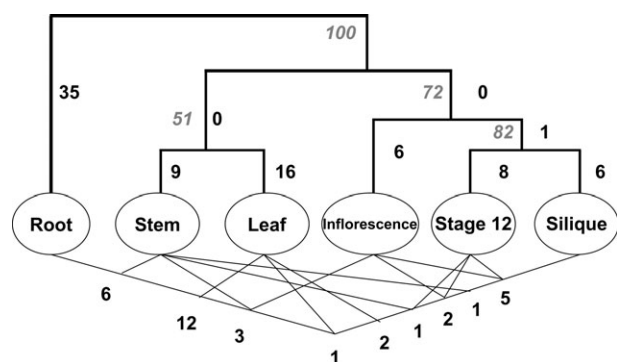


FIG. 4.—Complete linkage clustering and mapping of complementary expression patterns supports independence of regulatory modules. The tree above summarizes the results of a complete linkage cluster analysis of the six plant organs by expression level of regulatory genes and the analysis of paralogous gene pairs for complementary expression. Bootstrap support for the cluster analysis in gray and italicized. Of 280 paralogous pairs, 128 were found to have complementary expression. Black numbers indicate the number of paralogous pairs whose expression patterns differ only on organs below this branch. The network below the tree indicates the number of paralogous pairs that differ on the pairs or triples of organs connected at the nodes. For example, 35 sets exhibit complementation only in root compared with other organs; 12 sets exhibit complementation in leaf and root; and 1 set exhibits complementation in root, leaf, and silique. Eleven sets of paralogs are not shown on the graph because they have uncommon complementary expression patterns.

of expression between organs and common complementary expression patterns (fig. 4). This analysis resulted in expected patterns of organ similarity; however, bootstrap support for most nodes are relatively weak, except root versus all other organs, which has 100% bootstrap support. Reproductive organs are clustered together, with inflorescence and Stage-12 flower more similar. Almost all complementary expression patterns are mapped to the tips of the tree or have unusual patterns unrelated to organ similarity.

Discussion

Understanding Expression Pattern Shifts Using a Two-Way ANOVA

An expectation of both the classical and the DDC models for the fate of duplicated genes is that maintenance of duplicated genes will be accompanied by divergence in expression or protein structure (Ohno 1970; Force et al. 1999; Hughes 2002; Kondrashov et al. 2002; Wagner 2002). An ANOVA approach allowed us to identify and classify instances of expression divergence as they relate to the DDC model in regulatory genes following gene duplication using expression profiles for different organs. The two-way ANOVA results are interpreted as follows: (1) a G effect represents divergent quantitative expression levels between genes across all organs (e.g., fig. 2B), this is equivalent to expression divergence seen in previous correlation-based analyses and may be the result of differences in quantitative expression levels or may be the result of technical bias in microarray probe design; (2) an O effect represents different expression levels in each organ for both genes in the same paralogous pair (e.g., fig. 2C); and (3) a $G \times O$ effect represents the case that the expression levels for each locus in the paralogous pair are significantly different in spatial and quantitative terms (e.g., fig. 2E–H); this is interpreted as rep-

resentative of regulatory neofunctionalization and/or subfunctionalization. Another advantage for this approach is that ANOVA accounts for data quality because large variances will reduce the ability to identify significant effects.

Surprisingly, although most of the gene pairs in this study have evidently evolved differential gene expression patterns, few of the paralogous pairs are diverged in a way that is fully consistent with a classic subfunctionalization or neofunctionalization hypothesis of complete division of ancestral expression or acquisition of novel expression followed by loss of ancestral expression. In our results, it is uncommon to observe complete loss of expression in a given organ or treatment for one paralog while the sister paralog is expressed. Therefore, a simplistic binary model for expression shifts in regulatory duplicates, like the qualitative pathway of the DDC model (Force et al. 1999), is inadequate. For example, regulatory neofunctionalization as defined by gain of novel expression and loss of all ancestral expression seems to be extremely rare in duplicated *Arabidopsis* regulatory genes. More commonly, duplicated *Arabidopsis* regulatory genes in our study have altered the ancestral expression levels rather than strict division of ancestral expression (as seen in fig. 3). Though this type of expression shift is described as a quantitative shift in the DDC model, it is not considered to be a common pathway for divergence (Force et al. 1999). However, our results suggest that a quantitative pathway for regulatory subfunctionalization is frequently taken by duplicated regulatory genes. In other cases, altered expression levels include duplicate pairs where a single paralog is expressed at a higher level in a specific organ than the other paralogs or the ancestral gene (see *API/CAL* in fig. 3). This increase in expression may change the role of the gene, similar to regulatory neofunctionalization. However, this expression change is not entirely novel because the ancestral gene was expressed in the same organ. In addition, a paralog that does gain novel expression in a specific organ or condition may also retain part of the ancestral expression pattern. Such a mixture of subfunctionalization and neofunctionalization has been previously noted in studies using yeast and human duplicates (He and Zhang 2005) and computational models (Rastogi and Liberles 2005).

