Relaxed Selection Among Duplicate Floral Regulatory Genes in Lamiales

Jan E. Aagaard,[1,3] **John H. Willis,**[2] **Patrick C. Phillips**[1]

[1] Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR 97403, USA
[2] Department of Biology, Duke University, Durham, NC 27708, USA
[3] Department of Genome Sciences, University of Washington, Box 357730, Seattle, WA 98195, USA

Abstract. Polyploidization is a prevalent mode of genome diversification within plants. Most gene duplicates arising from polyploidization (paralogs) are typically lost, although a subset may be maintained under selection due to dosage, partitioning of gene function, or acquisition of novel functions. Because they experience selection in the presence of other duplicate loci across the genome, interactions among genes may also play a significant role in the maintenance of paralogs resulting from polyploidization. Previously, we identified duplicates of the genes *LFY/FLO* and *AP3/DEF* that directly interact in a floral regulatory pathway and are thought to be the result of ancient polyploidization in the Lamiales (> 50 mya). Although duplicates of MADS box genes including *AP3/DEF* are common throughout the angiosperm lineage, *LFY/FLO* duplicates in Lamiales are the first reported outside of tetraploid taxa. In order to explore hypotheses for the joint preservation of these interacting floral regulatory genes including novel *LFY/FLO* paralogs, here we clone *FLO* and *DEF* duplicates from additional Lamiales taxa and apply codon substitution models to test how selection acts on both genes following duplication. We find acceleration in the ratio of nonsynonymous-to-synonymous nucleotide substitutions for one (*FLO*) or both (*DEF*) paralogs that appears to be due to relaxed purifying selection as opposed to positive selection and shows a different pattern among functional domains of these genes. Several mechanisms are discussed that might be responsible for preservation of co-orthologs of *FLO* and *DEF* in Lamiales, including interactions among the genes of this regulatory pathway.

Key words: Gene duplication — Floral regulatory genes

Correspondence to: Jan E. Aagaard; *email:* jaagaard@gs.washington.edu

Introduction

Polyploidization occurs frequently in flowering plants. Estimates of the frequency of polyploidization based on cytological studies vary widely but suggest up to 70% of angiosperms may be polyploids (Wendel 2000; Otto and Whitton 2000). Analysis of cDNA and genomic sequences similarly shows that both recent and ancient polyploidization among crop and model plant taxa are common (Blanc and Wolfe 2004a). The relative importance of polyploidization to the evolution and diversification of flowering plants (angiosperms) has been extensively debated (e.g., Stebbins 1971; Levin 1983) and is reviewed elsewhere (Otto and Whitton 2000). It is clear, however, that polyploidization contributes significantly to the long-term genomic content of angiosperm taxa. For example, more than half of all genes from *Arabidopsis thaliana* are thought to be the result of ancient polyploidization (Maere et al. 2005), with approximately 20% of predicted genes resulting from a whole-genome duplication event that occurred 24–40 million years ago (mya; Blanc et al. 2003; Blanc and Wolfe 2004b).

Because the contribution of polyploidization to plant evolution is ultimately a function of the duplicate genes that are maintained within genomes, it is important to understand both the patterns and mechanisms of duplicate gene preservation. The majority of duplicates (paralogs) arising from polyploidization in plants are silenced. In the case of *A. thaliana*, nearly 70% of paralog pairs from the last polyploid event have lost one member (Blanc et al. 2003). Although there is some suggestion that paralogs resulting from polyploidization may persist longer (Otto and Whitton 2000; Moore and Purugganan 2003; Lynch and Conery 2001), this is generally consistent with patterns for smaller-scale duplication events (e.g., tandem or segmental duplications) which suggest that silencing due to loss of function mutations predominates (Walsh 1995), resulting in a half-life for fully redundant paralogs which is generally less than 10 million years (Lynch and Conery 2000). However, the probability that one member of a paralog pair is lost appears to differ markedly among functional classes of genes depending on the mechanism of gene duplication. For example, in *A. thaliana* paralogs of regulatory genes involved in transcription and signal transduction are retained at a higher rate if duplicates are the result of polyploidization rather than smaller-scale gene duplication events (Maere et al. 2005).

Paralogs that escape the typical fate of silencing and are maintained within genomes under selection may be preserved via a number of mechanisms. Theory suggests that preservation of paralogs is primarily due to parsing of the ancestral single-copy gene's functions among duplicates (subfunctionalization) rather than the acquisition of novel gene function driven by positive selection, based in large part on the predominance of degenerate rather than adaptive mutations (e.g., Force et al. 1999). Subfunctionalization can result from complementary degenerate mutations restricted to regulatory regions as in the model of Force et al. (1999) or involve functional domains within protein coding regions, resulting in an altered pattern of amino acid substitution between duplicates (Dermitzakis and Clark 2001). Because paralogs arising from polyploidization experience selection along with other duplicate loci across the genome, interactions among loci may also contribute significantly to their maintenance. Such interactions could result from simple dosage effects, which should predominate among genes functioning within metabolic pathways and regulatory genes involved in signal transduction and transcription (Birchler et al. 2001). Alternately, divergence in the timing or expression domains among pairs of paralogs could

have broad impact on duplicate gene preservation across the polyploid genome through concerted divergence of the interacting members of genetic pathways and gene networks (Blanc and Wolfe 2004b), essentially a multilocus extension of subfunctionalization models.

Previously, we identified duplicates of several regulatory genes important in floral development within the order of tricolpate plants Lamiales (Aagaard et al. 2005). These include co-orthologs of the floral meristem identity gene *FLO* (Coen et al. 1990) and floral homeotic MADS box gene *DEF* (Sommer et al. 1990) from the model Lamiales species *Antirrhinum majus*. Orthologs of *FLO* and *DEF* from *A. thaliana* (*LFY* and *AP3*, respectively [Weigel et al. 1992; Jack et al. 1992]) carry out similar functions in the floral regulatory pathway as in *A. majus*, where *LFY/FLO* positively regulates the expression of the downstream gene *AP3/DEF* (Ingram et al. 1997; Weigel and Meyerowitz 1993). Duplication of the *A. majus* genes *FLO* and *DEF* in Lamiales is thought to be the result of a whole-genome duplication (polyploidization) which occurred after the split between the ancestor of *A. majus* (Veronicaceae) and the lineage leading to many of the other families of Lamiales including Verbenaceae, Paulowniaceae, Phrymaceae, and Orobanchaceae (Aagaard et al. 2005). Co-orthologs of both genes (*FLOA* and *FLOB*, *DEFA* and *DEFB*) are expressed and appear functional in multiple taxa from these four families, and have been maintained within multiple lineages that diverged 30–50 mya (Wickstrom et al. 2001). Significantly, duplicates of *AP3/DEF* from several points along the angiosperm phylogeny are known (Theissen et al. 2000; Kramer et al. 1998, 2003), although Lamiales represents an as yet novel case of ancient *LFY/FLO* paralogs (Cronk 2001).

