Patterns of Molecular Evolution Among Paralogous Floral Homeotic Genes

Amy L. Lawton-Rauh, Edward S. Buckler IV, and Michael D. Purugganan
Department of Genetics, North Carolina State University

The plant MADS-box regulatory gene family includes several loci that control different aspects of inflorescence and floral development. Orthologs to the *Arabidopsis thaliana* MADS-box floral meristem genes *APETALA1* and *CAULIFLOWER* and the floral organ identity genes *APETALA3* and *PISTILLATA* were isolated from the congeneric species *Arabidopsis lyrata.* Analysis of these loci between these two *Arabidopsis* species, as well as three other more distantly related taxa, reveal contrasting dynamics of molecular evolution between these paralogous floral regulatory genes. Among the four loci, the *CAL* locus evolves at a significantly faster rate, which may be associated with the evolution of genetic redundancy between *CAL* and *AP1.* Moreover, there are significant differences in the distribution of replacement and synonymous substitutions between the functional gene domains of different floral homeotic loci. These results indicate that divergence in developmental function among paralogous members of regulatory gene families is accompanied by changes in rate and pattern of sequence evolution among loci.

Introduction

Developmental processes are largely controlled by regulatory loci that modulate gene expression patterns. Molecular genetic studies have identified a number of loci that regulate developmental processes, and many of these genes have been shown to be members of regulatory gene families. Developmental genetic investigations have demonstrated that these regulatory gene families evolve primarily by gene duplication and divergence (Scott and Weiner 1984; Purugganan 1998), leading to distinct paralogous loci regulating different aspects of organismal morphogenesis. The evolutionary divergence of gene function among regulatory gene family members provides a mechanism for the elaboration of developmental genetic networks and the increasing complexity of morphological structures. There has thus been continued interest in studying the evolutionary histories and dynamics of developmental regulatory genes, as well as in increasing efforts to investigate the evolution of homeodomain (Zhang and Nei 1996; Bailey et al. 1997), bHLH (Atchley and Fitch 1995), myb-class (Rosinski and Atchley 1998), and other regulatory gene families. It remains unclear, however, to what extent functional diversification among regulatory loci is mirrored by differences in the extent and patterning of sequence evolution between homologous developmental genes.

The plant MADS-box regulatory gene family comprises a group of developmental regulatory loci that encode sequence-specific DNA-binding transcriptional activators (Riechmann and Meyerowitz 1997). Plant MADS-box proteins are about 240–260 amino acids (aa) in length and are characterized by the presence of a highly conserved 57-aa MADS-box (see fig. 1). This MADS-box protein region is widely distributed among eukaryotic genomes within humans (*SRF*), *Drosophila* (*MEF2C*), and yeast (*MCM1*) transcriptional activators (Pollock and Treisman 1991). Plant MADS-box proteins also possess a moderately conserved 70-aa domain called the K-box, which may form coiled-coil structures that participate in protein–protein dimerization interactions (Ma, Yanofsky, and Meyerowitz 1991; Riechmann and Meyerowitz 1997). Plant MADS-box proteins also include the I- and C-regions that are poorly conserved at the sequence level (Purugganan et al. 1995).

Molecular studies indicate that homo- or heterodimerization of MADS-box proteins is necessary for sequence-specific DNA-binding activity of these transcriptional activators (Riechmann and Meyerowitz 1997). Domain analyses of *Arabidopsis* MADS-box proteins indicate that the MADS-box/I-region/K-box sequence (the MIK region) constitutes the functional core domain necessary for both dimerization and DNA binding (see fig. 1). Previous molecular studies have demonstrated that at least one *Arabidopsis* MADS-box transcriptional activator binds its DNA target as a homodimer (Krizek and Meyerowitz 1996; Riechmann, Krizek, and Meyerowitz 1996; Riechmann, Wang, and Meyerowitz 1996). Moreover, biochemical studies indicate that the *Arabidopsis* APETALA3 and PISTILLATA proteins bind to target promoter sequences as a heterodimer (Riechmann and Meyerowitz 1997). The dimerization domains differ between various MADS-box proteins; in vitro experiments suggest that the MADS-box and the I-region are both necessary for dimerization and that the K-box serves to stabilize protein–protein interactions (Riechmann and Meyerowitz 1997). Domain deletion studies of the related MADS-box transcriptional activator AGAMOUS indicates that a protein containing only the MADS-box, the I-region, and the K-box (the MIK region) can form efficient dimers (Mizukami et al. 1996).

Mutations at several *Arabidopsis* MADS-box genes result in floral phenotypes characterized by alterations in floral organ development (Bowman, Smyth, and Meyerowitz 1991). These loci are referred to as floral homeotic genes, and genetic and molecular studies indicate that these loci fall into two broad classes: floral meristem identity genes, specifying inflorescence and floral meristem identity, and floral organ identity genes, defining identities of organs at specific locations in the developing flower (Yanofsky 1995) (see fig. 2*A*). *APETALA1* and *CAULIFLOWER* are two MADS-box mer-

Key words: *Arabidopsis thaliana, Arabidopsis lyrata,* MADS-box, *APETALA1, APETALA3, CAULIFLOWER, PISTILLATA.*

Address for correspondence and reprints: Michael D. Purugganan, Department of Genetics, Box 7614, North Carolina State University, Raleigh, North Carolina 27695. E-mail: michaelp@unity.ncsu.edu.

Mol. Biol. Evol. 16(8):1037–1045. 1999
© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

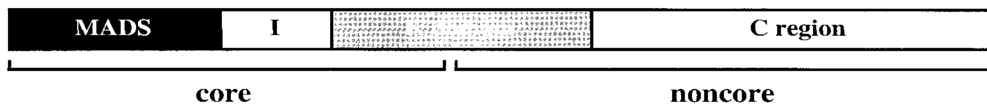

FIG. 1.—Structure of plant MADS-box proteins. The functional core and noncore domains are indicated in brackets.

istem identity loci. *AP1* and *CAL* perform partially redundant developmental functions, specifying the identity of the floral meristems flanking inflorescence shoots (Mandel et al. 1992; Bowman et al. 1993; Kempin, Savidge, and Yanofsky 1995). *AP1* appears to possess additional organ identity functions involving sepal and petal differentiation. Phylogenetic studies indicate that *AP1* and *CAL* arose from a recent gene duplication event, as the duplicate *CAL* locus is found only within the Brassicaceae family (see fig. 2*B*) (Purugganan 1997; Lowman and Purugganan 1999).

