Molecular Evolution of Flower Development: Diversification of the Plant MADS-Box Regulatory Gene Family

Michael D. Purugganan, Steven D. Rounsley, Robert J. Schmidt and Martin F. Yanofsky

Department of Biology, University of California at San Diego, La Jolla, California 92093

Manuscript received October 4, 1994
Accepted for publication February 6, 1995

ABSTRACT

Floral homeotic genes that control the specification of meristem and organ identity in developing flowers have been isolated from both *Arabidopsis thaliana* and *Antirrhinum majus*. Most of these genes belong to a large family of regulatory genes and possess a characteristic DNA binding domain known as the MADS-box. Members of this gene family display primarily floral-specific expression and are homologous to transcription factors found in several animal and fungal species. Molecular evolutionary analyses reveal that there are appreciable differences in the substitution rates between different domains of these plant MADS-box genes. Phylogenetic analyses also demonstrate that members of the plant MADS-box gene family are organized into several distinct gene groups: the AGAMOUS, APETALA3/PISTILLATA and APETALA1/AGL9 groups. The shared evolutionary history of members of a gene group appear to reflect the distinct functional roles these MADS-box genes play in flower development. Molecular evolutionary analyses also suggest that these different gene groups were established in a relatively short span of evolutionary time and that the various floral homeotic loci originated even before the appearance of the flowering plants.

ALTERATIONS in the spatial and temporal progression of developmental events are believed to underlie the evolutionary diversification of eukaryotic morphology (GOULD 1977). These developmental transformations proceed, in part, by variations in the structure and regulation of genes that control developmental mechanisms (RAFF and KAUFMANN 1983; DOEBLEY 1993). The central relevance of regulatory gene evolution in the process of organismal diversification has prompted renewed interest in the molecular evolution of loci that control pattern formation (DICKINSON 1991; SIDOW 1992; KAPPEN *et al.* 1993; PURUGGANAN and WESSLER 1994).

The development of floral structures provides an excellent system to investigate the molecular evolution of morphological differentiation in plants. Flowers are determinate sporophyll-bearing shoots that serve as reproductive structures among the angiosperms. Floral development begins with the transition of the shoot vegetative meristem to an inflorescence meristem. Floral primordia or meristems subsequently form at the flanks of this inflorescence meristem, differentiating to produce the perianth and reproductive organs of the flower. In the dicot *Arabidopsis thaliana*, which has served as a model system for the study of flower development, these organs are arranged in a characteristic four-whorled pattern. The first floral whorl of the Arabidopsis flower contains sepals followed by a whorl of petals. The third and fourth whorls contain the stamens and carpels, respectively. The precise number, size, and disposition of these floral organs vary enormously between plant taxa and are, in many cases, specifically adapted to particular pollination strategies (RICHARDS 1986). The broad organization of the flower, however, remains invariant between species. In all cases, the perianth structures (sepals and petals) are arranged at the periphery of the flower, while the reproductive organs (stamens and carpels) are centrally positioned.

Several floral homeotic loci have been identified in both Arabidopsis and *Antirrhinum majus* (snapdragon); mutations in these loci result in alterations in floral organ development (COEN 1991; COEN and MEYEROWITZ 1991). Genetic and molecular characterization of these homeotic loci indicate that these genes fall into two broad classes: meristem identity genes, which specify inflorescence and floral mersitem identity, and floral organ identity genes, which define the identities of the organs in the flower. The *APETALA2* (*AP2*), *APETALA3* (*AP3*), *PISTILLATA* (*PI*) and *AGAMOUS* (*AG*) genes, for example, regulate floral organ identity in the whorled Arabidopsis flower. The *LEAFY* (*LFY*), *CAULIFLOWER* (*CAL*) and *APETALA1* (*AP1*) loci, on the other hand, are examples of Arabidopsis meristem identity genes. Some genes appear to be involved in specifying both meristem and organ identity; the *AP1* locus, for example, participates in regulating both developmental functions (IRISH *et al.* 1990; BOWMAN *et al.* 1993).

Several floral homeotic genes have recently been isolated in Arabidopsis and Antirrhinum (SOMMER *et al.*

Corresponding author: Michael D. Purugganan, Department of Biology-0116, University of California at San Diego, La Jolla, CA 92093. E-mail: mikep@jeeves.uscd.edu

1990; YANOFSKY *et al.* 1990; HUIJSER *et al.* 1992; JACK *et al.* 1992; MANDEL *et al.* 1992; TROBNER *et al.* 1992; WEIGEL *et al.* 1992; BRADLEY *et al.* 1993; GOTO and MEYEROWITZ 1994). Most of the floral homeotic genes isolated to date encode proteins of ~260 amino acids that share regions of sequence similarity to each other, suggesting that these genes share a common evolutionary ancestor. Together, these loci are characterized by the presence of a highly conserved region of ~57 amino acids called the MADS domain and comprise a regulatory gene family referred to as the plant MADS-box gene family (MA *et al.* 1991; DAVIES and SCHWARZ-SOMMER 1994). The MADS domain was first identified in the *AGAMOUS* and Antirrhinum *DEFICIENS* (*DEF*) genes, as well as in transcription factors found in humans (SRF) and yeast (MCM1) (POLLOCK and TREISMANN 1991). Molecular studies have demonstrated that this domain acts as a sequence-specific DNA-binding moiety in these regulatory proteins.

These floral homeotic MADS-box genes also contain three other distinct regions of limited similarity outside of the MADS domain (MA *et al.* 1991) (see Figure 1). One distinctive region is a moderately conserved 70-amino acid region called the K-domain. Named for its similarity to regions of the keratin molecule, the structure of the K-domain appears capable of forming amphipathic helices. The K-domain could thus serve as a dimerization moiety, forming coiled-coil structures that mediate protein-protein interactions (DAVIES and SCHWARZ-SOMMER 1994).

Outside these two conserved domains are two regions of considerable sequence variability: a short intervening region (I region) that separates the MADS- and K-domains and the C-terminal region downstream of the K-domain, which exhibits relatively poor sequence conservation between different genes. An N-terminal sequence of variable size is also found immediately upstream of the MADS-domain of several genes, including the Arabidopsis *AG* and Antirrhinum *PLENA* loci (SOMMER *et al.* 1990; YANOFSKY *et al.* 1990).

Work in both Arabidopsis and Antirrhinum has demonstrated that members of the plant MADS-box gene family direct several crucial steps in the development of the flower. In Arabidopsis, for example, two of the three early acting genes that control floral meristem identity (*AP1* and *CAL*) are MADS-box genes. In the later steps of flower development, three overlapping homeotic functions, referred to as the **A**, **B** and **C** functions (COEN and MEYEROWITZ 1991), regulate the identity of organs in the different Arabidopsis floral whorls. Four of the genes necessary for these organ identity functions (*AP1*, *AP3*, *PI* and *AG*) are also MADS-box genes. Several other MADS-box genes have been shown to be expressed in the flower between the early acting meristem identity genes and the late-acting floral organ identity loci, suggesting that they mediate between the meristem specification and organ identity functions (A. MANDEL and M. YANOFSKY, unpublished data; B. SAVIDGE, S. ROUNSLEY and M. YANOFSKY, unpublished results).

Homologs to the Arabidopsis and Antirrhinum MADS-box floral homeotic genes have been isolated from several other dicot (*Lycopersicon esculentum*, *Petunia hybrida*, *Solanum tuberosum*, *Nicotiana tabacum*) and monocot (*Zea mays*, *Aranda deborah*) species (GARCIA-MAROTO *et al.* 1993; PNUELI *et al.* 1991; ANGENENT *et al.* 1992; HANSEN *et al.* 1993; SCHMIDT *et al.* 1993; PNUELI *et al.* 1994a). The precise functions of many of these genes remain uncertain, although the structure and expression patterns of several homologs indicate that these loci also participate in regulating floral development in distant taxa. Although most MADS-box genes display floral-specific expression patterns, some Arabidopsis genes are also expressed in vegetative tissues, suggesting that they are involved in other aspects of plant development as well (MA *et al.* 1991; PNUELI *et al.* 1991; ANGENENT *et al.* 1992; S. ROUNSLEY, G. DITTA and M. YANOFSKY, unpublished results).

