Duplication and Diversification in the *APETALA1/FRUITFULL* Floral Homeotic Gene Lineage: Implications for the Evolution of Floral Development

Amy Litt[1] and Vivian F. Irish

Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520-4108

Manuscript received March 21, 2003
Accepted for publication June 11, 2003

ABSTRACT

Phylogenetic analyses of angiosperm MADS-box genes suggest that this gene family has undergone multiple duplication events followed by sequence divergence. To determine when such events have taken place and to understand the relationships of particular MADS-box gene lineages, we have identified *APETALA1/FRUITFULL*-like MADS-box genes from a variety of angiosperm species. Our phylogenetic analyses show two gene clades within the core eudicots, eu*AP1* (including Arabidopsis *APETALA1* and Antirrhinum *SQUAMOSA*) and eu*FUL* (including Arabidopsis *FRUITFULL*). Non-core eudicot species have only sequences similar to eu*FUL* genes (*FUL*-like). The predicted protein products of eu*FUL* and *FUL*-like genes share a conserved C-terminal motif. In contrast, predicted products of members of the eu*AP1* gene clade contain a different C terminus that includes an acidic transcription activation domain and a farnesylation signal. Sequence analyses indicate that the euAP1 amino acid motifs may have arisen via a translational frameshift from the euFUL/FUL-like motif. The eu*AP1* gene clade includes key regulators of floral development that have been implicated in the specification of perianth identity. However, the presence of eu*AP1* genes only in core eudicots suggests that there may have been changes in mechanisms of floral development that are correlated with the fixation of floral structure seen in this clade.

THE products of MADS-box genes have been implicated in the regulation of a variety of plant developmental mechanisms and have been shown to be particularly important in the specification and development of the angiosperm flower (Coen and Meyerowitz 1991; Angenent *et al.* 1995; Rounsley *et al.* 1995; Alvarez-Buylla *et al.* 2000a; Ferrandiz *et al.* 2000). In *Arabidopsis thaliana* and other core eudicot species, MADS-domain-containing proteins are required for the proper transition from an inflorescence meristem to a floral meristem and for the correct specification of the identity of the four types of floral organs. The specification of floral organ identity has been codified in the ABC model (Coen and Meyerowitz 1991), which postulates three gene functions, A, B, and C, that act in overlapping concentric domains of the meristem to specify the floral organs. According to this model, based on work in the two model species *A. thaliana* and *Antirrhinum majus*, A-function specifies sepal identity in the outer domain of the meristem, A + B specifies petal, B + C specifies stamen (male reproductive organs), and C-function specifies carpel identity (female reproductive organs) in the innermost domain. Nearly all of the A-, B-, and C-function genes belong to the MADS-box family. Thus, understanding how different floral morphologies and developmental mechanisms evolved requires a determination of how these genes may have changed during the course of angiosperm diversification.

The history of the MADS-box gene family in plants is characterized by duplication events and subsequent divergence. For instance, phylogenies of the MADS-box gene family show that two lineages, which include the Arabidopsis B-function genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), arose by duplication from a single ancestral gene lineage and that the A-, B-, and C-function lineages themselves are probably all products of duplication events (Doyle 1994; Purugganan *et al.* 1995; Tandre *et al.* 1995; Hasebe and Banks 1997; Kramer *et al.* 1998; Krogan and Ashton 2000; Theissen *et al.* 2000). In addition to these duplications that preceded or occurred in conjunction with the origin of the angiosperms, MADS-box gene lineage duplications have also occurred within individual angiosperm lineages (Kramer and Irish 1999; Lowman and Purugganan 1999). The frequency of these events suggests that any comparative study of MADS-box genes requires as its foundation a comprehensive gene phylogeny that can be used to identify gene clades and to determine orthology (relationship through speciation) and paralogy (relationship through duplication) of various genes and lineages. Such a phylogeny provides a basis for defining orthologous genes for comparison and thereby provides a framework for comparative studies of gene structure, expression, and function across the angiosperms. We undertook a phylogenetic analysis of the *APETALA1/FRUITFULL* (*AP1/FUL*) MADS-box gene lineage (also

Sequence data from this article have been deposited with the GenBank Data Library under accession nos. AY306138–AY306195.

[1]*Corresponding author:* MCDB, 266 Whitney Ave., Yale University, New Haven, CT 06520-8104. E-mail: amy.litt@yale.edu

called the *SQUA* lineage; *e.g.*, Krogan and Ashton 2000; Theissen *et al.* 2000), members of which have been identified as key regulators of floral development in several model species, to identify duplication and sequence divergence events that occurred during the history of this gene lineage across the angiosperms.

In Arabidopsis, severe *apetala1* mutants have sepals transformed into bract-like structures that subtend secondary flowers. Petals are absent. The inner two whorls of organs, the stamens and carpels, are essentially normal. This pattern may be repeated in the secondary flowers with the formation of tertiary nested floral structures (Irish and Sussex 1990; Bowman *et al.* 1993). On the basis of this phenotype *AP1* has been implicated in the specification of floral meristem identity as well as of sepal and petal identity (A-function; *e.g.*, Irish and Sussex 1990; Bowman *et al.* 1993; Weigel and Meyerowitz 1994).