The purpose of the present study is to explore hypotheses regarding the evolutionary forces acting among these paralogs that may be responsible for the maintenance of duplicate copies. Here we clone additional co-orthologs of the floral regulatory genes *FLO* and *DEF* from four taxa within the Lamiales family Phrymaceae (*sensu* Beardsley and Olmstead 2002) to supplement our earlier work (Aagaard et al. 2005) in order to have sufficient power for likelihood-based statistical tests employing codon substitution models (Aagaard and Phillips 2005). We then test for rate heterogeneity and the role of selection in divergence among *FLO* and *DEF* paralogs following duplication in Lamiales using several different codon substitution models. Based on these results, we explore several mechanisms to explain the maintenance of the joint preservation of paralogs from multiple genes that are interacting members of the floral regulatory pathway.
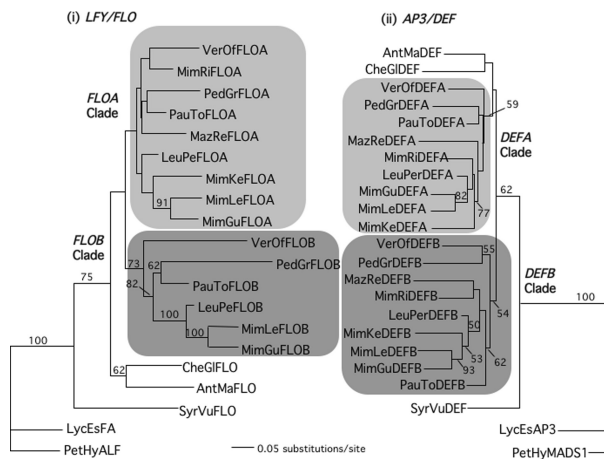
Fig. 1. Likelihood trees for homologs of (i) *LFY/FLO* and (ii) *AP3/DEF* cloned in this study and previously (see Supplementary Table S1). Paralogy groups of *LFY/FLO* and *AP3/DEF* within the Lamiales are designated *A* or *B* and are consistent with previous phylogenetic analyses of gene duplication within Lamiales (Aagaard et al. 2005). Likelihood searches were performed under the general time reversible model with four rate categories using PAUP (Swofford 2002). All ambiguous regions of alignments were excluded from analyses. Bootstrap support based on 100 bootstrap replicates is shown for nodes with ≥ 50% support.

Materials and Methods

Cloning Floral Regulatory Genes in Lamiales

We selected four taxa within the family Phrymaceae (Beardsley and Olmstead 2002) including *Mazus reptans*, *Mimulus ringens*, *Leucocarpus perfoliatus*, and *Mimulus kelloggii* (see Supplementary Table S1) to supplement paralogs of *FLO* and *DEF* previously cloned (Aagaard et al. 2005) from *Syringa vulgaris* (Oleaceae), *Chelone glabra* and *Antirrhinum majus* (Veronicaceae), *Verbena officinalis* (Verbenaceae), *Mimulus guttatus* and *Mimulus lewisii* (Phrymaceae), *Paulownia tomentosa* (Paulowniaceae), and *Pedicularis groenlandica* (Orobanchaceae). Source material was in horticultural propagation (*M. reptans*; J. Aagaard 2004-1, WTU) or was grown from seed stocks used in a previous study (*M. ringens*, *L. perfoliatus*, and *M. kelloggii* [Beardsley and Olmstead 2002]). From a single plant of each species genomic DNA was prepared from leaf material using the CTAB extraction of Kelly and Willis (1998), and total RNA prepared from flower buds across a range of developmental stages using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA). RACE Ready cDNA pools (5′- and 3′-cDNA) were made from floral bud total RNA (Clontech, Palo Alto, CA).

LFY/FLO- and *AP3/DEF*-like genes were cloned from all four species of Phrymaceae using the same protocol as in Aagaard et al. (2005). This involved a two-step approach for *LFY/FLO* that (i) screened for duplicate copies based on length polymorphism in introns of genomic DNA and (ii) tested for expression of genes while also extending coding sequences using 5′- and 3′-RACE from floral bud cDNA pools. Briefly, we used two degenerate PCR primer sets nested within conserved exon domains spanning the first (5′ primer, 5′-ATGAGGGATGAGGAGCTTGATSANATGATGRA-3′; 3′ primer, 5′-GCTCCGTCACGATAAANGGRTGYT-3′) and second (5′ primer, 5′-CGGCAGCGGGAGCAYCCNTTYAT-3′; 3′ primer, 5′-GCGTTGAAGATCGCRTCDATRTCC-3′) introns of *FLO* to amplify genomic fragments, which were gel purified, cloned, and sequenced (at least four clones per genomic PCR fragment). cDNAs corresponding to all genomic fragments were cloned and sequenced using 3′-RACE (four or more clones per 3′-RACE PCR). For unique 3′-RACE clones corresponding to genomic clones

spanning intron 2, 5′-RACE was used to extend sequence to the 5′-termini of coding regions. In contrast, *AP3/DEF*-like homologs were cloned directly from 3′-RACE cDNA pools. Briefly, a single degenerate primer at the 5′ terminus of the *DEF* coding region (5′-ATGGCTCGTGGGAAGATHCARAT-3′) was used in 3′-RACE, after which PCR products were directly cloned and sequenced (at least 12 clones sequenced per 3′RACE PCR).

Nucleotide alignments of the newly cloned *LFY/FLO*- and *AP3/DEF*-like genes from Phrymaceae as well as Solanales and Lamiales homologs (Supplementary Table S1) were carried out initially based on the translated nucleotide (protein) sequence using the ClustalX algorithm implemented in BioEdit (T. Hall, North Carolina State University, Raleigh), followed by visual alignment. All gapped regions were removed and sequences analyzed under likelihood criteria implemented in PAUP (Swofford 2002). Likelihood analyses employed the general time-reversible model with four rate categories, estimating the gamma shape parameter and the proportion of invariable sites ($GTR + 1 + \gamma$). Heuristic search criteria included TBR branch swapping with 10 random addition replicates. Support for nodes was estimated by 100 bootstrap replicates using the search criterion above. Based on phylogenetic analyses, *LFY/FLO*- and *AP3/DEF*-like genes cloned here were assigned to the paralog A or B clades identified in Aagaard et al. (2005).

Codon Substitution Models

Branch and Fixed-sites Models. We estimated nonsynonymous-to-synonymous substitution ratios (ω) at different points during the evolution of floral genes and among gene clades using two different classes of models. Both classes of models allow for positive selection ($\omega > 1$) but have low power to detect it when selection acts at a small proportion of codons because they average ω among sites (Yang 2001). First, we used the branch models of Yang (1998). A gene phylogeny was constructed *a priori* based on familial (Olmstead et al. 2001) and species (Beardsley and Olmstead 2002) relationships, and based on Aagaard et al. (2005) and the phylogenetic analyses of newly cloned *LFY/FLO*- and *AP3/DEF*-like genes described above (Fig. 1). The gene tree (Fig. 2) was partitioned into pro-orthologs (ω_0) and branches immediately-orthologs following the duplication event/events (ω_1) or paralog A and B clades (ω_2 and ω_3, respectively). The same sequence alignments used in phylogenetic analyses were analyzed using branch models constrained by the *a priori* gene phylogeny allowing for (model R1) a single ω ratio among branches and clades as well as nested models allowing two (model R2), three (model R3), or four (model R4) ω's for different clades in the gene phylogeny. All models were implemented using PAML, version 3.14 (Yang 1997). Likelihood ratio test (LRT) statistics were calculated as twice the difference between log-likelihoods for more complex versus nested less complex model, and compared with a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between models (Goldman and Yang 1994).

Second, we used the fixed-sites models of Yang and Swanson (2002) to compare ω's directly between paralogs. For this test, only sequences from species for which both paralogs had been cloned were used. This included six taxa for the comparison between *FLOA* and *FLOB* (*Verbena officinalis*, *Pedicularis groenlandica*, *Paulownia tomentosa*, *Leucocarpus perfoliatus*, *Mimulus lewisii*, and *Mimulus guttatus*) and nine taxa for the comparison between *DEFA* and *DEFB* (*Verbena officinalis*, *Pedicularis groenlandica*, *Paulownia tomentosa*, *M. reptans*, *M. ringens*, *Leucocarpus perfoliatus*, *M. kelloggii*, *Mimulus lewisii*, and *Mimulus guttatus*). Sequences were aligned as previously and analyzed using fixed-sites models C and E (Yang and Swanson 2002), which constrain both paralogs to a single ω (model C) or allow for separate ω's for paralogs (model E). LRT statistics were calculated as twice the
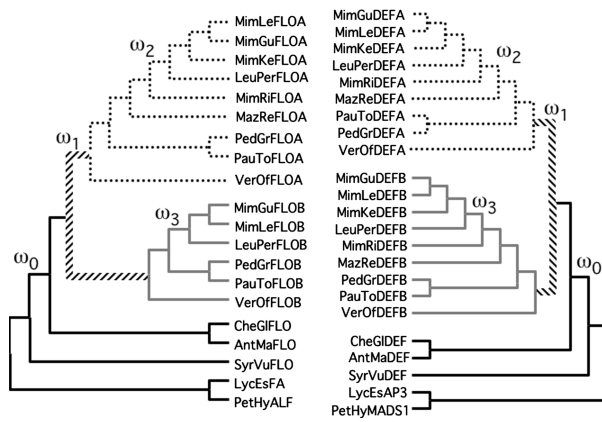
4