Another circumstance deviating from the DDC model concerns genes that are lowly expressed compared to their sister paralogs, but which still exhibit expression shifts that would contribute to maintenance (indicated by a $G \times O$ effect) and selective constraint on the coding sequence ($d_N/d_S < 1.0$). We have assigned these genes to a classification of “regulatory hypofunctionalization,” a special case of subfunctionalization that does not follow the typical pattern of subfunctionalization outlined by the DDC model. Examination of our data set, combined with additional expression data from TAIR, indicates that these are uncommon (22 pairs out of 280). This includes the *API/CAL*, *AGL6/AGL13*, *SEP1/SEP2*, and *SHP1/SHP2* paralogous pairs from the Type II MADS-box gene family, which are critical regulators of floral development (Becker and Theissen 2003; Irish 2003). Regulatory hypofunctionalization is a specialized case of subfunctionalization, where instead of splitting the expression pattern equally as predicted by the DDC model, expression in one of the paralogs is

greatly diminished (by at least two- to threefold) compared to the other paralogs in almost all organs and treatment conditions. In this case, it seems that the minor paralog is maintained through selection for genetic robustness (Moore, Grant, and Purugganan 2005), or loss of selective maintenance of the minor paralog has been so recent that the gene is not silenced across all organs. The maintenance of “redundant” duplicate genes can contribute to the robustness of the genetic network by reducing the fitness effect of deleterious mutations (Gu 2003). Given the importance of the floral developmental pathway, we may expect that regulatory hypofunctionalization as “protection” against deleterious mutations should be common among floral regulators. Another explanation is that expression differences between paralogs are among cells within an organ and, therefore, not detectable in typical microarray experiments. In any case, the minor paralog has persisted, but the evolutionary mechanism responsible for persistent low-level expression is not easily ascribed to regulatory subfunctionalization or neofunctionalization.

Complete regulatory nonfunctionalization is difficult to ascertain. Because of limited organ sampling, it is also possible that these putative silencing events are regulatory genes that have highly restricted expression patterns that could not be detected in the microarray results. However, given the breadth of the microarray experiments accessed through TAIR, the proportion of putative silent loci that are not truly silent is expected to be small. There is a significant difference in the median d_N/d_S ratio between paralogous pairs of actively expressed regulatory genes and paralogous pairs with a putatively silent regulatory gene, suggesting that actively expressed regulatory genes are evolving under purifying selection, whereas regulatory nonfunctionalization (silencing) is associated with a relaxation of evolutionary constraint on protein-coding portions of a regulatory gene.

What is the functional significance of these results? While some pairs show threefold differences or greater, others have more subtle differences. The level of divergence required to affect function is dependent on the particular gene, the cellular context, and the environment. Therefore, functional significance would need to be assessed in each case. Where functional studies are available, major expression difference is clearly correlated with functional differences. For example, in the Type II MADS-box gene family, *API* has higher levels of expression than does *CAL*, consistent with *API* playing a more prominent role in specifying floral meristem and organ identities than *CAL* (Mandel 1992; Kempin 1995). On the other hand, subtle expression differences may not be related to dramatic functional divergence detectable in laboratory conditions, as in the case of *SEP1* and *SEP2*, and *SHP1* and *SHP2*, which seem to have completely redundant roles in controlling flower and fruit development, respectively (Liljegren 2000; Pelaz 2000). Whereas the expression pattern shifts found in this study may not result in an altered phenotype in a single knockout mutant, it is important to recognize that laboratory conditions are only a small subset of conditions that the organism has experienced in its evolutionary history. Under natural conditions, these subtle expression pattern shifts could affect fitness over evolutionary time and therefore could be evolutionarily important expression

shifts (Weinig 2003). Alternatively, some of the subtle changes might represent evolutionarily “transient” states of truly redundant paralogs that will diverge functionally in the future.