The Arabidopsis genome contains two genes, *CAULIFLOWER* (*CAL*) and *FUL*, that are closely related to *AP1* and that share redundant functions for floral meristem specification. *CAL* has no phenotype on its own, but the *ap1 cal* double mutant shows an enhancement of the repeated branching pattern seen in *ap1* (Bowman *et al.* 1993; Kempin *et al.* 1995). An even more severe branching phenotype is seen in the Arabidopsis *ap1 cal ful* triple mutant, in which essentially all floral meristem character is lost and flowers are not formed (Ferrandiz *et al.* 2000). In contrast to *CAL*, *FUL* also has separate and nonredundant functions and is required for proper fruit and leaf development (Gu *et al.* 1998). In the Arabidopsis *ful* mutant, a lack of proper cell differentiation in the fruit walls abolishes fruit elongation, causing some fruits to rupture prematurely as the seeds develop. In addition, the cauline leaves of *ful* mutants are broader and rounder and have fewer cell layers than wild type.

To date little information is available regarding the function of members of the *AP1/FUL* gene family in other angiosperm species. *CAL* appears to be the result of a duplication specific to Brassicaceae (Purugganan 1997; Purugganan and Suddith 1998) and has been implicated in the cauliflower phenotype of *Brassica oleracea* (Kempin *et al.* 1995). Putative *FUL* orthologs are widespread throughout angiosperms, but their roles in other species have not yet been defined. The loss-of-function mutation of the Antirrhinum *AP1* ortholog, *SQUAMOSA* (*SQUA*), shows a more complete loss of floral meristem identity than *ap1* shows and rarely produces flowers, a phenotype similar to the *ap1 cal ful* triple mutant. Notably, when *squa* flowers are produced, the specification of organ identity is normal (Huijser *et al.* 1992). The proliferating inflorescence meristem phenotype in *Pisum sativum*, caused by a mutation in the *AP1* ortholog *PEAM4*, has been described as similar to that of *squa* mutants (Taylor *et al.* 2002). To date, *AP1* is the only gene in this lineage that has been shown to confer A-function in its native species.

We investigated the history of the *AP1/FUL* lineage by constructing a phylogeny that included sequences from a variety of angiosperm species. Previous analyses had suggested that *AP1* and *FUL* themselves belong to separate closely related gene clades that were the result of a duplication event that occurred sometime after the divergence of the monocot lineage (Hasebe and Banks 1997). The *AP1* and *FUL* clades have been included in a single gene family, generally called the *AP1* or *SQUA* family (*e.g.*, Southerton *et al.* 1998; Hasebe 1999; Theissen *et al.* 2000; Pelucchi *et al.* 2002), which has been further grouped with the *SEPALLATA* genes in the *AP1/AGL9* family (*e.g.*, Purugganan *et al.* 1995; Purugganan 1997; Buchner and Boutin 1998; Moon *et al.* 1999; Lawton-Rauh *et al.* 2000). *AP1* and *FUL* genes share significant sequence similarity, and thus the orthology of published genes is often difficult to ascertain. Our goal therefore was to determine to which group individual sequences belong in order to clarify orthology and paralogy and to provide a framework for gene comparisons. We also wanted to ascertain where, with respect to angiosperm phylogeny, the *AP1-FUL* duplication occurred. The results of our phylogenetic analysis indicate that there were several duplications in the evolution of the *AP1/FUL* gene family and that the *AP1-FUL* duplication is correlated with the diversification of the core eudicots and the concurrent fixation of floral structure. Sequence comparisons also identify conserved amino acid motifs that allow us to differentiate AP1-like and FUL-like sequences. These data allow us to formulate hypotheses regarding the evolution of floral developmental mechanisms across the angiosperms.

MATERIALS AND METHODS

Unique *AP1*- and *FUL*-like sequences available during the course of this study were identified by BLAST searches (Altschul *et al.* 1997) and were included in the analysis (see supplemental data at http://www.genetics.org/supplemental/ for accession numbers). *SEPALLATA*-like sequences and Arabidopsis *AGL6* and gymnosperm *DAL1*-like sequences, identified in published analyses as most closely related to the *AP1/FUL* lineage (Purugganan 1997; Hasebe 1999; Theissen *et al.* 2000), were included as outgroups.

New species were selected for inclusion in the analysis according to phylogenetic position (Figure 1) and availability of floral bud material. Species used and genes cloned from each are listed in Table 1. The species sampled include core eudicots as well as a variety of non-core eudicots and non-eudicots. Total RNA was extracted from ~1 g of floral buds of varied ages using the standard Trizol (Invitrogen, Carlsbad, CA) protocol. For *P. sativum*, the RNeasy kit (QIAGEN, Valencia, CA) was used; for *Heuchera americana* and *Corylopsis sinensis*, Concert Plant RNA reagent (Invitrogen) was used to eliminate starch coprecipitation. Poly(A)$^+$ RNA was isolated from total RNA using Magnetight particles (Novagen, Madison, WI). The purification procedure was performed twice on each RNA sample for cleaner separation of poly(A)$^+$ RNA. cDNA was synthesized using Superscript II (Invitrogen) according to the manufacturer's instructions.