The *Arabidopsis* loci *APETALA3* and *PISTILLATA* are floral organ identity genes required for petal and stamen development (see fig. 2*A*) (Jack, Brockmann, and Meyerowitz 1992; Goto and Meyerowitz 1994). Like *AP1* and *CAL, AP3* and *PI* are closely related to one another and arose via a gene duplication event. However, the *AP3-PI* duplication event occurred before the major angiosperm diversification, considerably earlier than the *AP1-CAL* duplication (see fig. 2*B*) (Purugganan 1997; Kramer, Dorit, and Irish 1998). Additionally, phylogenetic studies indicate that the last common ancestor of the AP3 and PI group of genes and the AP1/CAL gene subgroup predates the angiosperm/gymnosperm separation 285 MYA (Purugganan 1997).

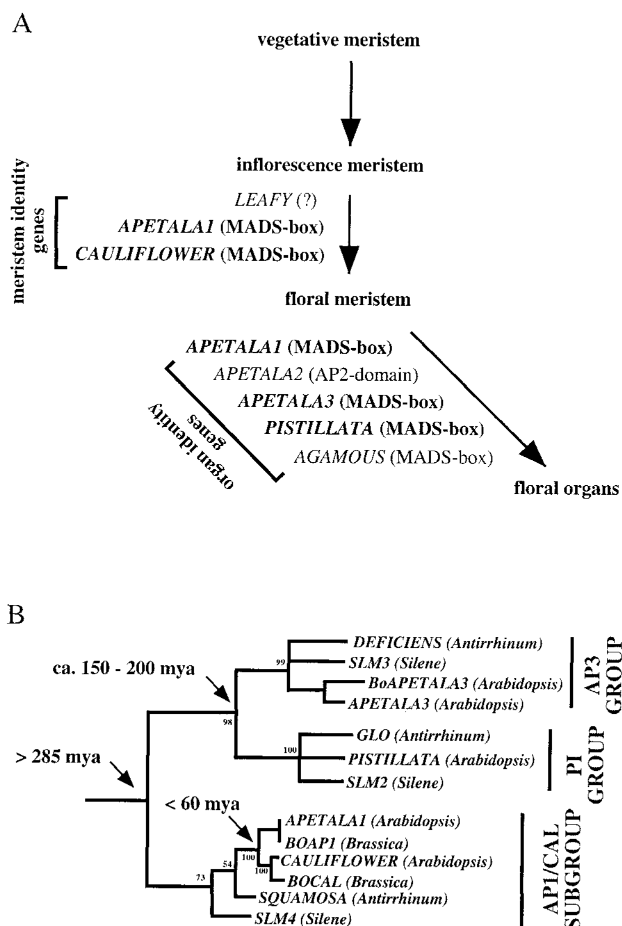

FIG. 2.—*A,* Schematic diagram of the *Arabidopsis thaliana* flower developmental pathway. The genes utilized in this study are indicated in boldface. *B,* Portions of the plant MADS-box gene family phylogeny, showing the relationships of the genes used in this study. Molecular-clock estimates of divergence times are indicated (Purugganan 1997).

Orthologs to the *Arabidopsis* floral homeotic genes *AP3, PI,* and *AP1/CAL* have been identified in several other angiosperm species (see table 1), including *Antirrhinum majus* (Scrophulariaceae) (Coen 1991; Saedler and Huijser 1993) and *Silene latifolia* (Caryophyllaceae) (Hardenack et al. 1994). In *Antirrhinum,* the *DEFICIENS, GLOBOSA,* and *SQUAMOSA* genes are orthologs to the *Arabidopsis AP3, PI,* and *AP1* loci, respectively (Sommer et al. 1990; Huijser et al. 1992; Trobner et al. 1992). Experiments with *Antirrhinum* indicate that *DEF* and *GLO* are also involved in petal and stamen differentiation in snapdragon flowers, while mutations in *SQUA* result in the formation of bract-forming shoots in place of flowers. In *Silene,* a number of MADS-box genes (*SLM2–SLM4*) have been shown to be orthologs to different *Arabidopsis* floral homeotic loci (Hardenack et al. 1994; Purugganan 1997) (see table 1). Genetic studies among distantly related flowering plant species with distinct floral morphologies suggest evolutionary conservation of basic developmental function between these orthologous floral homeotic loci (Coen 1991).

An analysis of the molecular evolution of the MADS-box floral homeotic genes constituting the flower developmental pathway may provide insights into evolutionary patterns that accompany diversification of genes within a regulatory network. This analysis would benefit from the significant amount of structural and functional information on these floral homeotic loci that has been obtained with both genetic and molecular approaches (Riechmann and Meyerowitz 1997). Moreover, the use of paralogous members of a single regulatory gene family provides a common structural and functional context to compare evolutionary patterns across developmentally distinct loci. In this paper, we describe patterns of sequence evolution among the genes *APETALA3, PISTILLATA, APETALA1,* and *CAULIFLOWER.* By utilizing genes from species that have diverged at various times, one can analyze changes in the rates of molecular evolution across different temporal scales. Our analysis includes comparisons of loci within the Brassicaceae (5–40 Myr divergence times) and among genes found between Brassicaceae, Scrophulariaceae, and Caryophyllaceae species (>60 Myr divergence) (Crane, Friis, and Pedersen 1995).

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

Table 1
Floral Homeotic Genes Used in this Study

Arabidopsis thaliana	*Arabidopsis lyrata*	*Brassica oleracea*	*Antirrhinum majus*	*Silene latifolia*
APETALA1	*AlAPETALA1*	*BoAPETALA1*	*SQUAMOSA*	*SLM4*
CAULIFLOWER	*AlCAULIFLOWER*	*BoCAULIFLOWER*	*SQUAMOSA*[a]	*SLM4*[a]
APETALA3	*AlAPETALA3*	*BoAPETALA3*	*DEFICIENS*	*SLM3*
PISTILLATA	*AlPISTILLATA*	NA	*GLOBOSA*	*SLM2*

[a] Since the *AP1* and *CAL* genes are duplicated within the Brasicaceae, these loci are orthologous to both Brassicaceae *AP1* and *CAL*.