Fig. 2. Representative gene phylogenies for *FLO* and *DEF* homologs from Lamiales and Solanales (see Supplementary Table S1). The relationships among genes are based on familial (Olmstead et al. 2001) and species (Beardsley and Olmstead 2002) relationships and based on the study of other Lamiales homologs (Aagaard et al., submitted) and phylogenetic analyses of genes cloned here (Fig. 1). Branches are partitioned into those for pro-orthologs of duplicate *FLO* and *DEF* genes (ω_0), branches immediately following the duplication event (ω_1), and those for A or B paralogs (ω_2 and ω_3, respectively). The phylogenetic relationships depicted in these trees are used in all likelihood analyses of codon substitution models, though specific subscripts for ω only correspond to those for branch and fixed-sites models (Table 1).

difference between model E and C and compared with a chi-squared distribution with two degrees of freedom.

Sites and Branch-site Models. We used two classes of codon substitution models that have increased power relative to branch or fixed-sites models (Yang 2001) to test for the action of positive selection during the evolution of floral genes. First, we applied sites models (Nielsen and Yang 1998; Yang et al. 2000; Yang and Swanson 2002) that hold ω constant among all branches of the gene phylogeny while allowing ω to take on multiple values among site classes. Sites models included M1a, M2a, M7, and M8. Second, we applied the modified branch-sites models of Zhang et al. (2005). For branch-sites models, the branches immediately following the duplication events resulting in the *FLOA* and *FLOB* or *DEFA* and *DEFB* clades (ω_1; Fig. 1) were specified as foreground branches, thus allowing for different ω's both over a subset of branches in the gene phylogeny and sites. This constraint has been used previously to test for positive selection following gene duplication (Bielawski and Yang 2003) and reflects a model of duplicate gene evolution whereby positive selection acts initially on duplicates followed by purifying selection in ancestral lineages (Hughes 1999). We used the modified branch-sites Model A of Zhang et al. (2005) in order to avoid problems with an elevated type 1 error rate reported for older branch-sites models (Yang and Nielsen 2002), including a nested model with $\omega_2 = 1$ fixed (Model A_{null}). A likelihood test comparing Model A with Model A_{null} constitutes test 2 of Zhang et al. (2005). Sites and branch-sites models were implemented using PAML, version 3.14 (Yang 1997). Because of multiple local optima in likelihood searches, sites and branch-sites models were run a minimum of three times using different initial ω's.

We also used sites models in combination with an empirical Bayes approach along with fixed-sites and branch models to investigate how ω differs among functional regions of *FLO* and *DEF* paralogs. Functional domains of MADS box genes including *DEF* are well characterized (e.g., Reichman and Meyerowitz 1997) including the MADS, I, K, and C-terminus regions; for analyses we treated the DNA binding domain (MADS) as distinct from other

regions (I, K, and C-terminus). Functional domains of *FLO* are not as well characterized, but *FLO* is a transcription factor (Coen et al. 1990) and the highly conserved C-terminus region of the *Arabidopsis thaliana* ortholog *LFY* is reported to bind DNA (Gocal et al. 2001) and has been suggested to constitute the DNA binding domain (Bomblies et al. 2003). For our analyses we treated the putative DNA binding domain as distinct from the N-terminus region of *FLO*. Based on ML estimates for 11 rate categories (10 constrained between 0 and 1, and 1 allowing $\omega > 1$) from sites model M8, posterior probabilities for each site (codon) were estimated using the empirical Bayes approach of Nielsen and Yang (1998), implemented in PAML version 3.14 (Yang 1997). Estimated ω's are calculated as the average over site classes weighted by the posterior probability that the codon belongs to that site class. The mean of these ω's within function domains provides a qualitative assessment of rate variation among regions. For statistical tests of variation among regions, we used fixed-sites models C and E comparing DNA binding versus all other domains. Fixed sites models were applied to all sequences (Supplementary Table S1) of *FLO* and *DEF* as well as limited to comparisons between paralogs (*FLOA* versus *FLOB*). In addition, branch models R1 and R2, including only the MADS DNA binding domain or only the non-MADS regions (I, K, and C), were fit to the *DEF* data from all sequences.

Results

LFY/FLO- and AP3/DEF-like Genes in Phrymaceae

We cloned five *LFY/FLO*-like and eight *AP3/DEF*-like genes among four taxa in the Phrymaceae (see Supplementary Table S1). Genomic fragments from two copies of *LFY/FLO*-like genes were cloned from *Leucocarpus perfoliatus* using the first intron primer, whereas a single copy was cloned from *Mazus reptans*, *Mimulus ringens*, and *Mimulus kelloggii* using first intron primers and for all four species using second intron primers. Sequence comparison of 3′-RACE cDNAs corresponding to genomic fragments identified all second intron fragments as identical to first intron genomic fragments cloned previously. Thus, we cloned two *LFY/FLO*-like genes from *L. perfoliatus* and a single copy from each of *M. reptans*, *M. ringens*, and *M. kelloggii*. All five *LFY/FLO*-like genes appear functional based on correctly spliced cDNAs lacking missense mutations or premature stop codons. The two copies of *LFY/FLO*-like genes found for *L. pefoliatus* are 37% divergent at silent sites, similar to the average between *FLOA* and *FLOB* co-orthologs found in other families of Lamiales (44% [Aagaard et al. 2005]). Phylogenetic analyses placed the single copies cloned from *M. reptans*, *M. ringens*, and *M. kelloggii* with *FLOA* paralogs (Fig. 1). One copy from *L. perfoliatus* also clustered with *FLOA* paralogs, whereas the second copy clustered with *FLOB* paralogs. As proposed by Aagaard et al. (2005), these analyses are consistent with ancient duplication of *FLO* within the Lamiales after the divergence between Veronicaceae (*A. majus* and *C. glabra*) and the lineage leading to many of the other families of Lamiales including

Verbenaceae, Lamiaceae, Phrymaceae, Paulowniaceae, and Orobanchaceae. We propose a nomenclature for these newly cloned *FLO* homologs as in Aagaard et al. (2005) that reflects the precedence of *Antirrhinum majus FLO* (Coen et al. 1990) with co-orthologs further delineated as *FLOA* and *FLOB* (Fig. 1; Supplementary Table S1).

Our assay for *AP3/DEF* copy number identified two copies from each of the four taxa. Copies were on average 34% divergent at silent sites, similar to the *DEFA* and *DEFB* co-orthologs identified previously among other families of Lamiales (38% [Aagaard et al. 2005]), and all appear functional based on correctly spliced cDNAs lacking missense mutations or premature stop codons. In addition, for *M. kelloggii* two putative alleles having 0.4% and 0.5% of silent sites with substitutions were found for *DEFA* and *DEFB* (see below), respectively. These putative alleles were reproducible over duplicate 3′ RACE experiments. Because of their high degree of similarity, only one from each *DEF* paralog was used in phylogenetic analyses and codon substitution models below. Phylogenetic analyses placed one copy from each species within the *DEFA* clade and the other copy within the *DEFB* clade (Fig. 1). The overall topology of this tree is nearly identical to that of Aagaard et al. (2005). Using a parametric bootstrapping approach, Aagaard et al. (2005) were unable to reject a null hypothesis placing the duplication of *DEF* within Lamiales at the same relative position as for *FLO* (after divergence of the lineage leading to Veronicaceae) consistent with simultaneous duplication of both genes, possibly as a result of an ancient whole genome duplication event within Lamiales. Accordingly, we follow the nomenclature for these newly cloned *DEF* homologs as in Aagaard et al. (2005) that reflects the precedence of *A. majus DEF* (Sommer et al. 1990), with co-orthologs further delineated as *DEFA* and *DEFB* (Fig. 1; Supplementary Table S1).