Amplification of target genes was carried out in two stages

TABLE 1

Genes isolated in this study, listed by species

Species	Family/order	Gene name[a]	Accession no.
Michelia figo (banana shrub)	Magnoliaceae/Magnoliales	*MfAGL6A*	AY306157
		MfAGL6B	AY306158
		MfFL	AY306159
Peperomia caperata (emerald ripple peperomia)	Piperaceae/Magnoliales	*PcFL1*	AY306167
		PcFL2	AY306168
Allium sp. (onion)	Alliaceae/Asparagales	*AlFL*	AY306138
Tradescantia virginiana (spiderwort)	Commelinaceae/Poales	*TvFL1*	AY306190
		TvFL2	AY306191
		TvFL3	AY306192
		TvFL4	AY306193
		TvSEP3	AY306189
Ranunculus bulbosus (bulbous buttercup)	Ranunculaceae/Ranunculales	*RbAGL6*	AY306184
		RbFL1	AY306179
		RbFL2	AY306180
		RbFL3	AY306182
		RbFL4	AY306183
R. acris (common buttercup)	Ranunculaceae/Ranunculales	*RaFL*	AY306181
Papaver nudicaule (Iceland poppy)	Papaveraceae/Papaverales	*PapnSEP3*	AY306174
		PapnFL1	AY306175
		PapnFL2	AY306176
P. somniferum (opium poppy)	Papaveraceae/Papaverales	*PapsFL1*	AY306177
		PapsFL2	AY306178
Chelidonium majus (celandine)	Papaveraceae/Papaverales	*CmFL1*	AY306144
		CmFL2	AY306145
Pachysandra terminalis (pachysandra)	Buxaceae/Buxales	*PatSEP1*	AY306166
		PatFL1	AY306164
		PatFL2	AY306165
Phytolacca americana (pokeweed)	Phytolaccaceae/Caryophyllales	*PaFL1*	AY306161
		PaFL2	AY306162
		PaFUL	AY306163
		PaAP1	AY306160
Heuchera americana (coral bells)	Saxifragaceae/Saxifragales	*HeaSEP1*	AY306151
		HeaFL	AY306149
		HeaFUL	AY306150
		HeaAP1	AY306148
Corylopsis sinensis (Chinese winter hazel)	Hamamelidaceae/Saxifragales	*CsFUL*	AY306146
		CsAP1	AY306147
Clarkia concinna (pink ribbons)	Onagraceae/Myrtales	*CcFL*	AY306143
Pisum sativum (pea)	Fabaceae/Fabales	*PisFUL*	AY306169
Syringa vulgaris (lilac)	Oleaceae/Scrophulariales	*SvSEP1*	AY306187
		SvSEP3	AY306186
		SvAP1	AY306185
		SvAGL6	AY306188
Antirrhinum majus (snapdragon)	Scrophulariaceae/Scrophulariales	*AmSEP3A*	AY306140
		AmSEP3B	AY306141
		AmSEP3C	AY306142
		AmFUL	AY306139
Petunia hybrida (petunia)	Solanaceae/Solanales	*PhSEP1*	AY306173
		PhSEP3	AY306171
		PhFL	AY306170
		PhFUL	AY306172
Lycopersicon esculentum (tomato)	Solanaceae/Solanales	*LeSEP1*	AY306152
		LeSEP3	AY306153
		LeFUL1	AY306155
		LeFUL2	AY306156
		LeAP1	AY306154
Paeonia suffruticosa (peony)	Paeoniaceae/Saxifragales	*PsMDS2*	AY306195

[a] Genes are named with two to four letters (the first uppercase and the rest lowercase) denoting the species (*e.g.*, *Mf* for *M. figo*), followed by an abbreviation indicating the gene clade to which they belong according to the results of this analysis (*AGL6*, *AGAMOUS*-like6; *SEP1*, *SEPALLATA1*; *SEP3*, *SEPALLATA3*; *FL*, *FUL*-like; *FUL*, eu*FUL*; *AP1*, eu*AP1*. *PsMDS2* was cloned by Elena Kramer (Harvard University).

using a protocol designed to recover all possible genes belonging to the *AP1/FUL* gene lineage. First, a forward degenerate primer (AP1MDS3, GTNCARYTNARRMGNATNGARAAYAA GAT), designed to anneal to the MADS-box of *AP1*- and *FUL*-like sequences, was used with a poly(T) reverse primer [poly(T), GACTCGAGTCGACATCGA$(T)_{17}$V]. The reaction was run for 30–35 cycles with an annealing temperature of 42° on a GeneAmp 2400 thermocycler (Perkin-Elmer/Applied Biosystems, Norwalk, CT). Products with discrete bands of 500–1000 bp were cloned (TOPO-TA cloning kit, Invitrogen). In addition, the product of this first amplification reaction was diluted 1:25 and used as template in successive PCR reactions. These reactions used combinations of three nested forward primers (AP1MDS1, GCICWTGARMTNTCNRTNYTNTGYGATGC; AP1MDS2, TGG NYTKNTSAAGAARGCTCATGA; SQUA, TCWGTKCTTTGTGA TGCTGAAGT) and three nested reverse primers (AP1R2, ATA-SASTGGTTCCAGMGTWAGGTC; SQUAR, GCAAAGCATCCM AKATGGCATG; AGL8R, AGRTGRYKAASCATCCAIKGIGGCA) as well as the two primers used in the initial amplification reaction. These reactions were run for 30–40 cycles at an annealing temperature of 46°. All products showing a 500- to 1000-bp band on an agarose gel were cloned.