In this report, we describe the molecular evolution of the plant MADS-box regulatory gene family. Analyses reveal that there is substantial heterogeneity in the rates of evolution of the different domains within these regulatory genes. Phylogenetic analyses also demonstrate that most plant MADS-box genes are organized into distinct gene groups and each group contains loci that share similar functions in regulating floral development. These floral homeotic gene lineages appear to have diversified rapidly and may have originated even before the evolution of the flowering plants.

MATERIALS AND METHODS

The Arabidopsis genes *AGL11–AGL15* and *AGL17* were recently isolated and the sequences deposited in Genbank (U20182–U20186, U20193) (S. ROUNSLEY, G. DITTA and M. YANOFSKY, in preparation). Sequences for *AGL8* and *AGL9* were provided by A. MANDEL and M. YANOFSKY. The sequence for the unpublished portion of *AGL3* was obtained from H. MA (Cold Spring Harbor Laboratory). All other genes used in these analyses were obtained from GENBANK. Alignments of both nucleotide and protein sequences for the plant MADS-box genes were made using PILEUP of the UWGCG package, and refined visually taking both amino acid and nucleotide sequences into consideration.

Most of the distance calculations were done using the Molecular Evolutionary Genetic Analyses package (KUMAR *et al.* 1994). Levels of nucleotide substitutions were calculated using the Tajima and Nei method (TAJIMA and NEI 1984). For nonsynonymous and synonymous substitution calculations, both the NEI and GOJOBORI (1986) and and LI *et al.* (LWL) (1985) substitution estimates were used. An unbiased estimation procedure for the LWL method implemented in the LI93 program was utilized (LI 1993). Both the Nei-Gojobori and LWL nonsynonymous substitution estimates were not significantly different from each other.

Phylogenetic analyses using parsimony methods were undertaken using the PAUP program (SWOFFORD 1993). Only the first and second codon positions were utilized, and transitions and transversions were weighted equally. The transver-

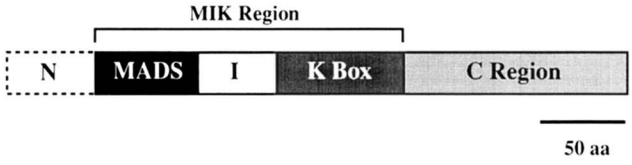

FIGURE 1.—Different regions within the plant MADS-box genes. The MADS-box encodes a putative DNA-binding domain, and is the most conserved sequence found between members of this gene family. The K-box and C-terminal region are more divergent sequences. A short region (the I region) links the MADS- and K-boxes. In several genes closely related to the Arabidopsis *AGAMOUS* gene, a variably sized N-terminal region is found upstream of the MADS-box. The bracket delineates the MIK region used in the phylogenetic analyses.

sion/transition substitution ratio derived from the sequence data ranges from 1 to 3. The analysis, however, appears relatively insensitive to transversion weights used; utilizing a transversion:transition weighting scheme of 3:1 and 10:1 does not significantly affect the tree topology. The tree construction was done with the PAUP heuristic search algorithm, with branch swapping performed using the tree-bisection-reconnection procedure. The PAUP MULPARS option was in effect. The neighbor-joining procedure in the MEGA package was used in distance-based phylogenetic analyses (SAITOU and NEI 1987). The Tajima-Nei nucleotide distance was used in the neighbor-joining reconstruction, with deletions ignored in pairwise comparisons and only first and second codon positions considered. The mammalian *MEF2C* MADS-box gene was used as an outgroup in both analyses (LEIFER *et al.* 1993).

RESULTS

Phylogenetic relationships between plant MADS-box genes: The phylogenetic relationships between different plant MADS-box genes were analysed using both parsimony and neighbor-joining methods in an effort to determine the rates and patterns of evolutionary diversification within this large gene family. A 507-bp sequence encompassing the MADS-box, the intervening I region and the K-box were utilized in the analyses (see Figure 1); the sequence of this region can be aligned reasonably well even between fairly distant plant MADS-box genes (Figure 2). Only the first and second codon positions were used in reconstructing the evolutionary history of this gene family.

Figure 3 shows a weighted parsimony-based phylogenetic tree depicting the relationships between the different plant MADS-box genes. Supports are estimated for different nodes from the results of 100 bootstrap replicates. The topology of this tree agrees with a phylogenetic tree derived using the neighbor-joining procedure, based on sequence distances calculated by the Tajima and Nei method. Figure 4 shows a bootstrapped neighbor-joining analysis, with node support estimated from 500 bootstrap replicates.

Phylogenetic analyses reveal that most of the members of the plant MADS-box regulatory gene family are organized into three major monophyletic groups. These groups are the AGAMOUS group, composed of 10 genes, including the Arabidopsis *AG*, the Antirrhinum *PLENA* and the maize *ZAG1* genes; the APETALA3/PISTILLATA group, also containing 10 genes and including the Arabidopsis *AP3* and *PI* genes, and the Antirrhinum *GLOBOSA* and *DEFICIENS* genes; and finally the APETALA1/AGL9 group, composed of 14 genes, including the Arabidopsis *AP1*, *CAL* and *AGL9* genes and the Antirrhinum *SQUAMOSA* locus.

Both the AP3/PI and AP1/AGL9 groups can be further subdivided into distinct monophyletic gene clades. The AP3/PI group includes two subgroups: an AP3 clade containing *AP3* and *DEFICIENS* and a PI clade containing *PI* and *GLOBOSA*. Both of these subgroups have very strong bootstrap support—99% in the parsimony analyses and 94% in the neighbor-joining trees. Genetic and molecular studies in Arabidopsis have shown that both the *AP3* and *PI* genes are expressed in the same regions of the flower and participate in specifying petal and stamen identity (JACK *et al.* 1992; GOTO and MEYEROWITZ 1993). Thus, the close evolutionary relationship between the AP3 and PI gene clades reflects the parallel regulatory functions these loci perform in floral development.

Within the AP1/AGL9 group are three distinct lineages. These include the AP1 clade, which contains the Arabidopsis *AP1*, *CAL* and *AGL8* genes and the Antirrhinum *SQUAMOSA* locus, the AGL9 clade, which contains the Arabidopsis *AGL9*, Lycopersicon gene *TM5* and the Petunia Floral Binding Protein 2 (*FBP2*) gene, and the AGL6/AGL13 clade, which contains the two Arabidopsis *AGAMOUS*-like genes *AGL6* and *AGL13*. The bootstrap support for these groups are 90% in the parsimony tree. In the neighbor-joining analysis, however, the AGL9 and AGL6/AGL13 clades form a monophyletic subgroup with moderate bootstrap support (84%).

Many of the plant species analyzed possess more than one loci from each gene group. The AP1 clade, for example, contains three distinct Arabidopsis genes: *AP1*, *CAL* and *AGL8*. The large number of genes found within each species' genome suggests that gene diversification via duplication has accompanied the molecular evolution of the different lineages of the plant MADS-box gene family.