We find that despite similarities in overall structure and basic functions of these different floral homeotic MADS-box genes, there appear to be significant differences in the rates of nucleotide substitution between functionally distinct loci. Genes that control differing aspects of floral development also show contrasting distributions in sequence changes across structural and functional domains. Altogether, this analysis provides a context for looking at sequence divergence among functionally distinct loci within a developmental regulatory gene family.

Materials and Methods

Isolation and Sequencing of Floral Homeotic Genes

Arabidopsis lyrata seed was provided by C. H. Langley. Tissue for the *Arabidopsis thaliana* Landsberg erecta ecotype was obtained from single-seed propagated leaf material provided by the Arabidopsis Biological Resource Center.

The *Arabidopsis lyrata AP3, PI, CAL,* and *AP1* genes were isolated by PCR amplification using primers designed from *A. thaliana* sequences. Miniprep DNA was isolated from young leaves as previously described (Ausubel 1992). PCR was performed, with 40 cycles of 1 min at 95°C, 1 min at 52°C, and 3 min at 72°C, followed by 15 min at 72°C. The error-correcting recombinant Tth polymerase XL formulation (Perkin Elmer) was used to minimize nucleotide misincorporation. The error rate for this polymerase formulation, based on multiple amplification and resequencing of known genes, is less than 1 in 7,000 bp (unpublished data).

The isolation of the *A. lyrata CAL* gene is reported elsewhere (Purugganan and Suddith 1998). PCR primers were designed based on genomic sequences provided by M. F. Yanofsky (*AP1* and *CAL*) and E. M. Meyerowitz (*AP3* and *PI*). The *AP3*-specific primers AP3F (for exon 1 forward) (5′-GAATATGGCGAGAG-GGAAGATCC-3′) and AP3R (for exon 7 reverse) (5′-GCCTTTAAT-TATTCAAGA-AGATGG-3′) and the PI-specific primers PI-1F (for exon 1 forward) (5′-GAGAAAGA-TGGGTAGAGGAAG-3′) and PI-1R (for exon 6 reverse) (5′-ATCTCGATGATCAA-TCGATGACC-3′) were used in PCR reactions to amplify the *A. thaliana* and *A. lyrata AP3* and *PI* genes, respectively. The *A. lyrata* and *A. thaliana AP1* genes were isolated as two overlapping fragments. This reaction utilized primers AP1FPCR (5′-ATGGGAAGGGGTAGGGTTCA-3′) and AP1X2R (5′-ATTAATT-CCTGCCACCGATCC-3′) for the 5′ fragment and primers AP1X2F (5′-GTAAAA-GGTACTATTGAGAG-3′) and (5′-AAGGTTGCA-GTTGTAAACGGG-3′) for the 3′ *AP1* fragment. Amplified DNA was cloned into pCR2.1 using the TA cloning kit (InVitrogen). DNA sequencing for both genes was conducted with the ABI377 automated sequencer using a series of nine nested internal sense and antisense primers. Cloned genes were sequenced at least twice, and ambiguous sites were visually rechecked from chromatograms. The DNA sequences are available from GenBank (accession numbers AF143379–AF143382).

Data Analysis

Sequences for floral homeotic genes from *A. thaliana, Brassica oleraceae, Antirrhinum majus,* and *Silene latifolia* were obtained from GenBank. Sequences between *A. thaliana* and *A. lyrata* were visually aligned; more distantly related sequences were aligned using published alignment frameworks for plant MADS-box genes (Purugganan et al. 1995).

The nucleotide substitution distances between sequences were estimated using the Tajima-Nei model (Tajima and Nei 1984). For coding region sequences, synonymous (K_s) and nonsynonymous (K_a) substitution distances were estimated with Jukes-Cantor corrections (Nei and Gojobori 1986). The statistical significance of nucleotide substitution distance differences was evaluated using a *t*-test. Synonymous codon usage for each gene was determined using the MEGA program package (Kumar, Tamura, and Nei 1994), and the effective number of codons was estimated (Wright 1990).

Phylogenies of the different floral homeotic genes were estimated using both maximum-parsimony (MP) (Swofford 1993) and neighbor-joining (NJ) (Saitou and Nei 1987) techniques. For the MP technique, the heuristic search algorithm with the tree bisection-reconnection procedure of the PAUP program was used (Swofford 1993) with random addition of genes and with the MULPARS and collapse options in effect. Node confidence was assessed with 500 bootstrap replicates of the data. For the NJ analysis, genetic distances were estimated using either the Tajima-Nei distances or nonsynonymous substitution distances. Confidence estimates were assessed with 500 bootstrap replicates of the data. The NJ analysis was undertaken using the MEGA program package.

Nucleotide substitutions were mapped onto the gene phylogenies using MacClade to determine the proportion of molecular change across various gene domains (Maddison and Maddison 1992). The number of inferred replacement changes along phylogenetic branches was plotted in a sliding-window analysis with a sequential overlapping 25-aa partition. Significance of

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

differences in the distribution of replacement changes at different structural regions between paralog pairs was determined by calculating D_i: D_i = |proportion of protein 1 − protein 2 replacement differences within window i| for each sliding window i, and $D = \max(D_i)$ over all i sliding window positions. The distribution of D for any pair of proteins was estimated by calculating the maximum sliding window difference for 1,000 permutations of the protein residues.

Contingency tests for independence of coding region substitution categories were conducted using Fisher's exact test to evaluate significance. The coding-region variation was partitioned into functional (core and noncore) domains (Riechmann and Meyerowitz 1997) for separate contingency analyses (Templeton 1996). The relative-ratio test (Muse and Gaut 1997) was used to examine locus-by-lineage effects in gene evolution. The test was conducted between gene pairs; for those pairs involving *PI,* we excluded the *B. oleracea* sequences. This test was implemented using the program PROPML (Proportional Maximum Likelihood) provided by S. V. Muse (North Carolina State University).