At least 50 clones were sequenced for most species. *AP1*-, *FUL*-, and *SEP*-like sequences from each species were aligned using GeneWorks (Oxford Molecular, Springfield, VA) or CLUSTALX (Thompson *et al.* 1994) to determine how many different sequences were present, and representative samples were sequenced on the reverse strand. Sequences from several species that were not exhaustively assessed were also included. These species include *Petunia hybrida*, *Lycopersicon esculentum*, *Clarkia concinna*, *P. sativum*, *Syringa vulgaris*, *Ranunculis acris*, and *Papaver nudicaule*. The following unpublished sequences were used with permission: *Lilium regale LrSQA, LrSQB* (A. Kanno, Tohoku University), *P. sativum PM9* (F. Madueno, Universidad Politecnica de Valencia-CSIC), and *Paeonia suffruticosa PsMADS9* (E. Kramer, Harvard University).

It was not possible to determine whether slightly different sequences represented alleles of a single gene or were in fact different genes; therefore the observed pattern of nucleotide variability was used to make this assessment. Groups of similar sequences were compared at variable sites; if a group could be divided into subgroups such that members of each subgroup shared the same nucleotide at each site, the subgroups were treated as separate genes and were all included in the analysis. If a group of similar sequences could not be so subdivided, and members showed nucleotide differences in a variable pattern across the gene, those sequences were taken to represent alleles of the same gene. In these cases a consensus sequence was used in the analysis.

Attempts to align nucleotide sequences produced inconsistent and significantly variable results. Amino acid sequences gave more reproducible results but contained insufficient information to produce well-resolved phylogenies. Therefore putative amino acid sequences were aligned in CLUSTALX and aa2dna (http://www.bio.psu.edu/People/Faculty/Nei/Lab/software.htm) was used to substitute the nucleotide sequences for the amino acids, thus producing a matrix of aligned nucleotide sequences (supplemental data at http://www.genetics.org/supplemental/). Mega version 2.1 (Kumar *et al.* 2001) was used to translate the alignment into a format for phylogenetic analysis. Phylogenetic analyses were performed in PAUP 4.0b10 (Swofford 2000) with Arabidopsis *AGL6* used to root the tree on the basis of previously published analyses (Purugganan 1997; Hasebe 1999; Theissen *et al.* 2000). Parsimony analyses were performed with heuristic search replicates (100 repetitions of random stepwise taxon addition with TBR branch swapping). Bootstrap support (Felsenstein 1985) for clades was estimated with 1000 heuristic search replicates (random stepwise taxon addition and TBR branch swapping). Analyses were performed using different representatives of the *SEPALLATA* and *AGL6* gene clades as outgroups (not shown); these different outgroup samples produced no appreciable changes in the topology of the ingroup. An analysis was also performed in which 200 nucleotides from the most variable region of the 3′ end of the sequences were eliminated from the aligned data matrix. The results (not shown) were the same except for the placement of *AtFL* and the monophyly of the non-core eudicot clade.

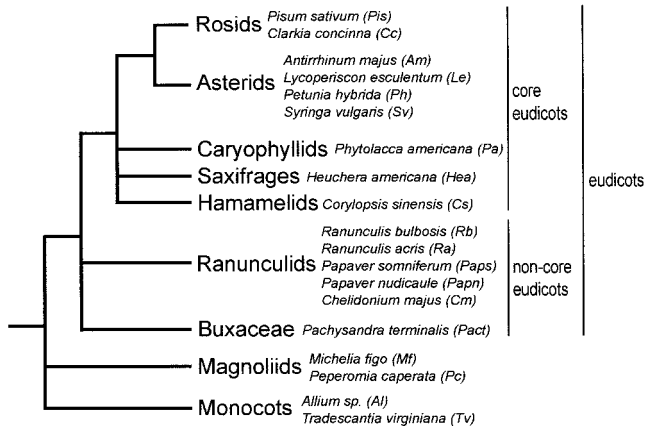

Figure 1.—Outline of angiosperm phylogeny. Arabidopsis is a rosid and Antirrhinum is an asterid. Species used in this study are indicated according to taxonomic group. This simplified phylogeny is based on analyses from Angiosperm Phylogeny Group (1998), Savolainen *et al.* (2000), and Soltis and Soltis (2000).

Core eudicot gene clades were named according to the Arabidopsis gene belonging to that clade (the eu*AP1* and eu*FUL* clades). Non-core eudicot and non-eudicot gene clades were designated "*FUL*-like" on the basis of the similarity of the sequences in these clades to those in the eu*FUL* clade. Genes were named according to species and to gene clade membership; thus *SvAP1* is the eu*AP1* gene isolated from the core eudicot *S. vulgaris* (lilac), *PaFUL* is the eu*FUL* gene isolated from the core eudicot *Phytolacca americana* (pokeweed), and *MfFL* is the *FUL*-like gene isolated from the magnoliid *Michelia figo*.

RESULTS

***AP1/FUL* gene phylogeny is congruent with angiosperm phylogeny:** To generate a phylogeny of the *AP1/FUL* genes, we cloned representatives of this gene lineage from 19 species representing major clades from across the angiosperms (Figure 1). The sequences were used in a parsimony analysis, along with *AP1/FUL* and outgroup (*SEP*, *DAL1*-like, and *AGL6*) sequences available in GenBank. The analysis found two most parsimonious trees that differ only in the relative positions of the three Brassicaceae *AP1* sequences. The consensus of the two trees is shown in Figure 2 and in simplified form in Figure 3. The structure of the monophyletic *AP1/FUL* clade in general mirrors angiosperm phylogeny (Figure 1), with successive branches leading to clades that consist of genes from successive branches of