The phylogenetic tree also reveals a class of genes, tentatively designated as "orphan genes," whose members do not belong to any of the three major gene groups. These include the Arabidopsis genes *AGL12*, *AGL14*, *AGL15* and *AGL17*, and the Lycopersicon genes *TM3* and *TM8*. The *AGL14* and *TM3* genes appear to belong to the same group, and it is likely that each of the other genes are also members of distinct MADS-box gene groups whose orthologues in other species have yet to be isolated. Both the parsimony and distance-based analyses do reveal that one orphan gene, the Lycopersicon gene *TM8*, represents the basal lineage of the plant MADS-box gene family. The basal

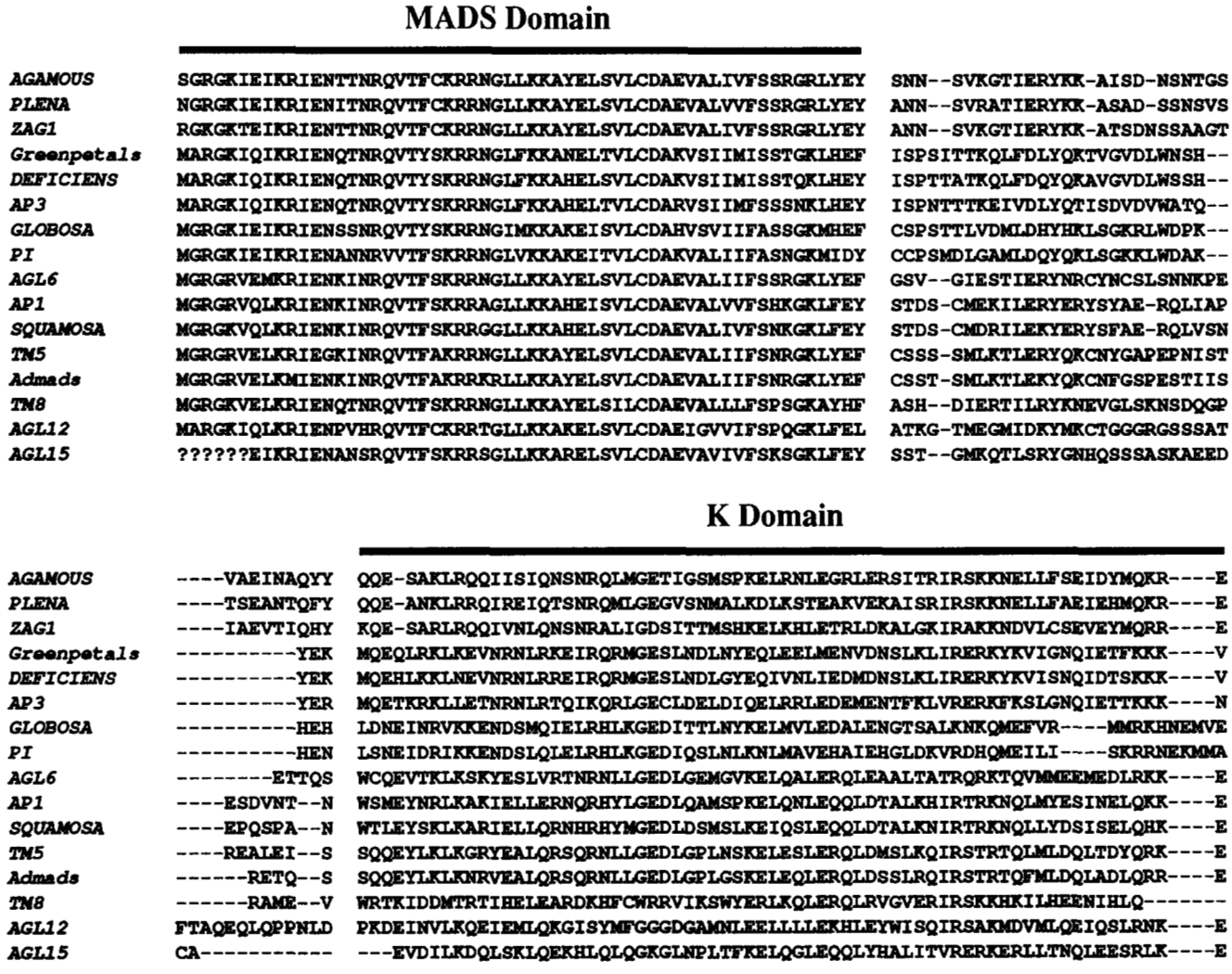

FIGURE 2.—Alignment of predicted protein sequences of the MIK region from several members of the plant MADS-box gene family. The conserved MADS-and K-domains are indicated. The variable I region is found between these two domains. Dashes denote gaps. The nucleotide sequence encoding this region (the MIK region) used in the phylogenetic analyses.

position of *TM8* has moderate bootstrap support (63%) in the parsimony analysis, although the support in the neighbor-joining tree is weak. The early divergence of the *TM8* lineage was also observed in other analyses using a smaller data set (DOYLE 1994). Like other orphan genes, the orthologues to *TM8* have not yet been characterized from other angiosperm species.

Rates of nucleotide substitutions between MADS-box genes: To compare the patterns of evolution between different plant nuclear genes, and between different gene domains, the rates of nucleotide substitutions for the plant MADS-box genes were calculated. The rate calculations are based on the levels of nucleotide substitutions in two independent sets of pairwise comparisons: the monocot *Zea ZAG1* and *ZAG2* genes *vs.* their dicot orthologues and the monocot orchid *Admads* gene with the dicot *AGL9*, *FBP2* and *TM5* loci (see Figure 4). The calculated substitution levels are shown on Tables 1 and 2. These pairwise comparisons provide separate estimates of levels of nucleotide divergence between monocot and dicot genes. The mean between the two estimates is an approximate value for the nucleotide substitution rate for the plant MADS-box gene family.

Nucleotide sequence distances were calculated for the entire gene, as well as separately for the MADS-box, K-box and C-terminal region. We have also calculated the nucleotide substitution levels for the region encompassing the MADS- and K-boxes (MIK region), which is the sequence used in reconstructing the evolutionary history of the MADS-box gene family (see Figure 1). Because of the relatively small size of the I region, this sequence was not separately analyzed. The nucleotide distance estimates were calculated using the Tajima-Nei method; levels of nonsynonymous substitutions per nonsynonymous site were also calculated (NEI and GOJOBORI 1986; LI 1993). To estimate the nucleotide substitution rates, a monocot/dicot divergence date of 135 mya was utilized. This date agrees quite well with angiosperm fossil evidence, which places the origins of the