Results and Discussion

Comparison of *A. thaliana* and *A. lyrata* Floral Homeotic Genes: Evolution at Short Temporal Scales

This study of the molecular evolution of developmental regulatory genes revolves around two questions: (1) Are there differing patterns of molecular evolution among orthologous developmental genes from species with diverse morphologies and (2) is the diversification of paralogous genes with specialized developmental functions accompanied by divergent patterns of molecular evolution?

Since the structural similarities among MADS-box floral homeotic genes reflect the biochemical similarity in their roles as transcriptional activators (Riechmann and Meyerowitz 1997), these four regulatory loci should display similar rates and patterns of molecular evolution. Alternatively, divergence in developmental function among these four regulatory loci in various species and different gene lineages may result in species- or lineage-specific variation in rates and patterns of molecular evolution between these developmental control genes (Purugganan 1998). In order to address this issue, the molecular evolution of four paralogous floral regulatory genes was analyzed. We isolated orthologs to the *A. thaliana AP1, CAL, AP3,* and *PI* loci from the congeneric Brassicaceae species *A. lyrata.* Orthologs to these *Arabidopsis* genes have also been identified in another Brassicaceae species (*B. oleracea*) as well as in *A. majus* (Scrophulariaceae) and *S. latifolia* (Caryophyllaceae) (table 1).

Neighbor-joining phylogenies of the different loci (see fig. 3) are congruent with previously established phylogenies (Purugganan et al. 1995; Purugganan 1997). The phylogenetic trees also show that the *AP1* and *CAL* genes duplicated before the split of *Arabidopsis* and *Brassica,* but after the separation of these Brassicaceae species from Scrophulariaceae and Caryophyllaceae (see

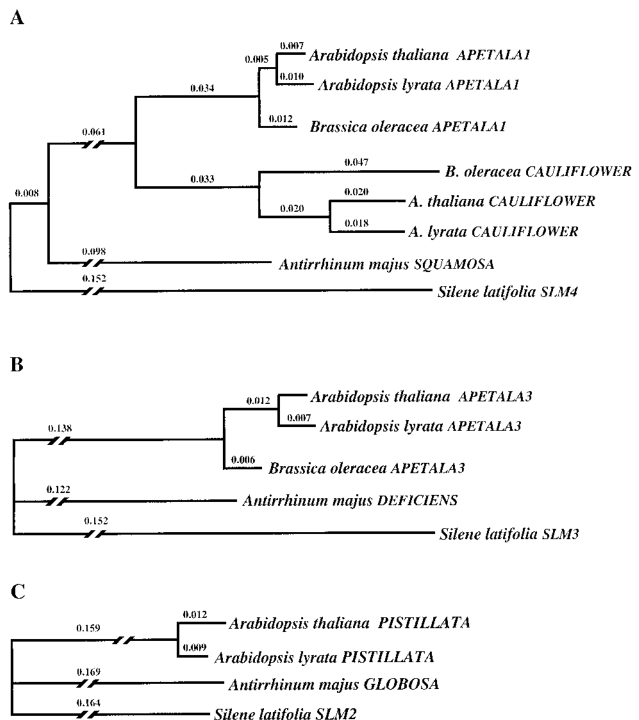

FIG. 3.—Phylogenies of the (*A*) *APETALA1* and *CAULIFLOWER,* (*B*) *APETALA3,* and (*C*) *PISTILLATA* genes. The branch lengths, estimated from neighbor-joining analysis, are indicated along each branch.

fig. 3). Comparison between *A. thaliana* and *A. lyrata* genes permits an analysis of regulatory gene diversification at relatively short evolutionary timescales (<5 MYA). The close relationship between *A. thaliana* and *A. lyrata* is reflected in the high sequence similarity between genes from these two species, with genomic sequences from the two taxa differing by approximately 3% at the nucleotide level. The intron/exon structures of the floral homeotic genes are also similar between the two species. There are several insertion/deletion (indel) differences within the intron regions of the various loci contributing to differences in genomic sequence lengths between the two *Arabidopsis* species. The *AP1* and *CAL* orthologs have 90 and 70 indels, respectively, ranging in size from 1 to 127 bp. There are fewer indels in *AP3* and *PI* genes between *A. thaliana* and *A. lyrata.* These two floral homeotic genes have 25 and 28 indels, respectively, from 1 to 50 bp in length.

The genes in these two species show similar patterns of moderate codon bias, with the effective number of codons (ENC) (Wright 1990) being approximately 56.7 for *CAL* and 55.9 for *AP3* and *PI.* The *AP1* locus in both species displays a slightly higher bias (ENC = 47.3) and is higher than the average for *Arabidopsis* nuclear genes (Miyashita et al. 1998). Since codon bias appears to be correlated with gene expression levels, this suggests that *AP1* may be expressed at higher levels in these species than are the three other floral regulatory genes (Sharp and Li 1986). Indeed, the multiple roles of *AP1* in floral meristem and organ identity (and the more limited developmental role of the three other loci)

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

Table 2
Sequence Distances Between *Aarabidopsis thaliana* and Other Brassicaceae Floral Homeotic Gene Orthologs

	VERSUS *Arabidopsis lyrata*				VERSUS *Brassica oleracea*			
	K (genomic)	K_a	K_s	K_a/K_s	*K* (genomic)[a]	K_a	K_s	K_a/K_s
APETALA1	0.079 ± 0.036	0.018 ± 0.006	0.109 ± 0.028	0.16	NA	0.024 ± 0.006	0.277 ± 0.049	0.09
CAULIFLOWER	0.120 ± 0.007	0.039 ± 0.008	0.142 ± 0.032	0.28	NA	0.085 ± 0.013	0.381 ± 0.061	0.22
APETALA3	0.089 ± 0.008	0.013 ± 0.005	0.098 ± 0.027	0.13	NA	0.024 ± 0.007	0.335 ± 0.057	0.07
PISTILLATA	0.087 ± 0.007	0.022 ± 0.007	0.163 ± 0.040	0.13	NA	NA	NA	NA

[a] Genomic sequence distance not available due to lack of intron information for *B. oleracea* genes.

provide support for the hypothesis that the higher codon bias for this gene may be associated with selection for translational efficiency.