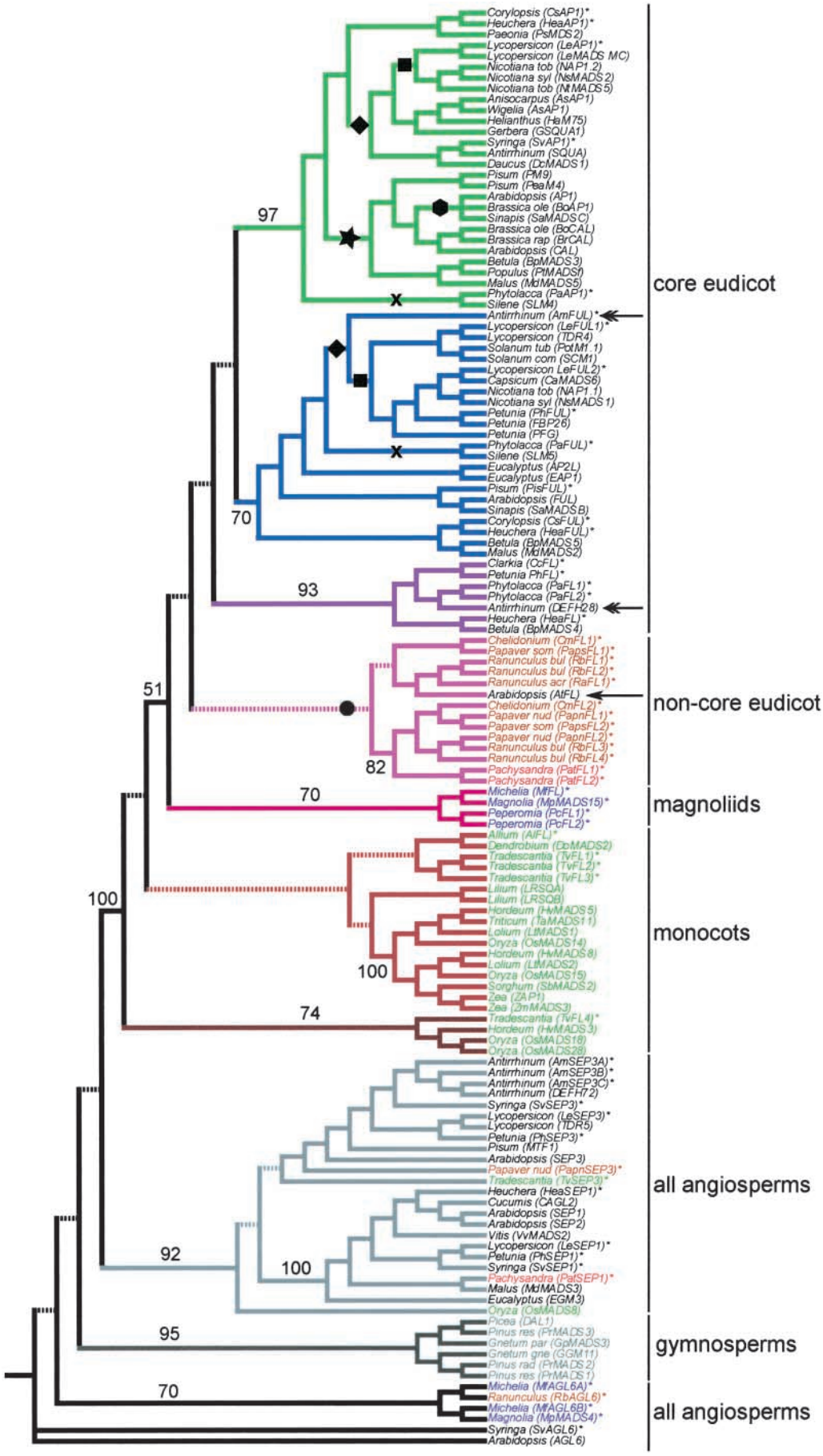

FIGURE 2.—Consensus of the two most parsimonious trees. The data included 133 total sequences. Bootstrap values of ≥50% are indicated for clades discussed in text. Dotted lines indicate bootstrap support of <50% for clades discussed in text. The individual trees were 13,666 steps; CI = 0.16 and RI = 0.58. The only disagreement between the two trees was the relationship of three Brassicaceae genes (black hexagon). Gene clades are indicated by color: green, eu*AP1*; blue, eu*FUL*; purple, core eudicot *FUL*-like; pink, non-core eudicot *FUL*-like; red, magnoliid *FUL*-like; light and dark brown, monocot *FUL*-like; light gray, *SEPALLATA*; dark gray, gymnosperm *DAL1*-like; black, angiosperm *AGL6*-like. Taxonomic affiliations of species from which individual gene sequences were obtained are indicated by font color: black, core eudicot; brown, ranunculid, orange, other non-core eudicot (Pachysandra); blue, magnoliid; green, monocot; gray, gymnosperm. The black circle indicates non-core eudicot clade, and the single arrow points to the Arabidopsis sequence that groups within this clade. Double arrows point to two Antirrhinum sequences with characteristics of *FUL*-like genes. Black squares indicate Solanaceae gene clades, X's indicate caryophyllid gene clades, black diamonds indicate asterid gene clades, and the black star indicates rosid eu*AP1* clade. Asterisks denote sequences generated in this study. See text for details.