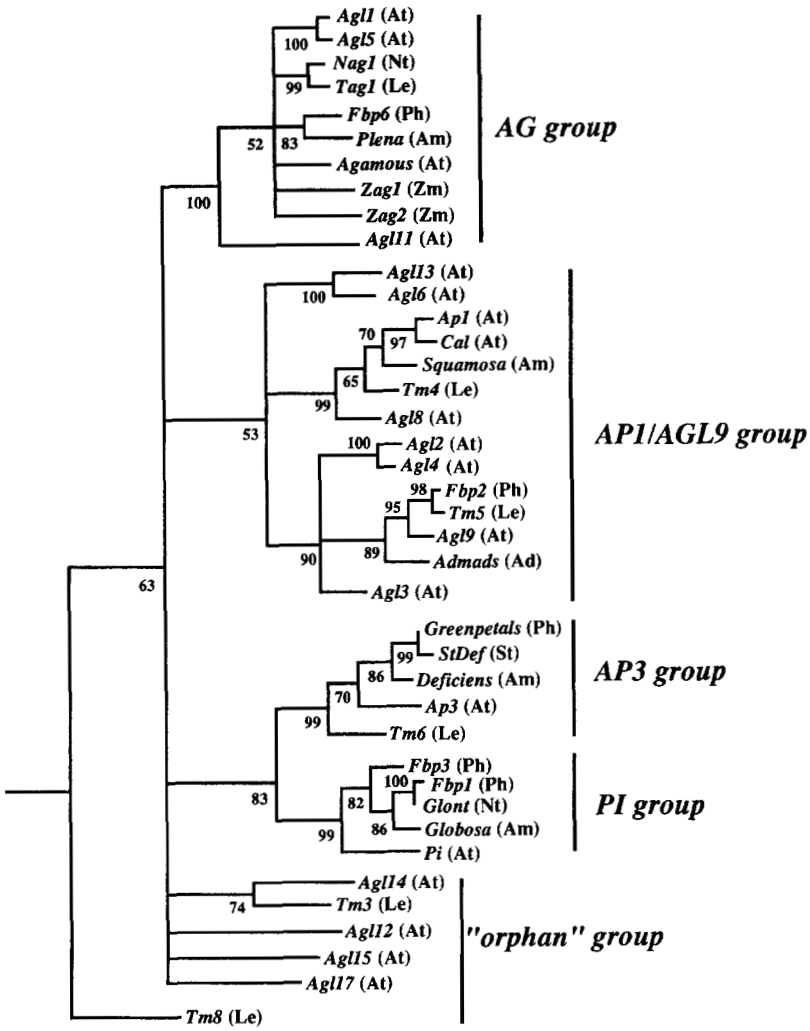

FIGURE 3.—Phylogenetic tree of the plant MADS-box gene family using parsimony analyses. The different gene groups are indicated on the right. A total of 507 nucleotide sites were used in the analysis. The numbers next to the nodes give bootstrap values from 100 replicates. The tree length is 1841 steps. Nodes with <50% bootstrap support are collapsed. The species origins for the different genes are indicated in the brackets: At, *Arabidopsis thaliana*; Am, *Antirrhinum majus*; Ph, *Petunia hybrida*; St, *Solanum tuberosum*; Nt, *Nicotiana tabacum*; Zm, *Zea mays*; Ad, *Aranda deborah*.

flowering plants in the early Cretaceous (DOYLE 1973). However, cpDNA sequence analysis has produced an estimate for the monocot/dicot divergence date as early as 200 mya (WOLFE *et al.* 1989). In calculating the rates for the MIK region, we have also used this early divergence date to provide an alternative substitution rate estimate.

All the rate estimates are shown in Table 3. The substitution rates for the entire gene using all nucleotide positions (2.5×10^{-9} substitutions/site/year) are somewhat higher than the rate when only the first two codon positions are utilized (1.5×10^{-9} substitutions/site/year). Because changes in the third codon positions result in synonymous changes and are less likely to be constrained by purifying selection, higher overall substitution rates are to be expected when third codon positions are used in the rate calculations. Indeed, the Tajima-Nei rate estimates for the first and second codons positions are significantly closer to the nonsynonymous substitution rates, because most changes in these two positions result in alterations in the amino acid sequence of the encoded protein.

The average DNA substitution rate estimate for the entire gene (2.5×10^{-9} substitutions/site/year) is half the average rate of neutral substitutions (5×10^{-9} substitutions/site/year) determined for plant nuclear genes (WOLFE *et al.* 1987). The rate of nonsynonymous substitutions/site/year is $\sim 1.4 \times 10^{-9}$ nonsynonymous substitutions/site/year, which is higher than the average rate of 0.87×10^{-9} nonsynonymous susbtitutions/site/year calculated for a large number of genes (LI *et al.* 1985). By comparison, the nonsynonymous nucleotide susbtitution rate for another plant regulatory gene, the *R* basic-helix-loop-helix protein gene, is 2.6×10^{-9} substitutions/site/year (PURUGGANAN and WESSLER 1994). This suggests that while the plant MADS-box genes are evolving more rapidly than typical eukaryotic loci, their sequences are changing more slowly than other regulatory genes that have thus far been analysed (see also TUCKER and LUNDRIGAN 1993).

Evolution of different domains in a plant regulatory gene: Differences in the levels of sequence constraint between various regions of a gene are reflected in disparities in substitution rates (LI *et al.* 1985). Analysis of the plant MADS-box gene families reveals significant differences in the rates of nonsynonymous nucleotide

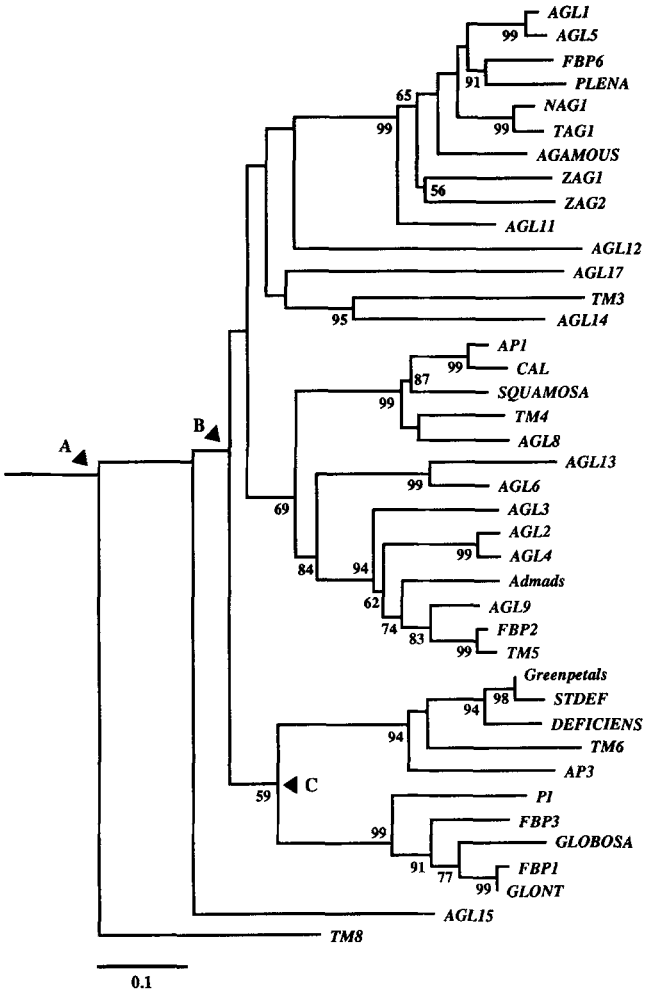

FIGURE 4.—Phylogenetic tree of the plant MADS-box gene family using the neighbor-joining method. The Tajima-Nei sequence distance was used to construct the tree. The numbers next to the nodes give bootstrap percentages after 500 replicates. Except for minor changes at the tips, the topology of this tree is concordant with the tree derived from parsimony analysis (Figure 3). The important nodes are labeled on the tree. Node A represents the earliest divergence event in the evolution of the plant MADS-box gene family. Nodes B and C represent the first and last divergence events in the establishment of the known floral homeotic gene lineages.

substitutions in different regions of the gene. Of the three large domains within these genes, the MADS-box, which encodes the putative DNA-binding domain, has the lowest substitution rate. The nonsynonymous substitution rate for this region is $\sim 3 \times 10^{-10}$ substitutions/site/year. The rate for the K-box is more than three times as large, with a rate of 1×10^{-9} substitutions/site/year. The rate for the K-box is comparable to the rate for the conserved region of the plant *R* regulatory gene family, whose conserved domain rate is 1.2×10^{-9} substitutions/site/year (PURUGGANAN and WESSLER 1994).

The C-terminal domain displays the greatest amount of protein sequence alterations between genes, with an estimated nonsynonymous substitution rate of 7.9×10^{-9} substitutions/site/year. The error on this estimate, however, is fairly large because only two pairwise comparisons were available for calculating this rate. The Tajima-Nei substitution rate for the first two codon positions (3.7×10^{-9} substitutions/site/year) may more accurately reflect the nonsynonymous substitution rate for the C-terminal domain. This rate estimate is 10 times greater than the rate for the MADS-box and is comparable to the estimated mean evolutionary rate of 4.7×10^{-9} substitutions/site/year for synonymous substitutions (LI *et al.* 1985). This suggests that the C-terminal region may be evolving at a near-neutral rate. The ratio of nonsynonymous to synonymous substitutions for this region, however, indicates that the C-terminal sequence is subject to a considerable degree of purifying selection (see below).