The rates of sequence substitution at these floral homeotic loci can be directly compared between the two species. The four floral homeotic genes evolve at different rates between the two species (see table 2). The Tajima-Nei sequence distance values (*K*) for three genes (*AP1, AP3* and *PI*) are comparable, ranging from 0.079 to 0.089 nucleotide substitutions per site across the entire gene. In contrast, the *CAL* locus evolves at a faster rate than the other regulatory loci ($K = 0.120$, $P < 0.05$). The high rate in *CAL* is the product of a high nonsynonymous substitution rate compared with those of the other loci ($P < 0.05$ for *AP1* and *AP3*).

The faster rate of protein evolution for *CAL* is also reflected in the ratio of nonsynonymous to synonymous substitutions (K_a/K_s). The *CAL* gene has a K_a/K_s of 0.28, while the other three floral homeotic loci have a K_a/K_s ratio of 0.13 to 0.16 (see table 1). These estimates indicate that the rate of protein evolution is greater for *CAL* than for the other three MADS-box floral homeotic genes between these two closely related *Arabidopsis* species. Similar results are seen when *A. thaliana* and *A. lyrata CAL* genes are compared with the *B. oleracea* ortholog (table 2). The accelerated evolution of *CAL* within members of the Brassicaceae family may reflect its more recent origin within this group. The duplication of *CAL* and *AP1* occurred sometime during the evolution of the Brassicaceae, possibly early in the history of this eudicot family (Purugganan 1997). Genetic studies indicate a degree of genetic redundancy between *AP1* and *CAL* in floral meristem identity function (Kempin, Savidge, and Yanofsky 1995). The redundancy of *CAL* to *AP1* may be reflected in the higher rate of evolution for this locus than for the other floral homeotic genes that are presumably under stronger stabilizing selection.

The Floral Homeotic Genes Do Not Display Significant Locus-by-Lineage Effects

Plant genes may evolve at different rates along different species lineages (Gaut et al. 1997). Several mechanisms, such as generation time, life history, and global mutation rates, will affect all loci within a genome to a similar extent. Alternatively, selective forces affecting individual loci that may vary between different species will result in uncorrelated evolutionary rates along specific taxonomic lineages (Gillespie 1991). The relative-ratio test evaluates whether the rates of molecular evolution are correlated among loci in various species lineages (Muse and Gaut 1997). Locus-by-lineage effects are characterized by uncorrelated variation among loci in the rates of evolution in different species lineages and may reflect variation in selective forces among different genes in different species.

The relative-ratio test does not reject the null hypothesis of correlated rates of molecular evolution among all four floral homeotic loci in different species lineages. The relative ratios of nonsynonymous distances between floral homeotic genes among different species lineages are remarkably similar, implying similar patterns of evolutionary rate among the genes between taxa. This correlation in rates is present even though the different genera utilized in the study (*Arabidopsis, Brassica, Antirrhinum,* and *Silene*) display considerable variation in inflorescence form, symmetry and size of flowers, and number, size, and shape of floral organs. This suggests that selection for floral morphological diversification between taxa is not associated with large, statistically significant differences in rates of molecular evolution among these specific taxonomic lineages.

Patterns of Divergence Between Functional Domains

Molecular genetic studies have delineated the functional core sequences of several MADS-box floral homeotic loci. This core region, which encompasses the MADS-box, the I-region, and the first 16 aa of the K-domain, has been shown to be necessary for dimerization and DNA-binding activities of these proteins (see fig. 1) (Riechmann and Meyerowitz 1997). The noncore region, which includes the 3′ half of the K-domain and all of the C-terminal region, does not appear to be important for DNA binding. The noncore region, however, includes sequences that may serve as the transcriptional activation domain (Riechmann and Meyerowitz 1997).

Previous work indicates that the levels of nonsynonymous substitutions are generally higher in the sequence encoding the C-terminal regions of plant MADS-box genes. Calibrated rates of nonsynonymous substitutions in the C-terminal region are found to be 79×10^{-10} nonsynonymous substitutions per site per year, compared with 3×10^{-10} nonsynonymous substitutions per site per year for the highly conserved MADS-box region, suggesting that the noncore region contains sequence elements that consistently display a greater degree of sequence divergence (Purugganan et al. 1995). However, mutations in the C-terminal region of several MADS-box genes are known to produce floral homeotic phenotypes (Kempin, Savidge, and Yanofsky 1995), in-