Expression Divergence Is Not Coupled with Global Protein Constraint

The classical and DDC models for the fate of duplicated genes assert that duplicated genes will only be maintained when changes occur in the regulation or protein activity of the paralogs which result in neofunctionalization or subfunctionalization in the case of the DDC model (Ohno 1970; Force et al. 1999; Hughes 2002; Kondrashov et al. 2002; Wagner 2002). Based on these models, we may expect retained paralogs without evidence of regulatory subfunctionalization and/or neofunctionalization to have significant changes in the protein sequence. However, the lack of significant difference in mean d_N/d_S between those paralogous pairs with a $G \times O$ effect and paralogous pairs without a $G \times O$ effect (with putative silencing events removed) suggests independent roles for gene regulation and protein activity in the maintenance of duplicated regulatory genes and that survival of a duplicate gene pair is not solely dependent on expression or protein changes. This agrees with previous studies that show no relation between expression divergence and protein divergence (Wagner 2000; Gu et al. 2002; Makova and Li 2003). It should be noted that this methodology for determining d_S and d_N (Yang 1997) has a low statistical power and only provides a general overview of protein constraint within the paralogous pair. It is still plausible for local adaptive protein evolution to occur within these paralogous pairs (Yang et al. 2000; Nam et al. 2005). Local adaptive evolution could complement changes at the regulatory level or eliminate selective pressure for expression pattern shifts.

Complementary Expression Patterns in Duplicated Regulatory Genes

By identifying the organs in which complementation occurs and comparing it to complete linkage clustering of organs based on expression levels, we can determine whether related organs may share the same regulatory module for expression or if regulatory modules are independent (fig. 4). In this case, the mapping of most complementary patterns to the tips of the tree supports independence of regulatory modules that coordinate a specific spatial or temporal expression pattern (Wray et al. 2003) and also suggests that many duplicated regulatory genes may have been expressed in a limited number of organs prior to duplication. These conclusions apply to approximately half of the paralogous pairs with a significant $G \times O$ effect.

The 122 paralogous pairs with a significant $G \times O$ effect but without complementation suggest that these non-complementary expression patterns evolve after duplication through a quantitative pathway; in this case, both copies are required at the same time in order to meet ancestral functional thresholds for proper regulation of genes downstream in a regulatory network (Force et al. 1999). Complementation may also not be detectable if subfunctionalization or neofunctionalization are occurring within organs included

in this study. However, there is also a possibility that the quantitative measures of expression are inconsistent, and a constant difference in expression across all organs may be a result of poor probe design or resolution for one of the paralogs. In this case, the ANOVA would detect significant G, O, and $G \times O$ effects, but a complementary expression pattern would not be present. Overall, the ANOVA approach, by identifying different components of variance in expression data, can identify unique relationships between expression patterns in a paralogous pair and is not confounded by technical bias in microarray data that can affect the quantitative expression level.

Role of Gene Duplication in the Evolution of Plant Regulatory Networks

Our hypothesis that duplicated regulatory genes in *Arabidopsis* will have evidence of expression pattern shifts that contribute to maintenance (regulatory subfunctionalization and/or neofunctionalization) is supported by the majority of paralogous pairs of regulatory genes which have a significant $G \times O$ effect. Therefore, we conclude that our global analysis supports a hypothesis that the molecular evolution of regulatory proteins in *Arabidopsis* is significantly impacted by regulatory subfunctionalization and neofunctionalization after duplication and that this hypothesis will apply to other angiosperms.

Given the prevalence of gene and genome duplication in the evolutionary history of plants, evolution of development in angiosperms may differ from organisms where genome duplication is rare and extensive expression pattern shifts after duplication would have a profound impact on the evolution of developmental and regulatory networks. This evolutionary scenario comes with testable predictions. Lineages with fewer detectable gene or genome-wide duplication events may be expected to have less specialized regulatory networks and regulatory genes with broader function.