Increased ω's Following Gene Duplication

Using two different codon substitution models, we found evidence that ω is increased for one of two co-orthologs of *FLO* following gene duplication in the Lamiales. Branch models allowing for two (R2) versus a single (R1) ω for branches in the gene phylogeny (Fig. 2, Table 1) provided a significantly better fit to the data based on the LRT (see Supplementary Table S2). The three-ratio model (R3) did not result in a better fit than R2, though a model estimating separate ω's for *FLOA* and *FLOB* clades (R4) did show a significantly lower log-likelihood score. Based on the four-ratio model, ω for *FLOB* (0.098) increased approximately threefold relative to single-copy lineages (ω = 0.034), or fourfold relative to *FLOA* (ω = 0.025). This is consistent with estimates from fixed-sites models; a fixed-sites model allowing separate ω's for *FLOA* and *FLOB* (Model E) versus a single ω (Model C) resulted in a significantly higher likelihood score (Table 1), estimating *FLOA* ω = 0.021 versus *FLOB* ω = 0.106.

Co-orthologs of *DEF* exhibit a somewhat different pattern than for *FLO*. Both *DEFA* and *DEFB* ω's are increased relative to single copy lineages but do not evolve at different rates relative to each other based on branch or fixed-sites models (Table 1). The two-ratio branch model (R2) resulted in a significantly better fit than the one-ratio model (R1), and adding additional parameters did not result in a significant decrease in log scores similar to results for fixed-sites models (Supplementary Table S2). Parameter estimates based on the two-ratio branch model suggest that co-orthologs of *DEF* (ω = 0.106) are evolving approximately 50% faster than single-copy lineages (ω = 0.073).

Variation in Substitution Rates Among Functional Domains of FLO and DEF

In addition to significant heterogeneity among branches of the gene phylogeny, we found evidence that ω's differ among functional domains within genes. In order to examine how among-site variation corresponds with functional domains of genes, we used two approaches. First, an empirical Bayes approach was used (Nielsen and Yang 1998; Yang et al. 2000) to assign posterior probabilities for each site class in M8 to codons in the *FLO* and *DEF* alignments (Fig. 2). Visual inspection of the distribution of ω's based on this naive empirical Bayes prediction suggests that the DNA binding region of *FLO* (mean ω = 0.05) and the MADS DNA binding region of *DEF* (mean ω = 0.05) are evolving under stronger selective constraint than the corresponding N-terminus (mean ω = 0.06) or I (mean ω = 0.18), K (mean ω = 0.14), and C-terminus (mean ω = 0.11) regions, respectively. This qualitative prediction is confirmed based on LRTs of fixed-sites models. Model E, allowing for different ω's in DNA-binding versus other functional domains, fit both *FLO* and *DEF* data significantly better than Model C as determined by LRTs ($2\Delta \ln L$ = 51.86 and 69.24, respectively; $p < 0.005$). Based on Model E (separate ω's for domains within a gene), the N-terminus of *FLO* is evolving 2.6-fold faster than the DNA binding domain (ω = 0.076 and 0.029, respectively). Similarly, the K, I, and C-terminus regions combined are evolving 4.5-fold faster than the MADS DNA binding domain of *DEF* (ω = 0.147 and 0.033, respectively).

We tested whether rate heterogeneity between a particular functional region was responsible for the observed differences in ω found using branch models. For *FLO* we used LRTs of fixed-sites models com-

Table 1. Parameter estimates (ω's) and likelihood scores for branch-specific models (R1–R4) and fixed-sites models (C and D)

Gene	Model	Number of parameters	Parameters for branches	Log-likelihood
LFY/FLO	One ratio (R1)	39	$\omega_0 = \omega_1 = \omega_2 = \omega_3 = 0.045$	−6967.539
	Two ratios (R2)	40	$\omega_0 = 0.038, \omega_1 = \omega_2 = \omega_3 = 0.051$	−6963.987[b]
	Three ratios (R3)	41	$\omega_0 = 0.033\ \omega_1 = 0.028\ \omega_2 = \omega_3 = 0.052$	−6963.369
	Four ratios (R4)	42	$\omega_0 = 0.034\ \omega_1 = 0.024\ \omega_2 = 0.025\ \omega_3 = 0.098$	−6933.873[b]
	Model C	30	$\omega_2 = \omega_3 = 0.061$	−5254.671
	Model E	32	$\omega_2 = 0.021, \omega_3 = 0.106$	−5224.876[b]
AP3/DEF	One ratio (R1)	45	$\omega_0 = \omega_1 = \omega_2 = \omega_3 = 0.096$	−5720.428[a]
	Two ratios (R2)	46	$\omega_0 = 0.073\ \omega_1 = \omega_2 = \omega_3 = 0.106$	−5717.543
	Three ratios (R3)	47	$\omega_0 = 0.073\ \omega_1 = 0.046\ \omega_2 = \omega_3 = 0.108$	−5716.537
	Four ratios (R4)	48	$\omega_0 = 0.073\ \omega_1 = 0.046\ \omega_2 = 0.109\ \omega_3 = 0.107$	−5716.532
	Model C	36	$\omega_2 = \omega_3 = 0.123$	−5148.544
	Model E	38	$\omega_2 = 0.121, \omega_3 = 0.125$	−5148.438

Note. Subscripts for ω's correspond to branches in Fig. 2 and are distinct from those used in sites and branch-sites models (Table 2). Models having significantly lower log-likelihood scores relative to a simpler nested model (e.g., R3 vs. R2, Model E vs. Model C) are given based on a χ^2 distribution (see Supplementary Table S2).
[a] Significant at $p \leq 0.05$.
[b] Significant at $p \leq 0.005$.

paring DNA binding regions alone between *FLOA* and *FLOB*, or N-terminus regions alone between *FLOA* and *FLOB*. Model E fit the data better than model C for both the DNA binding domain ($2\Delta\ln L = 29.84$, $p < 0.005$) and the N-terminus regions ($2\Delta\ln L = 22.194$, $p < 0.005$), indicating a significant difference between *FLOA* and *FLOB* in both regions (*FLOA* ω = 0.040 and 0.015 versus *FLOB* ω = 0.183 and 0.068 for N-terminus and DNA binding regions, respectively). For *DEF*, we used a branch model with one ratio (R1) versus two ratios (R2) identical to that used previously (Fig. 1, Table 2) but included only the MADS DNA binding domain in the alignment. In this case, adding an additional ω category ($\omega_1 = \omega_2 = \omega_3$) did not result in a significant test statistic ($2\Delta\ln L = 0.20$), indicating that the significant increase we found previously for *DEFA* and *DEFB* relative to single-copy lineages is due to K, I, and C-terminus domains alone.

No Evidence of Positive Selection on FLO and DEF in Lamiales

We found no evidence of positive selection acting within coding regions during the evolution of *FLO* or *DEF* within Lamiales based on either sites or branch-sites models. In most cases, sites models (Table 2) which allow for positive selection at a subset of sites within a molecule while holding ω constant over branches did not result in a significantly better fit than the corresponding nested nonselection model based on LRTs (Table 1; Supplementary Table S3). In the case of *FLO*, although the selection model M8 did show a significantly better fit, ω for the additional site class was estimated at 1.00 and constituted less than 1% of codons. Branch-site models similarly provide no evidence of positive selection acting on

FLO or *DEF*. We specified the two branches immediately following the duplication of *FLO* and *DEF* (ω_1) as foreground branches, and those prior to the duplication (ω_0) as well as daughter branches (ω_2, ω_3) as background branches (Fig. 1). Branch-sites models (Table 2) that allow positive selection only in the foreground branches (Model A) did not result in a better fit than the corresponding nested model (Model A_{null}). We also fitted models treating all branches within the *FLOB* or *DEFA* plus *DEFB* gene clades (where ω's were increased based on branch models) as the foreground branch and, similarly, found no evidence of positive selection using these branch-sites models (data not shown).