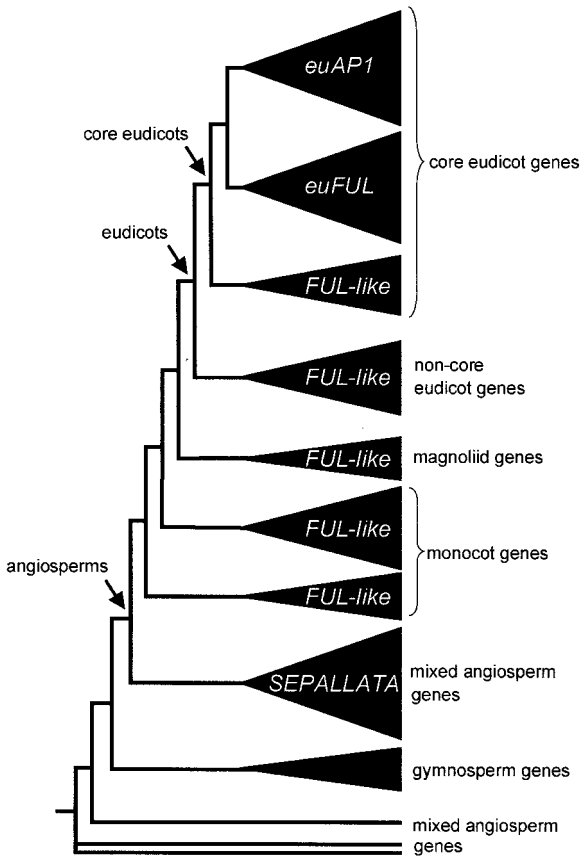

FIGURE 3.—Summary of the results of the phylogenetic analysis. Major gene clades are represented by triangles. The outgroups consist of *SEPALLATA* sequences and *AGL6*-like sequences from gymnosperms and angiosperms.

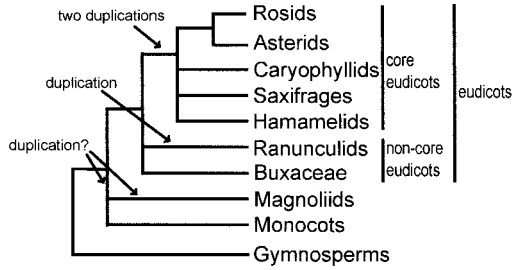

FIGURE 4.—Relationship of duplication events to angiosperm phylogeny. The positions of the duplication events identified in the phylogenetic analysis of the *AP1/FUL* gene lineage are indicated on the angiosperm phylogeny shown in Figure 1. See text for explanation.

the angiosperm phylogenetic tree. The results of the bootstrap analysis (Figure 2) show that although there is strong support for most of the major clades in the *AP1/FUL* phylogeny, there is little support (<50%) for the arrangement of these clades relative to each other. This suggests that conclusions that rely on the order of branching of the major gene clades must be made with caution. However, the congruence of the most parsimonious gene trees with established angiosperm phylogeny provides corroborating evidence for this topology.

Examination of the consensus tree turns up only one inconsistency in the correlation between the topology of the *AP1/FUL* gene tree and the angiosperm phylogenetic tree: the presence of an Arabidopsis sequence (*AtFL*) in the midst of an otherwise non-core eudicot clade (Figure 2). Analysis of the predicted protein sequence of this open reading frame, which is represented in GenBank only as a genomic fragment, shows that it is most likely a highly divergent paralog. In addition, four of the putative *AP1/FUL* sequences generated in this study (*MfAGL6A* and *MfAGL6B* from *M. figo*, *RbAGL6* from *R. bulbosus*, and *SvAGL6* from *S. vulgaris*) proved to be more similar to Arabidopsis *AGL6* than to the *AP1/FUL* genes.

The overall topology of the *AP1/FUL* gene tree corresponds to angiosperm phylogeny, but within the two major clades of core eudicot *AP1/FUL* genes (the eu*AP1* and eu*FUL* gene clades, Figure 3), this congruence breaks down. There are subclades composed of genes from specific angiosperm lineages, for instance, Solanaceae and Caryophyllidae (Figure 2); however, not all subclades are present in both the eu*AP1* and eu*FUL* clades. For example, each has a subclade of asterid sequences, but only the eu*AP1* clade has a subclade of rosid sequences (Figure 2). In addition, the relative positions of some subclades differ; for instance, in the eu*FUL* clade the caryophyllid sequences are nested within the clade, whereas in the eu*AP1* clade they are sister group to the rest of the clade (Figure 2). The lack of congruence between the topologies of the eu*AP1* and eu*FUL* gene clades probably reflects uneven taxonomic sampling in the two clades, but may reflect gene loss in one or both clades.

Phylogeny of the *AP1/FUL* lineage shows several duplication events: Inspection of the consensus tree reveals evidence for the occurrence of several duplications during the history of the *AP1/FUL* gene lineage. The monocot *AP1/FUL* sequences (monocot *FUL*-like genes) fall into two successively branching clades (Figures 2 and 3). This suggests a duplication in the gene lineage either prior to the origin of the monocots, with loss of one of the paralogs in later branching angiosperm lineages, or within the monocots, with unequal rates of divergence in the two resulting gene clades (Figure 4). However, the bootstrap analysis shows poor support (<50%) not only for the placement of the two monocot gene clades relative to each other, but also for the monophyly of the larger of the two monocot clades.