Supplementary Material

A full listing of the paralogous pairs used in this study, along with their respective ANOVA *P* values, status of complementary expression, d_N , d_S , d_N/d_S ratio, amino acid identity, DNA sequence identity, and gene family annotation are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Jiong Wang and Anthony Omeis for plant care. We are grateful for helpful comments from Kateryna Makova, Victor Albert, and anonymous reviewers. The support of the National Science Foundation Plant Genome Research Program for the Floral Genome Project (DBI-0115684) and a university fellowship from The Pennsylvania State University (J.M.D.) are gratefully acknowledged.

Literature Cited

Adams, K. L., R. Cronn, R. Percifield, and J. F. Wendel. 2003. Genes duplicated by polyploidy show unequal contributions

- to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. USA* **100**:4649–4654.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Anderbert, M. R. 1973. Cluster analysis for applications. Academic Press, New York.
- Becker, A., and G. Theissen. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* **29**:464–489.
- Blanc, G., K. Hokamp, and K. H. Wolfe. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* **13**:137–144.
- Blanc, G., and K. H. Wolfe. 2004a. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**:1679–1691.
- . 2004b. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**:1667–1678.
- Bomblied, K., R. L. Wang, B. A. Ambrose, R. J. Schmidt, R. B. Meeley, and J. Doebley. 2003. Duplicate FLORICAULA/LEAFY homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development* **130**: 2385–2395.
- Delongchamp, R. R., J. F. Bowyer, J. J. Chen, and R. L. Kodell. 2004. Multiple-testing strategy for analyzing cDNA array data on gene expression. *Biometrics* **60**:774–782.
- Doebley, J., and L. Lukens. 1998. Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**:1075–1082.
- Edgar, R. C. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113.
- Enright, A. J., V. Kunin, and C. A. Ouzounis. 2003. Protein families and TRIBES in genome sequence space. *Nucleic Acids Res.* **31**:4632–4638.
- Enright, A. J., S. Van Dongen, and C. A. Ouzounis. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**:1575–1584.
- Felsenstein, J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**:164–166.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**:1531–1545.
- Force, A., M. Lynch, and J. Postlethwait. 1999. Preservation of duplicate genes by subfunctionalization. *Am. Zool.* **39**:78A.
- Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, and J. Gentry. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**:R80.
- Gu, X. 2003. Evolution of duplicate genes versus genetic robustness against null mutations. *Trends Genet.* **19**:354–356.
- . 2004. Statistical framework for phylogenomic analysis of gene family expression profiles. *Genetics* **167**:531–542.
- Gu, X., Z. Zhang, and W. Huang. 2005. Rapid evolution of expression and regulatory divergences after yeast gene duplication. *Proc. Natl. Acad. Sci. USA* **102**:707–712.
- Gu, Z. L., D. Nicolae, H. H. S. Lu, and W. H. Li. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* **18**:609–613.
- He, X., and J. Zhang. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**:1157–1164.
- Hekstra, D. 2003. Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. *Nucleic Acids Res.* **31**:1962–1968.
- Hileman, L. C., and D. A. Baum. 2003. Why do paralogs persist? Molecular evolution of CYCLOIDEA and related floral