Discussion

Pattern of Duplication for FLO and DEF

The *FLO* and *DEF* paralogs we cloned here and previously likely resulted from ancient polyploidization in Lamiales more than 50 mya (Aagaard et al. 2005). Both *FLO* and *DEF* are thought to have duplicated after the split between the lineage leading to Veronicaceae (including the single-copy model species *Antirrhinum majus* [Olmstead et al. 2001]) and that leading to many other Lamiales families including Verbenaceae, Phrymaceae (*sensu* Beardsley and Olmstead 2002), Paulowniaceae, and Orobanchaceae. Aagaard et al. (2005) proposed polyploidization as the source of *FLO* and *DEF* duplicates in Lamiales based on concordant gene phylogenies, proportional levels of silent substitution among *FLO* and *DEF* paralogs as well as other paralagous floral regulatory genes in Lamiales, and linkage relationships (Fishman et al. 2001). Duplication of *DEF* in Lamiales complements a pattern of repeated duplications that have been main-

Table 2. Parameter estimates (ω's and proportions for site classes) and likelihood scores for sites models (M0–M8) and branch-sites models (A and B)

Gene	Model	Number of parameters	Parameter estimates	Log-likelihood
LFY/FLO	Neutral (M1a)	40	$\omega_0 = 0.038$, $p_0 = 0.974$ $\omega_1 = 1$, $p_1 = 0.026$	−6926.818
	Selection (M2a)*	42	$\omega_0 = 0.039$, $p_0 = 0.974$ $\omega_1 = 1$, $p_1 = 0.002$ $\omega_3 = 1.000$, $p_3 = 0.024$	−6926.818
	β (M7)	40	$p = 0.316$, $q = 5.096$	−6856.728
	β & ω (M8)*	42	$p = 0.361$, $q = 6.756$, $p_0 = 0.993$ $\omega = 1.000$, $p_1 = 0.007$	−6852.236*
	Model A*	42	Site class 0 1 2a 2b Proportion 0.980 0.020 0 0 Background ω 0.043 1.000 0.043 1.000 Foreground ω 0.043 1.000 1.000 1.000	−6957.405
	Model A$_{null}$*	41	Site class 0 1 2a 2b Proportion 0.980 0.020 0 0 Background ω 0.043 1.000 0.043 1.000 Foreground ω 0.043 1.000 1.000 1.000	−6957.405
AP3/DEF	Neutral (M1a)	46	$\omega_0 = 0.066$, $p_0 = 0.856$ $\omega_1 = 1$, $p_1 = 0.134$	−5637.999
	Selection (M2a)*	48	$\omega_0 = 0.066$, $p_0 = 0.866$ $\omega_1 = 1$, $p_1 = 0.080$ $\omega_3 = 1.000$, $p_3 = 0.054$	−5637.999
	β (M7)	46	$p = 0.344$, $q = 2.619$	−5569.889
	β & ω (M8)*	48	$p = 0.386$, $q = 3.502$, $p_0 = 0.984$ $\omega = 1.000$, $p_1 = 0.016$	−5568.048
	Model A*	48	Site class 0 1 2a 2b Proportion 0.865 0.135 0 0 Background ω 0.070 1.000 0.070 1.000 Foreground ω 0.070 1.000 1.000 1.000	−5659.618
	Model A$_{null}$*	47	Site class 0 1 2a 2b Proportion 0.865 0.135 0 0 Background ω 0.070 1.000 0.070 1.000 Foreground ω 0.070 1.000 1.000 1.000	−5659.618

Subscripts for ω's are distinct from those used in branch and fixed-sites models (Fig. 2, Table 1). Models having significantly lower log-likelihood scores relative to a simpler nested model (see Supplementary Table S3) are given based on a χ^2 distribution. *Models allowing positive selection.

[a] Significant at $p \leq 0.05$.

[b] Significant at $p \leq 0.005$.

tained under selection throughout the *AP3/DEF* gene lineage. Duplication of *AP3/DEF* homologs has occurred in seed plants (Theissen et al. 2000), within angiosperms at the base of the tricolpates (Kramer et al. 1998), and, more recently, among lineages of basal angiosperms (Stellari et al. 2004) and other tricolpate families such as Ranunculaceae (Kramer et al. 2003). In sharp contrast, duplicates of *FLO* in Lamiales are the first angiosperm paralogs reported outside of tetraploid taxa (e.g., *Zea mays* [Bomblies et al. 2003]) or recent duplications within *Brassicaceae* (Baum et al. 2005), despite surveys for copy number across angiosperms (e.g., Frohlich and Parker 2000). Thus, *FLO* and *DEF* paralogs in Lamiales represent an as yet novel example of the joint retention of duplicates of these floral regulatory genes.

Mechanisms of Duplicate Gene Preservation

Coding Regions of FLO *and* DEF *Show Evidence of Subfunctionalization.* In order to explore hypotheses for the preservation of these duplicate interacting floral regulatory genes, including the apparently novel ancient *FLO* paralogs we identified, we used several different models of codon evolution to test how selection acted within coding regions following

gene duplication in Lamiales. Two distinct classes of codon models are consistent in showing that positive selection within coding regions has not played a role in the maintenance of *FLO* or *DEF* paralogs. First, sites models indicate that ω is significantly less than one (Fig. 3, Table 2), consistent with strong purifying selection acting on duplicate copies. Second, a signature of positive selection is absent from branch-sites models, which have increased power to detect adaptive evolution acting during discrete intervals of a gene's evolutionary history (Bielawski and Yang 2003). For the modified branch-sites Model A (Table 2), we specified the two branches immediately following the duplication of *FLO* or *DEF* (ω_1) as foreground branches, and those prior to the duplication (ω_0) as well as daughter branches (ω_2, ω_3) as background branches (Fig. 2). This reflects a model whereby duplicated genes may initially experience positive selection, followed by the return of purifying selection along later branches in the gene phylogeny (Bielawski and Yang 2003) which could result from (i) gene duplicates acquiring new functions distinct from the ancestral single-copy gene (neofunctionalization *sensu* Force et al. [1999]) or (ii) adaptive evolution acting on subfunctionalized protein coding domains constrained to suboptimal performance in
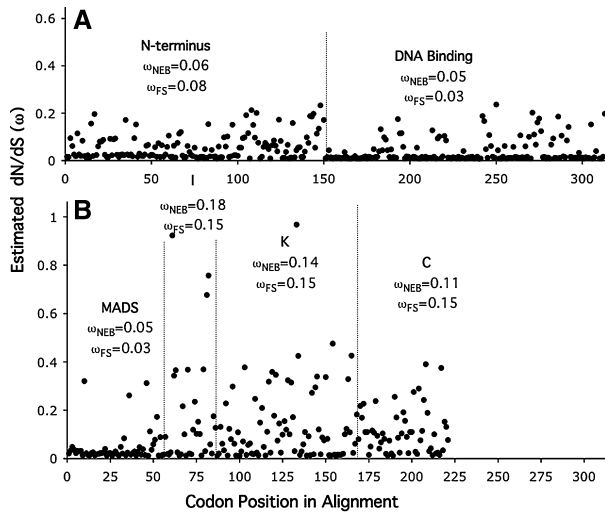
Fig. 3. Distributions of estimated ω values relative to the positions of codons and functional domains for (**A**) *FLO* and (**B**) *DEF*. Values for codon positions of the alignments used in all codon substitution models are calculated using an empirical Bayes approach (Yang 1998; Yang and Nielsen 2000) as the mean over site classes (M8) weighted by the posterior probabilities for sites. The mean ω over codons within functional domains using this naive empirical Bayes approach (ω_{NEB}) as well as the ω estimated for functional domains from fixed-sites models (ω_{FS}) are shown.

single-copy genes carrying out multiple functions (Hughes 1999; and see below). The absence of any evidence for positive selection in these branch-sites models or similar models specifying other portions of the gene phylogeny as foreground branches (see Results) again strongly supports continuous purifying selection within coding regions during the evolution of *FLO* and *DEF* paralogs in Lamiales.