Different domains can experience differing levels of selection at various times, leading to variations in the relative rates of nonsynonymous substitutions between distinct gene regions when homologous loci are compared (HUGHES 1993). Comparing levels of nonsynonymous substitutions provides an indication of the functional divergence of specific regions of a locus. For our analysis of the plant MADS-box genes, the proportion of nonsynonymous differences (p_n) was computed in pairwise comparisons of different genes between groups (NEI and GOJOBORI 1986). The p_n values were calculated separately for the MADS-box, K-box and C-terminal region. The proportion of nonsynonymous differences was used because the distant comparisons in the analysis make it difficult to compute nonsynonymous substitution levels, particularly for the highly divergent C-terminal domain. The mean p_n was computed between gene groups, and the relative rates of evolution among the different domains were compared (HUGHES 1993). In these calculations, the AP3 and PI groups are considered separately. The data reveal that relatively little nonsynonymous differences occur between MADS-boxes (see Table 4). It appears that some of the amino acid sequence divergence between these different plant MADS-box genes occur through the diversification of the K-box and C-terminal domain between gene groups (see Table 4). In intergroup comparisons, the K-box and C-terminal domains show a 2–3 times greater mean p_n than the MADS-box.

The degree of sequence constraint in various regions can also be evaluated by determining the ratio of nonsynonymous to synonymous substitutions in pairwise comparisons. Because the synonymous site substitutions are expected to be neutral (or nearly neutral), this ratio corrects for possible rate alterations that result from variations in mutation rate (LI *et al.* 1985). Differences in sequence constraint appear to reflect changes in the mode and levels of selective forces that act on sequences. To study levels of sequence constraint across the MADS-box genes, pairwise comparisons were undertaken between closely related genes within each of the

TABLE 1

Tajima-Nei distance estimates for the MIK region used in calculating nucleotide substitution rates

Genes	*AGL1*	*AGL5*	*NAG1*	*TAG1*	*AG*	*FBP6*	*PLE*	*AGL9*	*TM5*	*FBP2*
Entire gene										
ZAG1	0.512	0.490	0.423	0.458	0.517	0.427	0.459	n.a.	n.a.	n.a.
ZAG2	0.448	0.471	0.448	0.496	0.485	0.485	0.485	n.a.	n.a.	n.a
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.325	0.342	0.335
MADS box										
ZAG1	0.089	0.090	0.101	0.132	0.070	0.071	0.101	n.a.	n.a.	n.a.
ZAG2	0.063	0.075	0.087	0.111	0.52	0.042	0.075	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.072	0.081	0.720
K box										
ZAG1	0.441	0.443	0.339	0.382	0.342	0.385	0.432	n.a.	n.a.	n.a.
ZAG2	0.319	0.370	0.364	0.416	0.374	0.467	0.473	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.200	0.201	0.213
C-terminal region										
ZAG1	1.212	1.280	0.846	0.873	1.017	0.867	1.001	n.a.	n.a.	n.a.
ZAG2	1.092	1.141	1.115	1.241	1.209	1.199	1.065	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.972	0.996	0.839
MIK region										
ZAG1	0.287	0.279	0.249	0.276	0.228	0.255	0.284	n.a.	n.a.	n.a.
ZAG2	0.260	0.285	0.266	0.297	0.270	0.300	0.324	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.182	0.186	0.192

Rates are for first and second codon positions only; n.a., not applicable.

three major gene groups (Table 5); more distant comparisons could not be made because saturation of synonymous sites occurs even between monocot and dicot genes.

The results indicate that the different gene groups have similar levels of sequence constraint, with nonsynonymous/synonymous substitution ratios ranging from 0.119 to 0.185. These values are approximately equivalent to the average value of 0.14 for various monocot and dicot plant genes (MARTIN *et al.* 1989; HUANG *et al.* 1992) and 0.189 for a collection of 42 mammalian genes (LI *et al.* 1985). The sequence of members of the PI group, however, appear to be somewhat less constrained than other floral homeotic gene groups. Moreover, the various gene regions display significant differences in their levels of nonsynonymous/synonymous

TABLE 2

Estimates of levels of nonsynonymous substitutions in the MIK region used in calculating nonsynonymous substitution rates

Genes	*AGL1*	*AGL5*	*NAG1*	*TAG1*	*AG*	*FBP6*	*PLE*	*AGL9*	*TM5*	*FBP2*
Entire gene										
ZAG1	0.451	0.433	0.374	0.399	0.480	0.399	0.431	n.a.	n.a.	n.a.
ZAG2	0.424	0.436	0.441	0.463	0.464	0.462	0.456	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.308	0.314	0.312
MADS box										
ZAG1	0.075	0.081	0.081	0.099	0.065	0.053	0.089	n.a.	n.a.	n.a.
ZAG2	0.060	0.065	0.079	0.099	0.050	0.036	0.065	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.064	0.072	0.052
K box										
ZAG1	0.376	0.371	0.314	0.322	0.281	0.363	0.417	n.a.	n.a.	n.a.
ZAG2	0.319	0.346	0.377	0.381	0.331	0.448	0.430	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.150	0.155	0.168
C-terminal region										
ZAG1	–	–	–	–	2.770	–	1.518	n.a.	n.a.	n.a.
ZAG2	–	–	–	–	–	–	–	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	–	–	–
MIK region										
ZAG1	0.251	0.244	0.221	0.236	0.203	0.240	0.269	n.a.	n.a.	n.a.
ZAG2	0.254	0.267	0.268	0.278	0.253	0.295	0.298	n.a.	n.a.	n.a.
AdMads	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.151	0.154	0.156

n.a., not applicable; –, no substitution level estimate available.

TABLE 3

Rates of nucleotide substitutions for MADS-box genes in plants

Domains	Tajima-Nei substitution rate		Nonsynonymous substitution rate
	All positions	1st and 2nd positions	
Entire gene	2.5 ± 0.2	1.5 ± 0.2	1.4 ± 0.2
MADS box	1.1 ± 0.2	0.3 ± 0.1	0.3 ± 0.1
K box	1.9 ± 0.3	1.1 ± 0.2	1.0 ± 0.2
C-terminal region	4.6 ± 0.7	3.7 ± 0.7	7.9 ± 6.0
MIK region	1.6 ± 0.1	0.9 ± 0.1[a]	0.8 ± 0.1
MIK region[b]	1.1 ± 0.1	0.6 ± 0.1	0.5 ± 0.1

Nucleotide substitution rates are given as $\times 10^9$ substitution/site/year, based on a monocot/dicot divergence time of 135 mya. Values are means ± SD.

[b] Substitution rate used to calculate divergence times for major floral homeotic gene groups.

[a] Alternative substitution rates for MIK region calibrated using a monocot/dicot divergence time estimate of 200 mya.

substitutions. Among the different gene regions, the MADS-box displays the highest level of constraint, with nonsynonymous/synonymous substitutions ratios ranging from 0.036 to 0.096. The K-box and the C-terminal region are considerably more variable, with sequence constraints ranging from 0.109 to 0.399. The constraint levels in these variable regions, however, are still considerably below the levels seen for some other regulatory genes. The ratio of nonsynonymous/synonymous substitutions for the variable domains of the plant *R* gene is close to 1.0 (PURUGGANAN and WESSLER 1994).

For both the AP3 and PI groups, the nonsynonymous/synonymous substitution ratios for the K-box and C-terminal regions are roughly equivalent. This indicates that, within these two gene groups, both the K-box and C-terminal regions are under similar levels of selective constraint. In the AG and AP1/AGL9 groups, however, there appears to be a marked increase in the nonsynonymous/synonymous substitution ratio for the C-terminal region when compared to the K-box, although this increase may simply be due to the presence of several paralogous genes within these groups. Interestingly, the nonsynonymous/synonymous ratios for the MADS-and K-boxes in the AP1/AGL9 genes are comparable, indicating that both these domains fix amino acid-replacing substitutions at about the same rate.