- symmetry genes in Antirrhineae (Veronicaceae). *Mol. Biol. Evol.* **20**:591–600.
- Hughes, A. L. 2002. Adaptive evolution after gene duplication. *Trends Genet.* **18**:433–434.
- Irish, V. F. 2003. The evolution of floral homeotic gene function. *Bioessays* **25**:637–646.
- Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**:249–264.
- Jin, W., R. M. Riley, R. D. Wolfinger, K. P. White, G. Passador-Gurgel, and G. Gibson. 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* **29**:389–395.
- Kempin, S. A. 1995. Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* **267**:522–525.
- Kondrashov, F. A., I. B. Rogozin, Y. I. Wolf, and E. V. Koonin. 2002. Selection in the evolution of gene duplications. *Genome Biol.* **2**:8.1–8.9.
- Kramer, E. M., M. A. Jaramillo, and V. S. Di Stilio. 2004. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **166**:1011–1023.
- Lassmann, T., and E. L. L. Sonnhammer. 2002. Quality assessment of multiple alignment programs. *FEBS Lett.* **529**:126–130.
- Lee, C., C. Grasso, and M. F. Sharlow. 2002. Multiple sequence alignment using partial order graphs. *Bioinformatics* **18**:452–464.
- Liljegren, S. 2000. SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**:766–770.
- Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS system for mixed models. SAS Institute Inc., Cary, N.C.
- Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151–1155.
- Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**:459–473.
- Maddison, D. R., and W. P. Maddison. 2001. MacClade 4.0: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Mass.
- Maere, S., S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu, M. Kuiper, and Y. Van de Peer. 2005. Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. USA* **102**:5454–5459.
- Makova, K. D., and W. H. Li. 2003. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13**:1638–1645.
- Mandel, M. A. 1992. Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. *Nature* **360**:273–277.
- Martinez-Castilla, L. P., and E. R. Alvarez-Buylla. 2003. Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Natl. Acad. Sci. USA* **100**:13407–13412.
- Matsunaga, S., E. Isono, E. Kejnovsky, B. Vyskot, J. Dolezel, S. Kawano, and D. Charlesworth. 2003. Duplicative transfer of a MADS box gene to a plant Y chromosome. *Mol. Biol. Evol.* **20**:1062–1069.
- Moore, R. C., S. R. Grant, and M. D. Purugganan. 2005. Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **22**:91–103.
- Nam, J., K. Kaufmann, G. Theissen, and M. Nei. 2005. A simple method for predicting the functional differentiation of duplicate genes and its application to MIKC-type MADS-box genes. *Nucleic Acids Res.* **33**:e12.
- Ohno, S. 1970. Evolution by gene duplication. Springer, New York.
- Osborn, T. C., J. C. Pires, J. A. Birchler et al. (11 co-authors). 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* **19**:141–147.
- Pelaz, S. 2000. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405**:200–203.
- Rastogi, S., and D. A. Liberles. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* **5**:28.
- Simillion, C., K. Vandepoele, M. C. E. Van Montagu, M. Zabeau, and Y. Van de Peer. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **99**:13627–13632.
- Spath, H. 1980. Cluster analysis algorithms. Ellis Horwood, Chichester, United Kingdom.
- Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genome-wide experiments. *Proc. Natl. Acad. Sci. USA* **100**:9440–9445.
- Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.
- Thompson, J. D., J. C. Thierry, and O. Poch. 2003. RASCAL: rapid scanning and correction of multiple sequence alignments. *Bioinformatics* **19**:1155–1161.
- Vandepoele, K., C. Simillion, and Y. Van de Peer. 2002. Detecting the undetectable: uncovering duplicated segments in *Arabidopsis* by comparison with rice. *Trends Genet.* **18**:606–608.
- Vision, T. J., D. G. Brown, and S. D. Tanksley. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* **290**:2114–2117.
- Wagner, A. 2000. Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc. Natl. Acad. Sci. USA* **97**:6579–6584.
- . 2002a. Selection and gene duplication: a view from the genome. *Genome Biol.* **3**:1012.1011–1012.1013.
- . 2002b. Asymmetric functional divergence of duplicate genes in yeast. *Mol. Biol. Evol.* **19**:1760–1768.
- Weinig, C. 2003. Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* **165**:321–329.
- Wendel, J. F. 2000. Genome evolution in polyploids. *Plant Mol. Biol.* **42**:225–249.
- Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman, and L. A. Romano. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**:1377–1419.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
- Yang, Z., N. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.
- Zahn, L. M., H. Kong, J. Leebens-Mack, S. Kim, P. S. Soltis, L. L. Landherr, D. E. Soltis, C. W. dePamphilis, and H. Ma. 2005a. The evolution of the SEPALLATA subfamily of MADS-box genes: a pre-angiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**:2209–2223.
- Zahn, L. M., J. Leebens-Mack, C. W. dePamphilis, H. Ma, and G. Theissen. 2005b. To B or not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J. Hered.* **96**:225–240.
- Zhang, X., B. Feng, Q. Zhang, D. Zhang, N. Altman, and H. Ma. 2005. Genome-wide expression profiling and identification of gene activities during early flower development in *Arabidopsis*. *Plant Mol. Biol.* **58**:401–419.

Douglas Crawford, Associate Editor

Accepted November 2, 2005