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

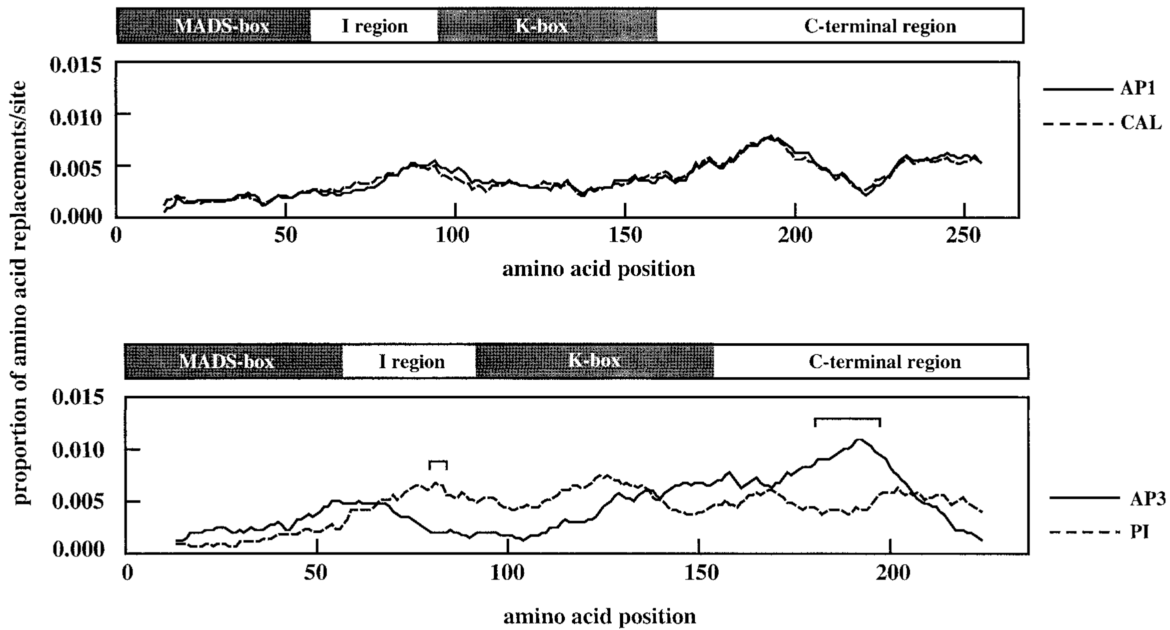

FIG. 4.—Amino acid site diversity profiles of paralogous floral homeotic proteins. The number of replacement substitutions was counted over a maximum-parsimony tree of the genes under study. The relative positions of the various protein structural domains are shown at the top of each profile. Regions of significant differences between the AP3 and PI paralog pair are shown in brackets ($P < 0.05$). There are no regions of significant difference between the AP1 and CAL proteins.

dicating that this domain is required for protein function despite its rapid evolutionary rate.

The distributions of nucleotide substitutions between these sequence regions provide information on the differences in selective forces that operate between these two functional domains. Similar selective constraints on all four floral homeotic genes should result in a similar distribution of replacement substitutions between core and noncore domains (Templeton 1996). A sliding-window analysis plotting the number of amino acid replacements detected within a gene from all five study species indicates that different loci have different patterns of conservation and divergence (see fig. 4). Both visual inspection of these graphs and permutation testing suggest that the closely related paralogs CAL and AP1 exhibit nearly identical patterns. This result may reflect both the relatively recent duplication of these loci and the use of the same *Silene* and *Antirrhinum* genes as outgroup sequences in the analysis of both CAL and AP1 variation. The paralogs AP3 and PI do show substantial differences from one another. A permutation test indicates that variation between portions of the AP3 and PI I- and C-terminal regions are significant ($P < 0.05$). These differences in amino acid replacement patterns suggest that each locus evolves differently despite similarities in overall structural organization between the paralogs.

A contingency analysis was undertaken to test whether relative levels of replacement and synonymous substitutions were similar between functional regions in all floral homeotic genes (Templeton 1996). These contingency tests indicate that the distribution patterns of nucleotide substitutions differed between floral homeotic loci when *A. thaliana* and *A. lyrata* were compared. The *AP3* gene appears to show substantial constraint in both the core and the noncore domains (see table 4). Only 2 of the 6 substitutions in the core domain and 2 of the 10 substitutions in the noncore domain were replacement changes; the differences between core and noncore regions are not statistically significant (Fisher's exact test, $P = 0.604$). In *AP1, CAL,* and *PI,* however, the noncore domain exhibited as many or more nonsynonymous substitutions relative to synonymous substitutions than did the core domains (table 4). The differ-

Table 3
Sequence Distances Between Coding Regions of *Arabidopsis thaliana* and Distant Floral Homeotic Gene Orthologs

	Versus *Brassica oleracea*	Versus *Antirrhinum majus*	Versus *Silene latifolia*
APETALA1	0.070 ± 0.010	0.365 ± 0.028	0.449 ± 0.034
CAULIFLOWER	0.140 ± 0.015	0.410 ± 0.031[a]	0.458 ± 0.034[a]
APETALA3	0.083 ± 0.012	0.437 ± 0.033	0.479 ± 0.036
PISTILLATA	NA[b]	0.489 ± 0.038	0.538 ± 0.041

[a] Estimates of CAL and *AP1* divergence between *A. thaliana* and non-Brassicaceae species are not independent due to recent *AP1/CAL* duplication.

[b] *Brassica oleracea PI* ortholog not available.

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

Table 4
Distribution of Coding Region Differences Between Functional Domains

	APETALA1		*CAULIFLOWER*		*APETALA3*		*PISTILLATA*	
	Core	Noncore	Core	Noncore	Core	Noncore	Core	Noncore
Replacement	0	10	5	12	2	2	1	9
Silent	6	7	10	10	4	8	8	9
	$P = 0.019$*		$P = 0.315$		$P = 0.604$		$P = 0.091$	

* Significant at 5% level.

ences between core and noncore domains, however, are significant only for *AP1,* for which 10 out of 17 differences in the noncore domain are nonsynonymous, while no substitutions in the core region result in amino acid replacements (Fisher's exact test, $P = 0.019$).

The patterns of nonsynonymous evolution differ between different evolutionary timescales (see table 5). In general, the rate of nonsynonymous evolution for the noncore region is approximately two to three times the rate for the core domain when very distantly related species (*A. thaliana* vs. *A. majus* or *S. latifolia*) are examined. This trend does not appear to hold for *AP3*; the noncore region has half as many nonsynonymous substitutions as the core region in comparisons between *A. thaliana* and *B. oleracea* (0.034 and 0.017 nonsynonymous substitutions per site for the *AP3* core and noncore domains, respectively), and over longer evolutionary distances, the situation is reversed in *AP3.*

Evolutionary Dynamics of a Regulatory Gene Pathway

It has been shown in several cases that gene families controlling morphogenesis typically contain both paralogous and orthologous members with diverged developmental functions (Ruddle et al. 1994). This variation in regulatory function, however, proceeds in the context of remarkable conservation of structure in key domains within such control loci (Gerhart and Kirschner 1997). This apparent structural conservation underlies the similarity in biochemical function (e.g., DNA-binding transcriptional activation) among members of a regulatory gene family. These observations suggest that changes in regulatory gene sequences play a secondary role during the evolution of eukaryotic developmental systems and that changes in regulatory gene expression patterns provide the major mechanism by which control genes diverge (Gerhart and Kirschner 1997).

If differential expression patterns and not the actual sequences of regulatory loci are crucial to the divergent functions, then the evolutionary dynamics of regulatory protein structure should be similar across homologous loci. Our analyses of floral homeotic genes of different taxa indeed suggest that there are no apparent significant differences in the rates of evolution of orthologous plant MADS-box loci. Relative-rate tests do not show any appreciable rate differences between orthologous loci (Purugganan 1997). Moreover, relative-ratio tests do not detect significant floral homeotic locus-by-lineage effects. This indicates that the rates of molecular evolution of different floral homeotic genes are similar across the flowering plant taxa analyzed.