Although we found no evidence of positive selection acting among paralogs, branch models of codon evolution show that purifying selection is relaxed within coding regions of one (*FLOB*) or both (*DEFA/B*) paralogs following duplication in Lamiales (Fig. 2, Table 1). Relaxed purifying selection is a common pattern following gene duplication, and often differentially affects one paralogue (e.g., Conant and Wagner 2002). For regulatory genes, relaxed purifying selection within functional domains of coding regions has been proposed to result from paralogous proteins binding a subset of interaction partners relative to the single-copy ancestral protein (e.g., Dermitzakis and Clark 2002). Alternately, subfunctionalization might involve regulatory domains of these genes similar to the general mechanism proposed by Force et al. (1999). In this case, the increase in ω we see for protein coding regions following duplication would be an indirect result, possibly reflecting fewer binding partners within a restricted expression domain. In the case of *FLOA*, fixed-sites models show that selection is relaxed across the gene including the DNA binding domain (Fig. 3A; also see Results). In contrast, for *DEFA/B* purifying selection

remains strong within the MADS DNA binding domain but is relaxed within the I, K, and C domains important in protein interaction and activation (Fig. 3B; also see Results). Taken together, results from branch and fixed-sites models are consistent with a subfunctionalization hypothesis for the maintenance of *FLO* and *DEF* paralogs in Lamiales that could involve distinctly different functional domains between the two genes.

Our results are consistent with the theory (Walsh 1995; Force et al. 1999) that subfunctionalization rather than adaptive evolution contributes more strongly to duplicate gene preservation. However, the relative contribution of these evolutionary mechanisms to maintenance of paralogs remains controversial. Both functional evidence and sequence divergence data for several other paralogous floral regulatory are consistent with subfunctionalization of paralogs. Functional evidence from *Maize* (*ZAG1* and *ZMM2* [Mena et al. 1998]) and *Gerbera* (*GRCD1* and *GRCD2* [Uimari et al. 2004]) show that recent duplicates of other MADS box genes that regulate floral development have partitioned expression domains in floral buds typical of subfunctionalization. Similarly, non-MADS box genes involved in establishing floral symmetry thought to be duplicated within the Lamiales lineage leading to *Antirhinum* (Veronicaceae) show (i) differences in patterns of gene expression suggestive of subfunctionalization and (ii) evidence of relaxed purifying selection within functional domains for one of two paralogs but no evidence of positive selection among paralogs (*CYC* and *DICH* [Hileman and Baum 2003]). Ancient paralogs of MADS box genes duplicated near the base of the tricolpate lineage which regulate male and female floral organ development also show evidence of partitioned expression domains and functional roles during floral development (Causier et al. 2005). Interestingly, in this case functional evidence from representative Rosid (*A. thaliana*; *AG* and *SHP1/2*) and Asterid (*A. majus*; *PLE* and *FAR*) model taxa shows that the regulatory role of individual paralogs has resolved differently within independently evolving lineages, lending further support to a subfunctionalization hypothesis for preservation of ancient MADS box paralogs. However, signatures of adaptive evolution are increasingly reported among coding regions of genes following duplication as statistical models improve (e.g., Bielawski and Yang 2003) and have been detected previously within noncoding regulatory sequences of several other MADS box genes derived from the most recent polyploidization event in *A. thaliana* (Moore and Purugganan 2005). In addition, subfunctionalization and adaptive evolution are not mutually exclusive hypotheses. A signature of adaptive evolution can result secondarily after constraints are relaxed following subfunction-

alization (Hughes 1999), and under this scenario adaptive mutation (a low probability event) is not prerequisite for the action of positive selection following duplication. In sum, future studies should investigate partitioning of expression domains or functional interactions as well as divergence within noncoding regulatory regions of *FLO* and *DEF* paralogs in Lamiales (see below) for further evidence supporting subfunctionalization.

Dosage is Unlikely to Maintain FLO *and* DEF *Paralogs in Lamiales.* Although dosage effects are seen among transcription factors (Birchler et al. 2001) and are thought to be particularly prominent following polyploidization (Blanc and Wolfe 2004b; Maere et al. 2005), dosage is unlikely to play a role in the maintenance of *FLO* or *DEF* paralogs in Lamiales. Dosage can contribute to the preservation of duplicate regulatory genes in two distinct ways. First, Moore and Purugganan (2005) suggest that a signature of adaptive evolution for fully redundant paralogs of MADS box genes in *A. thaliana* is consistent with selective maintenance of functional redundancy to reduce the developmental error rate. Concepts of functional redundancy contributing to developmental robustness have been explored extensively elsewhere (e.g., Cooke et al. 1997). Second, Blanc and Wolfe (2004b) and Maere et al. (2005) show duplicate genes that constitute subunits of protein complexes are retained at a higher rate than other functional classes of genes, possibly due to stoichiometric requirements or dominant negative phenotypes following loss of one copy (Veitia 2003). *AP3/DEF* constitutes one subunit of a multidomain protein (Pelaz et al. 2000) and thus might be disproportionately affected by such dosage effects. However, because these hypotheses predict retention of paralogs in an undifferentiated state and we see divergence between functional domains of *DEFA/B*, it is unlikely that they have contributed significantly to their preservation in Lamiales. Similarly, although dosage effects are seen for transcription factors (Birchler et al. 2001), dosage seems an unlikely explanation for *LFY/FLO*, which can regulate floral induction at very low levels (Coen et al. 1990; Weigel et al. 1992), because of partial redundancy with another floral regulatory gene (*SQUA/AP1* [Zik and Irish 2003]) and because of divergence within the DNA binding domain.

Could Transcriptional Regulatory Interactions Contribute to Duplicate Gene Preservation in Lamiales? Paralogs arising from polyploidization experience selection along with other duplicate loci across the genome, suggesting that interactions among loci might contribute to their maintenance. Because paralogs in Lamiales appear to be the result of ancient polyploidization, an intriguing possibility for our novel finding of duplicate copies of *FLO* involves the interactions among genes in the floral regulatory pathway. *LFY/FLO* is one of the earliest expressed genes in the developing floral meristem, regulating the expression of many downstream genes including floral homeotic genes of the MADS box family of transcription factors such as *AP3/DEF* (reviewed in Zik and Irish 2003) as well as several genes having uncharacterized functions during floral development (William et al. 2004). *LFY/FLO* is known to positively regulate the expression of *AP3/DEF* in both *A. majus* (Ingram et al. 1997) and *A. thaliana* (Weigel and Meyerowitz 1993). In the case of *A. thaliana*, *LFY* activation of *AP3* is known to involve transcriptional activation via direct binding to the *AP3* promoter (Lamb et al. 2002) and transcriptional activation requires a complex of proteins (Zhao et al. 2001), some of which are known to play a similar role in *A. majus* (Ingram et al. 1997).