Estimated divergence times for the different floral gene groups: The phylogenetic relationships between the different plant MADS-box genes provide us with insights into the origins and early diversification of this regulatory gene family. Both the parsimony and neighbor-joining analysis suggests that the different floral homeotic gene groups were established fairly rapidly and relatively early in the evolution of land plants. The rapidity of lineage establishment is clearly evident in the neighbor-joining tree (Figure 4). Node B on the tree marks the start of diversification of the major gene groups, while node C represents the point at which the separate AP3 and PI lineages begin. This internodal length (node B to C) represents ~11% of the entire evolutionary history of the plant MADS-box genes (node A to tree tips), which indicates that the lineages for each of the floral homeotic gene groups were all established in a relatively brief evolutionary time span.

When did the evolutionary diversification of the floral homeotic gene groups occur? The divergence time for these floral homeotic gene lineages can be approxi-

TABLE 4

Mean proportion of nonsynonymous differences (p_n) in different gene regions between major floral homeotic gene groups

Comparison	MADS box	K box	C region
AG group			
vs. AP3 group	0.24 ± 0.04	0.50 ± 0.04	0.71 ± 0.04
vs. PI group	0.25 ± 0.04	0.47 ± 0.04	0.71 ± 0.04
vs. AP1/AGL9 group	0.18 ± 0.04	0.46 ± 0.04	0.68 ± 0.04
AP3 group			
vs. PI group	0.19 ± 0.03	0.51 ± 0.04	0.63 ± 0.04
vs. AP1/AGL9 group	0.23 ± 0.04	0.52 ± 0.04	0.68 ± 0.04
PI group			
vs. AP1/AGL9 group	0.23 ± 0.04	0.50 ± 0.04	0.71 ± 0.04

Values are means ± SD.

TABLE 5

Ratio of nonsynonymous to synonymous substitutions in different gene regions within major floral homeotic gene groups

Comparison	All	MADS box	K box	C region
AG	0.134 ± 0.007	0.036 ± 0.003	0.143 ± 0.023	0.286 ± 0.031
AP3	0.119 ± 0.026	0.043 ± 0.009	0.172 ± 0.047	0.183 ± 0.028
PI	0.185 ± 0.016	0.096 ± 0.014	0.268 ± 0.027	0.225 ± 0.029
AP1/AGL9	0.149 ± 0.022	0.071 ± 0.034	0.109 ± 0.023	0.399 ± 0.070

Values are means ± SD.

mated from the calculated mean substitution rate (Table 3) and the mean levels of nucleotide substitution between the different groups. The MIK region was used in calculating this divergence time estimate, since this region can still be reasonably aligned between distant genes. Furthermore, the MIK sequences between the monocot and dicot lineages used in these analyses appear to be evolving at equivalent rates ($P > 0.999$). To minimize the effects of synonymous site saturation, only the substitution rate using the first two codon positions were considered.

The mean level of nucleotide substitution between gene groups is 0.6105. Differences in the nucleotide susbstitution levels between the basal *TM8* gene and the various gene groups are not statistically significant in a *t*-test ($P > 0.10$). Using a substitution rate of 9×10^{-10} substitutions/site/year, we estimate that the approximate time in which the floral homeotic gene diversification event occurred was 340 mya. This crude estimate relies on the assumption that the substitution rate within this sequence has not substantially altered during the evolutionary history of these genes. This divergence time does suggest that the last common ancestor of the floral homeotic genes existed much earlier than the monocot/dicot divergence, which is concordant with the observed relationships between the different groups (Figures 3 and 4).

DISCUSSION

Morphological development in eukaryotes is controlled by sets of regulatory genes that interact to direct pattern formation. Genetic studies have shown that many of these genes are organized into regulatory gene families that are related to each other by descent from a common ancestral gene (DICKINSON 1991; KAPPEN *et al.* 1993). The isolation of many of these developmental loci from evolutionarily distant and morphologically distinct taxa allows us to study the conservation of developmental functions and provides us with opportunities to analyze the underlying genetic architecture of morphological diversification.

The plant MADS-box genes comprise a fairly large gene family, many of whose members regulate various aspects of flower development (COEN 1991; DAVIES and SCHWARZ-SOMMER 1994; M. YANOFSKY, unpublished data). A total of 40 reported plant MADS-box sequences from six dicot and two monocot species were used in these analyses. The largest species sample comes from studies with *A. thaliana* (Brassicaceae), in which 19 MADS-box genes have been sequenced.

Phylogenetic analysis reveals that most of these MADS-box genes are organized into distinct monophyletic gene groups whose members share common functional roles in floral development. These groups are roughly equivalent to clades identified in a smaller analysis of this gene family (DOYLE 1994). Three major MADS-box gene groups were established early during the evolution of this plant multigene family. The AGAMOUS group, members of which include the Arabidopsis *AG* and Antirrhinum *PLENA* genes, is one of these gene clades. Genetic and molecular analysis demonstrates that these two floral homeotic genes are required for stamen and carpel development in the inner two whorls (YANOFSKY *et al.* 1990; BRADLEY *et al.* 1993). Most of the other genes in this group also show expression in carpels and/or stamens, suggesting that these genes play related functional roles in different species (MA *et al.* 1991; PNUELI *et al.* 1991; BRADLEY *et al.* 1993; SCHMIDT *et al.* 1993).

The AP3/PI group includes the Arabidopsis *APETALA3* and *PISTILLATA* genes and the Antirrhinum *DEFICIENS* and *GLOBOSA* loci. These genes are expressed in whorls 2 and 3 of the Arabidopsis and Antirrhinum flowers, and are known to be necessary for petal and stamen development (SOMMER *et al.* 1990; JACK *et al.* 1992; TROBNER *et al.* 1992). The two major clades within this lineage define paralogous genes that are both required for activity. These paralogous loci may interact with each other to control stamen and petal development; the Antirrhinum DEF and GLO (SCHWARZ-SOMMER *et al.* 1992) as well as the Arabidopsis AP3 and PI (GOTO and MEYEROWITZ 1994) proteins, for example, have been shown to associate with each other to form heterodimers.

The AP1/AGL9 group is also composed of distinct gene clades. The AP1 clade includes the Arabidopsis *AP1* and *CAL* genes as well as the Antirrhinum *SQUAMOSA* locus. These genes have been shown to specify the identity of the floral meristem and to regulate the development of the perianth organs (MANDEL *et al.*

1992; KEMPIN *et al.* 1995). The function of the genes in the AGL9 and AGL6/AGL13 clades remains less clear. Both *TM5* (in the AGL9 clade) and *AGL6* show floral-specific expression (MA *et al.* 1991; PNUELI *et al.* 1991), consistent with possible roles for these genes in flower development. Interestingly, inactivation of the Petunia *FBP2* and Lycopersicon *TM5* genes results in both meristem and organ identity defects (ANGENENT *et al.* 1994; PNUELI *et al.* 1994), and it has been suggested that these two loci are candidates for genes that mediate between the meristem and organ identity functions (M. YANOFSKY, unpublished data). Two additional Arabidopsis genes, *AGL2* and *AGL4*, also belong to the AGL9 clade. These genes, like *AGL9*, are expressed early in flower development and may also act to mediate between the early- and late-acting floral regulatory genes (B. SAVIDGE, S. ROUNSLEY and M. YANOFSKY, unpublished results).