Such similarity in molecular evolutionary rates across different taxonomic lineages is in contrast to the significant variation in floral and inflorescence morphologies displayed by some of the species in this study. This suggests that selective differences between orthologous loci in these different taxa are not manifested in dramatic differences in the rates of nucleotide substitution. This would lend support to the assertion that differential regulation, and not differential structure, may be the major component of regulatory gene evolution (Gerhart and Kirschner 1997). However, it is also probable that orthologous regulatory genes may be less likely to display significantly different rates of molecular evolution as a result of species-specific variation in selection pressures. For example, if evolutionary changes in the sequences of these regulatory proteins are crucial to interspecific diversification of function, then these changes may be confined to only a few amino acid positions and would not significantly impact the overall between-species nucleotide substitution rates. There is evidence, moreover, for accelerated protein evolution rates in orthologous floral homeotic genes at short timescales in plant adaptive radiations (unpublished data).

While variation in the patterns of molecular substitutions is not evident between orthologous floral homeotic genes among our taxa, there appear to be significant differences in the evolutionary dynamics of paralogous regulatory genes performing distinct developmental functions. The divergence in floral developmental function of the four homeotic genes in this study appears to be accompanied by variation in the patterns of molecular evolution among these loci. Comparisons between the floral homeotic genes found in Brassicaceae species, for example, suggest that at least one locus

Table 5
Comparison of Nonsynonymous Substitutions Between Core and Noncore Domains

	Versus *Brassica oleracea*	Versus *Antirrhinum majus*	Versus *Silene latifolia*
APETALA1			
Core	0.004 ± 0.004	0.138 ± 0.025	0.184 ± 0.030
Noncore	0.039 ± 0.011	0.263 ± 0.033	0.350 ± 0.040
APETALA3			
Core	0.034 ± 0.012	0.178 ± 0.030	0.222 ± 0.034
Noncore	0.017 ± 0.008	0.376 ± 0.040	0.395 ± 0.050
PISTILLATA			
Core	NA[a]	0.178 ± 0.030	0.176 ± 0.030
Noncore	NA[a]	0.560 ± 0.070	0.570 ± 0.070

[a] The *B. oleracea PI* gene is not available.

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

(*CAL*) evolves at a significantly higher rate than other paralogous genes. This increased rate of *CAL* evolution is also observed at nonsynonymous sites, but not at synonymous nucleotide positions, suggesting that variation in the rate of protein evolution that exists among paralogous loci is not merely due to differences in the neutral mutation rates.

Our results also indicate that paralogous developmental regulatory genes may display differences in the patterns of nucleotide substitutions between functional domains, most likely as a result of differences in the selective pressures experienced by each gene. Similarity in selective constraints between loci would be expected to manifest itself in the conservation of nucleotide substitution patterns between functional domains among the four floral homeotic genes. Analyses of the distribution of nucleotide substitutions indicate that molecular changes are not partitioned equally among the functional domains. In general, the *AP1, CAL,* and *PI* genes display greater degrees of sequence constraint in the core domain, with the noncore regions showing higher levels of nonsynonymous substitutions. *AP3,* however, has consistently high levels of sequence constraint across both domains. There are differences in the patterns of substitution despite similarities in developmental function among the loci. The *AP3* and *PI* genes, for example, are both required for petal and stamen development, but the distribution of replacement substitutions between functional domains differs significantly between these two floral homeotic genes.

The emerging picture from this analysis is a contrasting portrait of regulatory gene evolution between paralogous members of a developmental gene family and orthologs of specific genes found in morphologically distinct taxa. Interspecific evolution of these floral regulatory loci does not appear to be associated with striking variation in evolutionary rates between orthologs. Divergence of developmental function between paralogous regulatory loci, on the other hand, appears to be associated with significant differences in the rates and patterns of molecular evolution. Our results suggest that diversification of regulatory gene families by duplication results in the formation of loci with distinct and contrasting evolutionary dynamics.

Acknowledgments

We would like to thank S. V. Muse and J. Thorne for stimulating discussions. We would also like to thank M. F. Yanofsky and E. M. Meyerowitz for providing us with sequence information. This work was supported in part by an NIH postdoctoral fellowship to E.S.B. and grants from the USDA NRICGP and the National Science Foundation to M.D.P.

LITERATURE CITED

Atchley, W. R., and W. M. Fitch. 1995. Myc and Max—molecular evolution of a family of proto-oncogene products and their dimerization partner. Proc. Natl. Acad. Sci. USA **92**:10217–10221.

Ausubel, F. 1992 Short protocols in molecular biology. John Wiley, New York.

Bailey, W. J., J. Kim, G. P. Wagner, and F. H. Ruddle. 1997. Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol. Biol. Evol. **14**:843–853.

Bowman, J. L., J. Alvarez, D. Weigel, E. M. Meyerowitz, and D. Smyth. 1993. Control of flower development in *Arabidopsis thaliana* by *AP1* and interacting genes. Development **119**:721–743.

Bowman, J. L., D. R. Smyth, and E. M. Meyerowitz. 1991. Genetic interactions among floral homeotic genes of *Arabidopsis.* Development **112**:1–20.

Carroll, S. B. 1995. Homeotic genes and the evolution of arthropods and chordates. **376**:479–485.

Coen, E. S. 1991. The role of homeotic genes in flower development and evolution. Annu. Rev. Plant Physiol. Plant Mol. Biol. **42**:241–279.

Crane, P. R., E. M. Friis, and K. R. Pedersen. 1995. The origin and early diversification of angiosperms. Nature **374**: 27–33.

Gaut, B. S., L. G. Clark, J. F. Wendel, and S. V. Muse. 1997. Comparisons of the evolutionary processes at *rbcL* and *ndhF* in the grass family. Mol. Biol. Evol. **14**:769–777.

Gerhart, J., and M. Kirschner. 1997. Cells, embryos and evolution. Blackwell Science, Malden, Mass.

Gillespie, J. 1991. The causes of molecular evolution. Oxford University Press, Oxford, England.

Goto, K., and E. M. Meyerowitz. 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA.* Genes Dev. **8**:1548–1560.