The cognate relationships among genes in the floral regulatory pathway could contribute to the joint retention of paralogs of genes such as *FLO* or *DEF* via a number of mechanisms. One possibility is simple dosage effects, which are thought to be prominent among regulatory genes involved in signal transduction and transcription (Birchler et al. 2001). However, as described above we believe this is an unlikely explanation in our case. Alternatively, Blanc and Wolfe (2004b) proposed divergence in the timing or expression domains among pairs of paralogs could have a broad impact on duplicate gene preservation across a polyploid genome, a process they termed concerted divergence of gene expression. This is essentially a multilocus extension of subfunctionalization models. For example, sequence divergence between regulatory elements of subfunctionalized *DEF* paralogs driving expression in different floral tissues or at different times during development might require *FLO* paralogs with divergent DNA binding domains. This is an intriguing possibility given the pattern of elevated nonsynonymous substitutions within the DNA binding domain of *FLOB* and because one of two *cis*-regulatory regions of *AP3* controlling early expression (Hill et al. 1998) corresponds to the *LFY* binding site (Lamb et al. 2002). Moreover, this hypothesis is testable, as it predicts increased regulatory activity between pairs of duplicate genes.

With the notable exception of more recent *LFY/FLO* duplicates from tetraploids (Bomblies et al. 2003; Kelly 1995; Ahearn et al. 2001) or among Brassicaceae taxa (Baum et al. 2005), to our knowledge no other paralogs as divergent as *FLOA/B* in Lamiales have been described. The conspicuous lack of diversification of the *LFY/FLO* gene lineage broadly among angiosperms via gene duplication has led to the suggestion that evolutionary processes may disfavor the retention of *LFY/FLO* paralogs. Al-

though negative selection on genomes harboring *LFY/FLO* paralogs is a possibility (Cronk 2001), Baum et al. (2005) argue that duplicates may simply be lost by drift more frequently than other floral regulatory genes due to differences in the transcriptional role or regulatory organization of *LFY/FLO*, leading to a lower probability of subfunctionalization. Whether the divergence of regulatory interactions among paralogs such as *DEFA/B* might explain our finding of novel *FLO* duplicates in Lamiales is speculative at this point, and will require future studies examining both the functional interactions among pairs of paralogs and divergence within noncoding regulatory regions. Because *LFY/FLO* duplicates are thought to be less susceptible to preservation via standard subfunctionalization mechanisms (Baum et al. 2005), such work may prove particularly valuable for understanding the importance of genetic interactions in the maintenance of duplicate genes following polyploidization

References

Aagaard J, Phillips P (2005) Accuracy and power of the likelihood ratio test for comparing evolutionary rates among genes. J Mol Evol 60:426–433

Aagaard J, Willis J, Phillips P (2005) Duplication of floral regulatory genes in the Lamiales. Am J Bot 92:1284–1293

Baum D, Yoon H, Oldham R (2005) Molecular evolution of the transcription factor *LEAFY* in Brassicaceae. Mol Phylogenet Evol 37:1–14

Beardsley P, Olmstead R (2002) Redefining Phrymaceae: the placement of *Mimulus*, tribe Mimuleae, and *Phryma*. Am J Bot 89:1093–1102

Bielawski J, Yang Z (2003) Maximum likelihood methods for detecting adaptive evolution after gene duplication. JSFG 3:201–212

Blanc G, Wolfe K (2004a) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

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

Blanc G, Hokamp K, Wolfe K (2003) A recent polyploidization superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res 13:137–144

Bomblies K, Wang R, Ambrose B, Schmidt R, Meeley R, Doebley J (2003) Duplicate FLORICAULA/LEAFY homologs *ZFL1* and *zfl2* control inflorescence architecture and flower patterning in maize. Development 130:2385–2395

Causier B, Castillo R, Zhou J, Ingram R, Xue Y, Schwarz-Sommer Z, Davies B (2005) Evolution in action: following function in duplicated floral homeotic genes. Curr Biol 15:1508–1512

Coen E, Romero J, Doyle S, Elliott R, Murphy G, Carpenter R (1990) Floricaula: a homeotic gene required for flower development in *Antirrhinum majus*. Cell 63:1311–1322

Conant G, Wagner A (2002) Asymmetric sequence divergence of duplicate genes. Genome Res 13:2052–2058

Cook J, Nowak M, Boerlijst M, Maynard-Smith J (1997) Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet 13:360–364

Cronk Q (2001) Plant evolution and development in a post-genomic context. Nat Rev Genet 2:607–619

Dermitzakis E, Clark A (2001) Differential selection after duplication in mammalian developmental genes. Mol Biol Evol 18:557–562

Fishman L, Kelly A, Morgan E, Willis J (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. Genetics 159:1701–1716

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

Frohlich M, Parker D (2000) The mostly male theory of flower evolutionary origins: from genes to fossils. Syst Bot 25:155–170

Gocal G, King R, Blundell C, Schwartz W, Andersen C, Weigel D (2001) Evolution of floral meristem identity genes: analysis of *Lolium temulentum* genes related to APETALA1 and LEAFY of *Arabidopsis*. Plant Physiol 125:1788–1801

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Hileman L, Baum D (2003) Why do paralogs persist? Molecular evolution of *CYCLOIDEA* and related floral symmetry genes in Antirrhineae (Veronicaceae). Mol Biol Evol 20:591–600

Hill T, Day C, Zondio S, Thackeray A, Irish V (1998) Discrete spatial and temporal *cis*-acting elements regulate transcription of the *Arabidopsis* floral homeotic gene APETALA3. Development 125:1711–1721

Hughes A (1999) Adaptive evolution of genes and genomes. Oxford University Press, Oxford

Ingram G, Doyle S, Carpenter R, Schultz E, Simon R, Coen E (1997) Dual role for *fimbriata* in regulating floral homeotic genes and cell division in *Antirrhinum*. EMBO J 16:6521–6534

Jack T, Brockman L, Meyerowitz E (1992) The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. Cell 68:683–697

Kelly A, Willis J (1998) Polymorphic microsatellite loci in *Mimulus guttatus* and related species. Mol Ecol 1998:769–774

Kramer E, Dorit R, Irish V (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. Genetics 149:765–783

Kramer E, Di Stilio V, Schluter P (2003) Complex patterns of gene duplication in the *AP3* and PI lineages of the Ranunculaceae. Int J Plant Sci 164:1–11

Kush A, Brunelle A, Shevell D, Chua N (1993) The cDNA sequence of two MADS box proteins in Petunia. Plant Physiol 102:1051–1052

Lamb R, Hill T, Tan Q, Irish V (2002) Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. Development 129:2079–2086

Levin D (1983) Polyploidy and novelty in flowering plants. Am Nat 122:1–25

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

Lynch M, Conery J (2001) Gene duplication and evolution. Science 293:1551

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

Mena M, Ambrose B, Meeley R, Briggs S, Yanofsky M, Schmidt R (1996) Diversification of C-function activity in maize flower development. Science 274:1537–1540

Molinero-Rosales N, Jamilena M, Zurita S, Gomez P, Capel J, Lozano R (1999) FALSIFLORA, the tomato orthologue of FLORICAULA and LEAFY, controls flowering time and floral meristem identity. Plant J 20:685–693

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

Moore R, Purugganan M (2005) Molecular population genetics of redundant floral-regulatory genes in Arabidopsis thaliana. Mol Biol Evol 22:91–103

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Olmstead R, dePamphilis C, Wolfe A, Young N, Elisons W, Reeves P (2001) Disintegration of the Scrophulariaceae. Am J Bot 88:348–361

Otto S, Whitton J (2000) Polyploid incidence and evolution. Annu Rev Genet 34:401–437

Pelaz S, Ditta G, Baumann E, Wisman E, Yanofsky M (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature 405:200–203

Reichman J, Meyerowitz E (1997) MADS domain proteins in plant development. J Biol Chem 378:1079–1101

Sommer H, Beltran J, Huijser P, Pape H, Lonnig W, Saedler H, Schwarz-Sommer Z (1990) Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. EMBO J 9:605–613

Souer E, van der Krol A, Kloos D, Spelt C, Bliek M, Mol J, Koes R (1998) Genetic control of branching pattern and floral identity during Petunia inflorescence development. Development 125:733–742

Stebbins G (1971) Processes of organic evolution in higher plants. Edward Arnold, London