The ranunculid *AP1/FUL* genes (non-core eudicot *FUL*-like genes) also fall into two clades, but in this instance the two clades together form a monophyletic group (Figure 2). This topology suggests an *AP1/FUL* lineage duplication within the ranunculid lineage (Figure 4). Bootstrap support for the sister-group relationship of the two subclades is <50%; however, most of the ranunculid species sampled are represented by at

least one sequence in each of the two clades, providing evidence for a duplication. The absence of a *Papaver nudicaule* sequence from one of the two clades may be due to incomplete sampling or to loss of one lineage in that species. The sequences of the *Pachysandra terminalis* group in one of the two ranunculid clades; this position is supported by a moderately high (82%) bootstrap value (Figure 2). Pachysandra is not a ranunculid but is in the same paraphyletic assemblage of non-core eudicots (Figure 1). The presence of Pachysandra genes in only one of the two clades may indicate that the duplication occurred within the ranunculids and that the sequences in the clade that lack the Pachysandra sequences diverged more rapidly from the ancestral preduplication eudicot sequence.

Within the core eudicots there is evidence for two duplications that produced three gene clades: the eu*AP1* clade, the eu*FUL* clade, and the core eudicot *FUL*-like clade (Figures 2–4). Representatives of the eu*AP1* and eu*FUL* clades have been identified from a wide variety of core eudicot species; however, to date core eudicot *FUL*-like genes have been identified in only six species. The core eudicot *FUL*-like clade has strong bootstrap support (93%) and includes genes from all major core eudicot lineages that were sampled for this analysis, suggesting that more intense sampling of other species may uncover additional members of this clade. The core eudicot *FUL*-like gene clade is sister group to a monophyletic group formed by the eu*AP1* and eu*FUL* gene clades, but bootstrap analysis shows weak support (<50%) for this position.

Amino acid alignment defines conserved C-terminal motifs: The predicted amino acid sequences of the genes included in this analysis have the typical "MIKC" structure of plant type II MADS-domain containing proteins (Alvarez-Buylla *et al.* 2000b). Some residues in the M, I, and K domains appear to be diagnostic for the AP1/FUL clade as compared to the SEP, DAL1-like, and AGL6-like sequences. In general, the sequences included in this analysis are highly conserved throughout the MADS and K domains and somewhat more variable but still conserved in the I domain and N-terminal portion of the C terminus. In contrast, much of the C-terminal domain is widely divergent, even in sequences from closely related species. Most of the AP1/FUL sequences contain C-terminal regions rich in glutamine, but regions rich in proline, serine, or glycine are also common. At the very C terminus all the FUL-like and euFUL sequences show a highly conserved hydrophobic six-amino-acid sequence (FUL-like motif: L/MPPWML). This is generally followed by either two basic residues or one polar and one basic residue (Figure 5A). The number of amino acids between the FUL-like motif and the C terminus of the protein varies from five to seven. A related conserved motif (IPGWML) has been reported in SEPALLATA sequences (Ampomah-Dwamena *et al.* 2002), and it can be seen in related DAL1-like sequences (MQGWMV; Figure 5A). The tryptophan in the fourth position is strictly conserved in all sequences included in this analysis, and the residue following the tryptophan is methionine in all but a few AP1/FUL and SEP sequences. The glycine in the third position is highly conserved in the DAL1-like and SEP sequences (see also Tandre *et al.* 1995) but is replaced by a proline in most FUL-like and euFUL sequences. In the SEP and DAL1-like sequences this motif nearly always terminates the proteins, in contrast to the five- to seven-amino-acid extension that follows the FUL-like motif in FUL-like and euFUL sequences.

The FUL-like motif, identified in the predicted protein products of monocot, magnoliid, ranunculid, and two groups of core eudicot genes (eu*FUL* and *FUL*-like), is absent from the predicted products of the eu*AP1* gene clade. These euAP1 sequences instead have a distinct C terminus with two short conserved motifs, RRNaLaLT/NLa, where "a" is an acidic residue (euAP1 motif), and CFAT/A (farnesylation motif), which terminates the protein (Figure 5A). A variable number of additional acidic residues are just upstream of the euAP1 motif (Figure 5A), and in the case of several euAP1 proteins, this acidic region (including both the euAP1 motif and the upstream region) has been shown to have transcriptional activation properties when tested in yeast (Cho *et al.* 1999). The four terminal amino acids of euAP1 predicted proteins form a farnesylation motif, which is a signal for the attachment of a farnesyl moiety to the cysteine residue; farnesylation causes proteins to be targeted to membranes. Yalovsky *et al.* (2000) showed that Arabidopsis AP1 is farnesylated *in vivo* and *in vitro*. The euFUL and FUL-like proteins, which lack this motif, are most likely not so modified.

DISCUSSION

The *AP1/FUL* genes are found only in angiosperms: Many of the genes that have been shown in Arabidopsis to be key regulators of floral development (*e.g.*, *AP1*, *AP3*, *PI*, *AG*, *SEP1, SEP2, SEP3*) belong to closely related paralogous lineages of the MADS-box gene family. These lineages appear to have arisen as a result of duplication events, although the exact relationship of the lineages to each other and the timing of the duplications is unclear (*e.g.*, Purugganan 1997; Winter *et al.* 1999; Krogan and Ashton 2000; Theissen *et al.* 2000). The *AP3/PI* (Arabidopsis B-function) and *AG* (Arabidopsis C-function) lineages are both present in gymnosperms (Tandre *et al.* 1995; Hasebe 1999; Sundström *et al.* 1999; Becker *et al.* 2000; Theissen *et al.* 2000); thus the origins of the genes responsible for specifying the identity of the male and female reproductive organs predate the origin of the flower. This is consistent with the apparent homology of the male and female gametophytes of angiosperms with those of gymnosperms (Endress 2001).