Hardenack, S., D. Ye, H. Saedler, and S. Grantnm. 1994. Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. Plant Cell **6**:1775–1787.

Huijser, P., J. Klein, W.-E. Lonnig, H. Meijer, H. Saedler, and H. Sommer. 1992. Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *SQUAMOSA* in *Antirrhinum majus.* EMBO J. **11**: 1239–1249.

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

Kempin, S. A., B. Savidge, and M. F. Yanofsky. 1995. Molecular basis of the cauliflower phenotype in *Arabidopsis.* Science **267**:522–525.

Kramer, E. M., R. L. Dorit, and V. F. Irish. 1998. Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *AP3* and *PI* MADS-box gene lineages. Genetics **149**:765–783.

Krizek, B. A., and E. M. Meyerowitz. 1996. Mapping the protein regions responsible for the functional specificities of the *Arabidopsis* MADS domain organ identity proteins. Proc. Natl. Acad. Sci. USA **93**:4063–4070.

Kumar, S., K. Tamura, and M. Nei. 1994 Molecular evolutionary genetics analysis package. Version 1.1. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park.

Lowman, A. C., and M. D. Purugganan. 1999. Duplication of the *Brassica oleracea APETALA1* floral homeotic gene and the evolution of domesticated cauliflower. J. Hered. (in press).

Ma, H., M. F. Yanofsky, and E. M. Meyerowitz. 1991. *AGL1-AGL6,* an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev. **5**:484–495.

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Mass.

MANDEL, M. A., C. GUSTAFSONBROWN, B. SAVIDGE, and M. F. YANOFSKY. 1992. Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature **360**: 273–277.

MIYASHITA, N. T., A. KAWABE, H. INNAN, and R. TERAUCHI. 1998. Intra- and interspecific DNA variation and codon bias of the alcohol dehydrogenase (*Adh*) locus in *Arabis* and *Arabidopsis* species. Mol. Biol. Evol. **15**:1420–1429.

MIZUKAMI, Y., H. HUANG, M. TUDOR, Y. HU, and H. MA. 1996. Functional domains of the floral regulator AG—characterization of the DNA binding domain and analysis of dominant negative mutations. Plant Cell **8**:831–845.

MUSE, S. V., and B. S. GAUT. 1997. Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. Genetics **146**:393–399.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

OHNO, S. 1970. Evolution by gene duplication. Springer-Verlag, New York.

POLLOCK, R., and R. TREISMAN. 1991. Human SRF-related proteins: DNA binding properties and potential regulatory targets. Genes Dev. **5**:2327–2341.

PURUGGANAN, M. D. 1997. The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. J. Mol. Evol. **5**:392–396.

———. 1998. The molecular evolution of development. Bioessays **20**:700–711.

PURUGGANAN, M. D., S. D. ROUNSLEY, R. J. SCHMIDT, and M. F. YANOFSKY. 1995. Molecular evolution of flower development—diversification of the plant MADS-box regulatory gene family. Genetics **140**:345–356.

PURUGGANAN, M. D., and J. I. SUDDITH. 1998. Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95**:8130–8134.

RIECHMANN, J. L., B. A. KRIZEK, and E. M. MEYEROWITZ. 1996. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins AP1, AP3, PI and AG. Proc. Natl. Acad. Sci. USA **93**:4793–4798.

RIECHMANN, J. L., and E. M. MEYEROWITZ. 1997. MADS domain proteins in plant development. Biol. Chem. **378**:1079–1101.

RIECHMANN, J. L., M. Q. WANG, and E. M. MEYEROWITZ. 1996. DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins AP1, AP3, PI and AG. Nucleic Acids Res. **24**:3134–3141.

ROSINSKI, J. A., and W. R. ATCHLEY. 1998. Molecular evolution of the *myb* family of transcription factors: evidence for polyphyletic origin. J. Mol. Evol. **46**:74–83.

RUDDLE, F. H., J. L. BARTELS, K. L. BENTLEY, C. KAPPEN, M. T. MURTHA, and J. W. PENDLETON. 1994. Evolution of Hox genes. Annu. Rev. Genet. **28**:423–442.

SAEDLER, H., and P. HUIJSER. 1993. Molecular biology of flower development in *Antirrhinum majus.* Gene **135**:239–243.

SAITOU, N., and M. NEI. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SCOTT, M. P., and A. J. WEINER. 1984. Structural relationships among genes that control development: sequence homology between the *Antennapedia, Ultrabithorax,* and *fushi tarazu* loci of *Drosophila.* Proc. Natl. Acad. Sci. USA **81**:4115–4119.

SHARP, P., and W.-H. LI. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. **24**:28–38.

SOMMER, H., J. P. BELTRAN, P. HUIJSER, H. POPE, W.-E. LOENNIG, and H. SAEDLER. 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.

SWOFFORD, D. 1993. Phylogenetic analysis using parsimony. Version 3.1. Illinois Natural History Survey, Champaign.

TAJIMA, F., and M. NEI. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**: 269–285.

TEMPLETON, A. R. 1996 Contingency tests of neutrality using intra/interspecific gene trees—rejection of neutrality for the evolution of the mitochondrial *cytochrome oxidase II* gene in the hominoid primates. Genetics **144**:1263–1270.

TROBNER, W., L. RAMIREZ, P. MOTTE, I. HUE, P. HUIJSER, W. E. LONNIG, H. SAEDLER, H. SOMMER, and Z. SCHWARZ-SOMMER. 1992. *GLOBOSA*—a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. EMBO J. **11**:4693–4704.

WRIGHT, F. 1990. The 'effective number of codons' used in a gene. Gene **87**:23–29.

YANOFSKY, M. F. 1995. Floral meristems to floral organs—genes controlling early events in *Arabidopsis* flower development. Annu. Rev. Plant Physiol. Plant Mol. Biol. **46**:167–188.

ZHANG, J. Z., and M. NEI. 1996. Evolution of *Antennapedia*-class homeobox genes. Genetics **142**:295–303.

MARCY K. UYENOYAMA, reviewing editor
Accepted April 22, 1999

Downloaded from https://academic.oup.com/mbe/article/16/8/1037/2925481 by guest on 30 July 2021