The rest of the other MADS-box genes do not appear to belong to any of these three major evolutionarily distinct groups. Intriguingly, several of these genes, including the Lycopersicon *TM3* (PNUELI *et al.* 1991) and Arabidopsis *AGL12*, *AGL14*, *AGL15* and *AGL17* (S. ROUNSLEY, G. DITTA and M. YANOFSKY, unpublished results), are expressed in the vegetative tissues of the plant. The *AGL12*, *AGL14* and *AGL17* genes are preferentially expressed in roots, while *AGL15* is expressed in the Arabidopsis embryo. The RNA accumulation patterns of these orphan genes suggest that these loci are members of gene groups involved in other, extrafloral aspects of plant development.

Functional divergence between regulatory genes appears to arise from localized evolution of different gene domains. This is particularly evident in regulatory genes, where both rapid and slow evolving sequences are frequently juxtaposed (TUCKER and LUNDRIGAN 1993; WHITFIELD *et al.* 1993; PURUGGANAN and WESSLER 1994). It has been suggested that the heterogeneity in substitution rates may play a role in the evolution of novel developmental functions within a regulatory gene family (PURUGGANAN and WESSLER 1994). Like many regulatory loci, the plant MADS-box genes display significant variation in levels of sequence similarity across their length. Among the different regions of these floral regulatory genes, the MADS-box exhibits the highest level of sequence conservation. The rates of nonsynonymous nucleotide substitution within this sequence is significantly slower than the rate for other regions and is well below the mean rate of sequence evolution for other genes. The ratio of nonsynonymous to synonymous substitutions in the MADS-box also shows that this DNA-binding region is under severe sequence constraint, with less than one nonsynonymous substitution occuring for every 10 synonymous ones (see Table 4). The MADS-box sequence, however, does show several amino acid replacements, particularly between gene groups. Despite the high sequence constraint for this domain within gene groups, it is possible that changes in this region contribute to the diverse functions of the different plant MADS-box loci.

Both the K-box and C-terminal regions, however, are evolving at markedly faster rates. The rates of evolution of these domains differ among various gene groups. The ratio of nonsynonymous/synonymous substitutions for the K-box of the AP1/AGL9 genes, for example, is comparable to that of the MADS-box, suggesting that the AP1/AGL9 K-box is under greater levels of sequence constraint than K-boxes in other gene groups. The C-terminal domain also appears to be evolving more rapidly in members of the AGAMOUS and AP1/AGL9 groups, possibly due to functional divergence between paralogous MADS-box loci within these gene groups. Interestingly, the ratios of nonsynonymous to synonymous substitutions for both the K-box and C-terminal domains are significantly lower than those observed for the variable regions of other regulatory genes. Moreover, the relatively high levels of nonsynonymous substitutions in these regions do not appear to reflect the dispensability of these sequences. In several floral homeotic loci, mutant phenotypes are associated with mutational lesions that alter or eliminate these 3′ sequences (JACK *et al.* 1992; DAVIES and SCHWARZ-SOMMER 1994; A. MANDEL and M. YANOFSKY, unpublished results). The highly divergent sequences in the C-terminal regions between floral homeotic gene groups, and to a lesser extent the K-domains, suggest that changes in these domains are also responsible for the functional divergence of these homeotic loci.

The phylogeny of the plant MADS-box loci allows us to reconstruct the early evolutionary history of those genes that are known to control pattern formation in the angiosperm flower. It appears that the major floral homeotic gene lineages were established fairly rapidly, in a span of time representing approximately one-ninth of the entire evolutionary history of the plant MADS-box gene family. This may indicate that the functions encoded by genes in the AGAMOUS, AP3, PI and AP1/AGL9 groups were also fixed in a relatively short evolutionary period. These lineages were certainly well established by the time the monocotyledenous and dicotyledenous angiosperms diverged ~135 mya. It is difficult to estimate with any precision the time of divergence for the major floral homeotic gene clades, especially because the nucleotide substitution rates for these genes may have changed significantly during their evolutionary history. An approximate substitution rate estimates for the MADS- box, I region and K-box, however, suggests that these floral homeotic gene lineages apparently split from each other sometime in the mid- to late-Paleozoic 340 mya.

This date represents a particularly dynamic time in the evolution of the vascular flora. The primitive trimerophytes present in the Devonian landscape were giving way to transitional ancestors of the seed plants,

including a group known as the progymnosperms (STEWART 1983). The progymnosperms, which exhibited characteristics intermediate between the now extinct trimerophytes and the spermatophytes, eventually gave rise to the lineage leading to the cycads, gingkos, conifers, gnetophytes and, eventually, the angiosperms (STEWART 1983). If our estimated divergence time is correct, then the establishment of these floral homeotic gene lineages predated the evolution of the flowering plants, and members of these gene groups should be present not only in flowering plants but in the cone-bearing gymnosperms as well.

Could these homeotic genes also exist in other plant species? The uncertainty surrounding the calculated rate estimates leaves open the possibility that these major MADS-box gene groups may have sprung into existence even earlier. The mid-Paleozoic era was also the time when the ferns began to diverge from the line leading to the seed plants (STEWART 1983); this split may well have occurred after the diversification of the plant MADS-box genes. Moreover, if the monocot/dicot divergence occurred as far back as 200 mya, as argued by WOLFE *et al.* (1989), then the floral homeotic gene diversification event may have occurred considerably earlier. Indeed, using this alternate date to calibrate our rate estimates, we arrive at a date of *ca.* 500 mya for the divergence of these homeotic gene groups, predating the time in the Ordovician when the first land plants began to evolve from their aquatic green algal ancestors. If this latter, very early date, is correct, then orthologues to the angiosperm floral homeotic genes should be present in all land plant species and perhaps even in the chlorophyte algal groups. Isolation of MADS-box homologs from various nonvascular plant and even algal groups may allow us to date the initial diversification event for the floral homeotic genes more precisely than the approximate times suggested by our molecular clock estimates.

Unlike the MADS-box gene family, flowers are a relatively recent evolutionary innovation among the vascular plants, appearing in the fossil record only in the Cretaceous period. It is probable that the direct ancestors of the floral homeotic genes participated in the reproductive development of nonflowering plants and were only later recruited to perform the critical task of controlling the development of the angiosperm flower. It is surprising that many of these gene lineages were not established sequentially over a long period of evolutionary time, but all arose in a relatively brief period. It is intriguing to think that this rapid diversification occured during the evolutionary experimentation that accompanied the rise of the seed plants during the late Paleozoic, including several gymnosperm groups (STEWART 1983). Our study on the mode and tempo of molecular evolution of these floral homeotic genes indicates that these regulatory loci continue to evolve at a relatively rapid pace, paralleling the continued morphological elaborations displayed between angiosperm flowers. As we continue the study of the plant MADS-box regulatory gene family, we should begin to understand the genetic bases behind the development of floral structures, and examine more intensively the relationship between molecular evolution and morphological diversification.

The authors would like to thank E. WATERS for helpful discussions, and E. A. KELLOGG, B. MORTON, L. MCDADE and M. FROHLICH and two anonymous reviewers for their critical comments on this manuscript. We are also grateful to M. A. MANDEL and H. MA for providing unpublished sequence information and W.-H. LI for providing a copy of LI93. This work was supported in part by an Alfred P. Sloan Foundation Postdoctoral Fellowship in Molecular Evolution to M.D.P., grants from the U.S. Department of Agriculture to R.J.S. and M.F.Y., and a grant from the National Science Foundation to M.F.Y. (DCB-9018749).

LITERATURE CITED

ANGENENT, G. C., M. BUSSCHER, J. FRANKEN, J. MOL and A. J. VAN TUNEN, 1992 Differential expression of two MADS-box genes in wild-type and mutant petunia flowers. Plant Cell **4:** 983–993.

ANGENENT, G. C., J. FRANKEN, M. BUSSCHER, D. WEISS and A. J. VAN TUNEN, 1994 Co-suppression of the petunia homeotic gene *fbp2* affects the identity of the generative mersitem. Plant J. **5:** 233–244.