Stellari G, Jaramillo M, Kramer E (2004) Evolution of the APETALA3 and PISTILLATA lineages of MADS-box-containing genes in the basal angiosperms. Mol Biol Evol 21:506–519

Swofford D (2002) PAUP*: phylogenetic analysis using parsimony (and other methods), 4.0 Beta for Macintosh. Sinauer Associates, Sunderland, MA

Theissen G, Becker A, Di Rosa A, Kanno A, Kim J, Munster T, Winter K, Saedler H (2000) A short history of MADS-box genes in plants. Plant Mol Biol 42:115–149

Uimari A, Kotilainen M, Elomaa P, Yu D, Albert V, Teeri T (2004) Integration of reproductive meristem fates by a SEPALLATA-like MADS-box gene. Proc Natl Acad Sci USA 101:15817–15822

Veita R (2003) Nonlinear effects in macromolecular assembly and dosage sensitivity. J Theor Biol 220:19–25

Walsh J (1995) How often do duplicated genes evolve new functions? Genetics 139:421–428

Weigel D, Meyerowitz E (1993) Activation of floral homeotic genes in Arabidopsis. Science 261:1723–1726

Weigel D, Alvarez J, Smyth D, Yanofsky M, Meyerowitz E (1992) Leafy controls floral meristem identity in Arabidopsis. Cell 69:843–859

Wendel J (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Wikstrom N, Savolainen V, Chase M (2001) Evolution of the angiosperms: calibrating the family tree. Proc R Soc Lond B 268:2211–2220

William D, Su Y, Smith M, Lu M, Baldwin D, Wagner D (2004) Genomic identification of direct ta rget genes of LFY. Proc Natl Acad Sci USA 101:1775–1780

Yang Z (1997) PAML: a program for package for phylogenetic analysis by maximum likelihood. CABIOS 15:555–556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yang Z (2001) Adaptive molecular evolution. In: Balding D, Bishop M, Cannings C (eds) Handbook of statistical genetics. Wiley, New York, pp 327–350

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917

Yang Z, Swanson W (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol 19:49–57

Yang Z, Nielsen R, Goldman N, Petersen A (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479

Zik M, Irish V (2003) Flower development: initiation, differentiation, and diversification. Annu Rev Cell Dev Biol 19:19–40

Table S1. Homologs *of LFY/FLO and AP3/DEF* cloned previously or in this study. Gene names conform to the cited reference for previously cloned homologs, or follow the nomenclature *for Antirrhinum mqjus* if cloned in this study. For species with more than one homolog, an A or B designation describes the paralogy group within which genes cluster in phylogenetic analyses (Aagaard et al. 2005). Species with multiple A or B copies are given a further numeric designation. An abbreviation incorporating both species and gene names is used to label terminal branches in gene trees.

Species	Gene	Abbreviation	Reference	Accession
Petunia hybrida	*ALF*	PetHyALF	Souer et al. 1998	AF030171
	PMADS1	PetHyPMADSl	Kush et al. 1993	X69946
Solatium esculentum	*FA*	LycEsFA	Molinero-Rosales et al. 1999	AF 197935
	AP3	LycEsAPS	Kramer et al. 1998	AF052868
Syringa vulgaris	*FLO*	SyrVuFLO	Aagaard et al. 2005	AY524037
	DEF	SyrVuDEF	Aagaard et al. 2005	AY524007
Antirrhinum majus	*FLO*	AntMaFLO	Coen et al. 1990	M55525
	DEF	AntMaDEF	Sommer et al. 1990	S12378
Chelone glabra	*FLO*	CheGlFLO	Aagaard et al. 2005	AY524029
	DEF	CheGlDEF	Aagaard et al. 2005	AY524008
Verbena officinalis	*FLOA*	VerOfFLOA	Aagaard et al. 2005	AY524030
	FLOB	VerOfFLOB	Aagaard et al. 2005	AY524038
	DEFA	VerOfDEFA	Aagaard et al. 2005	AY524009
	DEFB	VerOfDEFB	Aagaard etal. 2005	AY524014
Pedicularis groenlandica	*FLOA*	PedGrFLOA	Aagaard et al. 2005	AY524033
	FLOB	PedGrFLOB	Aagaard et al. 2005	AY524039
	DEFA	PedGrDEFA	Aagaard et al. 2005	AY524010
	DEFB	PedGrDEFB	Aagaard et al. 2005	AY524017
Paulownia tomentosa	*FLOA*	PauToFLOA	Aagaard et al. 2005	AY524034
	FLOB	PauToFLOB	Aagaard et al. 2005	AY524040
	DEFA	PauToDEFA	Aagaard etal. 2005	AY524011
	DEFB	PauToDEFB	Aagaard et al. 2005	AY524018
Mazus reptans	*FLOA*	MazReFLOA	*this study*	AY530546
	DEFA	MazReDEFA	*this study*	AY530538
	DEFB	MazReDEFB	*this study*	AY530542
Mimulus ringens	*FLOA*	MimRiFLOA	*this study*	AY530547
	DEFA	MimRiDEFA	*this study*	AY530539
	DEFB	MimRiDEFB	*this study*	AY530543
Leucocarpus perfoliatus	*FLOA*	LeuPerFLOA	*this study*	AY530548
	FLOB	LeuPerFLOB	*this study*	AY530550
	DEFA	LeuPerDEFA	*this study*	AY530540
	DEFB	LeuPerDEFB	*this study*	AY530544
Mimulus kelloggii	*FLOA*	MimKeFLOA	*this study*	AY530549
	DEFA	MimKeDEFA	*this study*	AY530541
	DEFB	MimKeDEFB	*this study*	AY530545
Mimulus lewisii	*FLOA*	MimLeFLOA	Aagaard et al. 2005	AY524035
	FLOB	MimLeFLOB	Aagaard etal. 2005	AY524041
	DEFA	MimLeDEFA	Aagaard et al. 2005	AY524013
	DEFB	MimLeDEFB	Aagaard et al. 2005	AY524019
Mimulus guttatus	*FLOA*	MimGuFLOA	Aagaard et al. 2005	AY524036
	FLOB	MimGuFLOB	Aagaard et al. 2005	AY524042
	DEFA	MimGuDEFA	Aagaard et al. 2005	AY524012
	DEFB	MimGuDEFB	Aagaard et al. 2005	AY524020

Table S2. Likelihood ratio test (LRT) statistics comparing branch-specific models (R1-R4) and fixed-sites models (C and E). The test statistic (2ΔInL) is calculated as twice the difference between log-likelihood values for models, and compared with a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between models (Table 1).

Gene	Models Compared	2ΔInL
LFY/FLO	One-ratio (Rl) vs. two-ratio (R2)	7.104**
	Two-ratio (R2) vs. three-ratio (R3)	1.236
	Three-ratio (R3) vs. four-ratio (R4)	58.992**
	Model C vs. Model E	59.590**
AP3/DEF	One-ratio (Rl) vs. two-ratio (R2)	5.770*
	Two-ratio (R2) vs. three-ratio (R3)	2.012
	Three-ratio (R3) vs. four-ratio (R4)	0.010
	Model C vs. Model E	0.212

*Significant at $p = 0.05$. **Significant at $p \leq 0.005$.

Table S3. Likelihood ratio test (LRT) statistics comparing sites models (Mla-M8) and branch-sites models (A and B). The test statistic (2ΔInL) is calculated as twice the difference between log-likelihood values for models, and compared with a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between models (Table 2).

Gene	Models Compared	2Δln$L$
LEY/FLO	Neutral (Ml a) vs. selection (M2a)	0.000
	Beta (M7) vs. Beta & ω (M8)	8.984*
	ModelA vs. ModelA$_{null}$	0.000
AP3/DEF	Neutral (Ml a) vs. selection (M2a)	0.000
	Beta (M7) vs. Beta & ro (M8)	3.682
	ModelA vs. ModelA$_{null}$	0.000

*Significant at $p \leq 0.05$. **Significant at $p \leq 0.005$.