BOWMAN, J. L., J. ALVAREZ, D. WEIGEL, E. M. MEYEROWITZ and D. SMYTH, 1993 Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. Development **119:** 721–743.

BRADLEY, D., R. CARPENTER, H. SOMMER, N. HARTLEY and E. COEN, 1993 Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of Antirrhinum. Cell **72:** 85–95.

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

COEN, E., and E. M. MEYEROWITZ, 1991 The war of the whorls: genetic interactions controlling flower development. Nature **353:** 31–37.

DAVIES, B., and Zs. SCHWARZ-SOMMER, 1994 Control of floral organ identity by homeotic MADS-box transcription factors. Results Probl. Cell Differ. **20:** 235–258.

DICKINSON, W., 1991 The evolution of regulatory genes and patterns in *Drosophila*. Evol. Biol. **25:** 127–174.

DOEBLEY, J., 1993 Genetics, development and plant evolution. Curr. Opin. Genetics Dev. **3:** 865–872.

DOYLE, J. A., 1973 Fossil evidence on early evolution of the monocotyledons. Q. Rev. Biol. **48:** 339–413.

DOYLE, J., 1994 Evolution of a plant homeotic multigene family: towards connecting molecular systematics and molecular developmental genetics. Syst. Biol. **43:** 307–328.

GARCIA-MAROTO, F., F. SALAMINI and W. ROHDE, 1993 Molecular cloning and expression patterns of the *Deficiens*-homologous gene *St-Deficiens* from *Solanum tuberosum*. Plant J. **4:** 771–780.

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

GOULD, S. J., 1977 *Ontogeny and Phylogeny*. Harvard University Press, Cambridge.

HANSEN, G., J. J. ESTRUCH, H. SOMMER and A. SPENA, 1993 *NTGLO*: tobacco homologue of the *GLOBOSA* floral homeotic gene of *Antirrhinum majus*, cDNA sequence and expression pattern. Mol. Gen. Genet. **239:** 310–312.

HUANG, N., G. L. STEBBINS and R. RODRIGUEZ, 1992 Classification and evolution of the alpha-amylase genes in plants. Proc. Natl. Acad. Sci. USA **89:** 7526–7530.

HUGHES, A. L., 1994 Phylogeny of the C3/C4/C5 complement-component gene family indicates that C5 diverged first. Mol. Biol. Evol. **11:** 417–425.

HUIJSER, P., J. KLEIN, W.-E. LONNIG, H. MEIJER, H. SAEDLER *et al.*, 1992 Bractomania, an inflorescence anomaly, is caused by the

loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. EMBO J. **11:** 1239–1249.

IRISH, V., and I. M. SUSSEX, 1990 Function of the *APETALA1* gene during Arabidopsis floral development. Plant Cell **2:** 741–753.

JACK, T., L. L. BROCKMANN 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.

KAPPEN, C., K. SCHUGHART and F. H. RUDDLE, 1993 Early evolutionary origin of major homeodomain sequence classes. Genomics **18:** 54–70.

KEMPIN, S., B. SAVIDGE and M. F. YANOFSKY, 1995 Molecular basis of the *cauliflower* phenotype in Arabidopsis. Science **267:** 522–525.

KUMAR, S., K, TAMURA and M. NEI, 1994 *Molecular Evolutionary Genetic Analysis Package 1.1*. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, State College, PA.

LEIFER, D., D. KRAINC, Y. YU, J. MCDERMOTT, R. E. BREITBART *et al.*, 1993 *MEF2C*, a MADS/MEF2-family transcription factor expressed in a laminar distribution in cerebral cortex. Proc. Natl. Acad. Sci. USA **90:** 1546–1550.

LI, W. H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

LI, W. H., C. I. WU and C. L. LUO, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2:** 150–174.

LU, Z. X., M. WU, C. S. LOH, C. Y. YEONG and C. J. GOH, 1993 Nucelotide sequence of a flower-specific MADS box cDNA clone from orchid. Plant Mol. Biol. **23:** 901–904.

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.

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

MARTIN, W., A. GIERL and H. SAEDLER, 1989 Molecular evidence for pre-Cretaceous angiosperm origins. Nature **339:** 46–48.

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

PNUELI, L., M. ABU-ABEID, D. ZAMIR, W. NACKEN, Zs. SCHWARZ-SOMMER *et al.*, 1991 The MADS box gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from *Antirrhinum* and *Arabidopsis*. Plant J. **1:** 255–266.

PNUELI, L., D. HAREVEN, L. BRODAY, C. HURWITZ and E. LIFSCHITZ, 1994b The *TM5* MADS-box gene mediates organ differentiation in the three inner whorls of tomato flowers. Plant Cell **6:** 175–186.

PNUELI, L., D. HAREVEN, S. ROUNSLEY, M. YANOFSKY and E. LIFSCHITZ, 1994a Isolation of the tomato *A6AMOUS* gene *TAG1* and analysis of its homeotic role in transgenic plants. Plant Cell **6:** 163–173.

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., and S. R. WESSLER, 1994 Molecular evolution of the plant *R* regulatory gene family. Genetics **138:** 849–854.

RAFF, R., and T. KAUFMANN, 1983 *Embryos, Genes and Evolution*. Indiana University Press, Bloomington, IN.

RICHARDS, A. J., 1986 *Plant Breeding Systems*. Allen and Unwin, Winchester, MA.

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

SCHMIDT, R. J., B. VEIT, M. A. MANDEL, M. MENA, S. HAKE *et al.*, 1993 Identification and molecular characterization of *ZAG1*, the maize homolog of the *Arabidopsis* floral homeotic gene *AGAMOUS*. Plant Cell **5:** 729–737.

SCHWARZ-SOMMER, Zs., I. HUE, P. HUIJSER, P. J. FLOR, R. HANSEN *et al.*, 1992 Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: Evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J. **11:** 251–263.

SIDOW, A., 1992 Diversification of the *Wnt* gene family in the ancestral lineage of vertebrates. Proc. Natl. Acad. Sci. USA **89:** 5098–5102.

SOMMER, H., J. P. BELTRAN, P. HUIJSER, H. PAPE, W. LONNIG *et al.*, 1990 *Deficiens*, a homeotic gene involved in the control of floral morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. EMBO J. **9:** 605–613.

STEWART, W., 1983 *Paleobotany and the Evolution of Plants*. Cambridge University Press, Cambridge, UK.

SWOFFORD, D., 1993 *Phylogenetic Analysis using Parsimony*. Illinois Natural History Survey, Champaign, IL.

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

TROBNER, W., L. RAMIREZ, P. MOTTE, I. HUE, P. HUIJSER *et al.*, 1992 *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. EMBO J. **11:** 4693–4704.

TUCKER, P. K., and B. LUNDRIGAN, 1993 Rapid evolution of the sex-determinig loci in Old World mice and rats. Nature **364:** 715–717.

WEIGEL, D., J. ALVAREZ, D. R. SMYTH, M. F. YANOFSKY and E. M. MEYEROWITZ, 1992 *LEAFY* controls floral mersitem identity in *Arabidopsis*. Cell **69:** 843–859.

WHITFIELD, L., R. LOVELL-BADGE and P. GOODFELLOW, 1993 Rapid sequence evolution of the sex- determining gene *SRY*. Nature **364:** 713–715.

WOLFE, K., M. GOUY, Y. YANG, P. SHARP and W. H. LI, 1989 Date of the monocot dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86:** 5201–5205.

WOLFE, K. H., W.-H. LI and P. M. SHARP, 1987 Rates of nucleotide susbtitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84:** 9054–9058.

YANOFSKY, M., H. MA, J. L. BOWMAN, G. N. DREWS and E. MEYEROWITZ, 1990 The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature **346:** 35–39.

Communicating editor: W.-H. LI