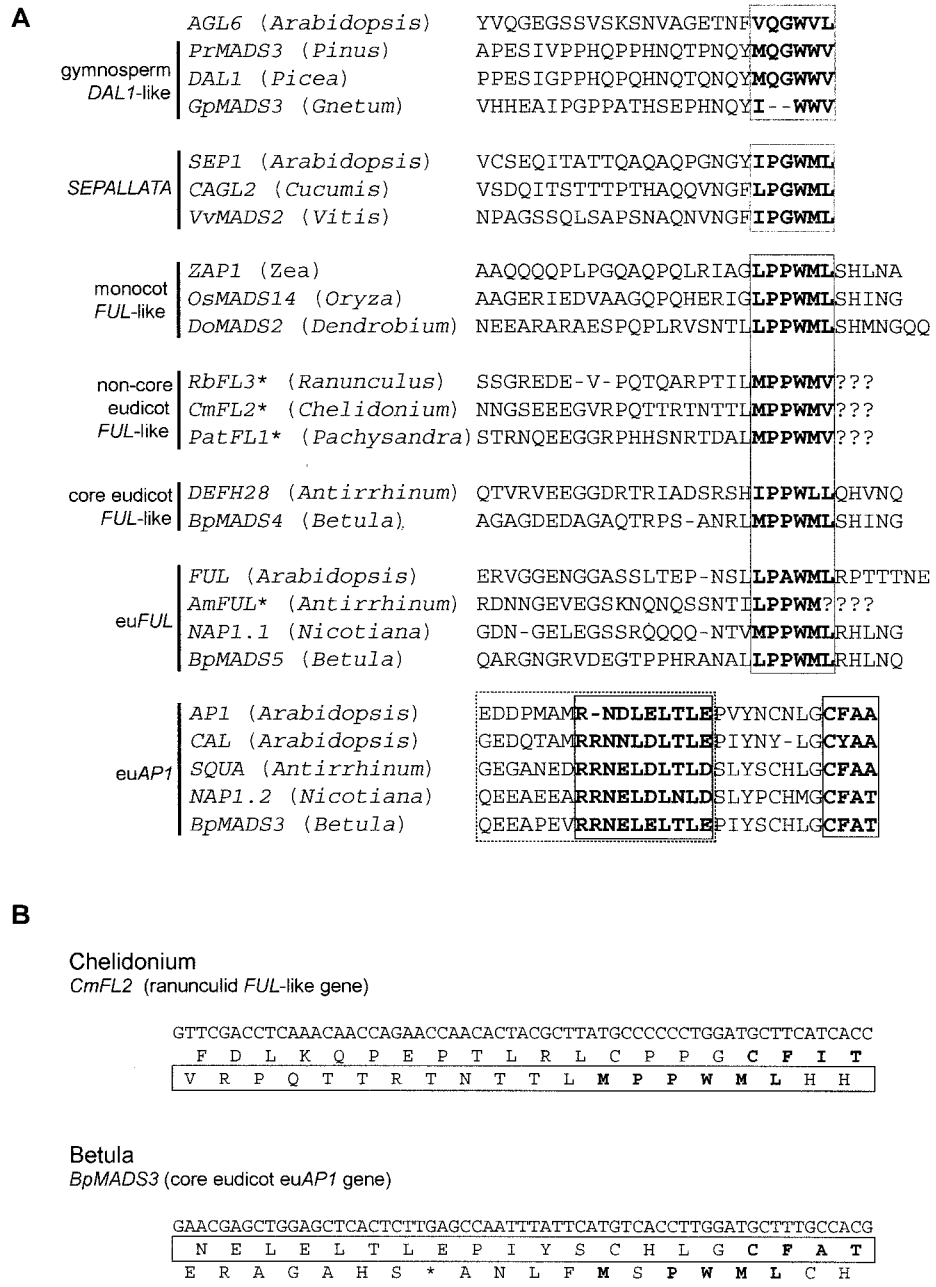

FIGURE 5.—Conserved C-terminal motifs. (A) C terminus of representative outgroup, FUL-like, euFUL, and euAP1 predicted protein sequences. Sequences marked with an asterisk were generated for this study; those sequences are incomplete at the C terminus, indicated by question marks. Conserved sequence motifs are in boldface type and boxed with a solid line (see text for details). Dotted box shows region of euAP1 sequences with a high percentage of acidic amino acids (see text for details). (B) Frameshift relationship between euFUL and euAP1 motifs. Representative sequences showing evidence that the farnesylation motif present in the predicted protein sequences of eu*AP1* genes may have evolved from the FUL-like motif through a translational frameshift. In the two examples presented, the top line is the nucleotide sequence, and the following two lines are two of the three possible translation frames. The predicted correct translations are boxed. In the case of *CmFL2*, the correct frame has the FUL-like motif, but the farnesylation motif can be seen in the incorrect frame. In the case of *BpMADS3*, the correct frame has the farnesylation motif but five of the six amino acids of the FUL-like motif can be seen in the incorrect frame.

In contrast, our data and previous studies (*e.g.*, TANDRE *et al.* 1995; HASEBE 1999; SUNDSTRÖM *et al.* 1999; BECKER *et al.* 2000; THEISSEN *et al.* 2000) indicate that the *AP1/FUL* genes, which have been shown in model species to be required for the specification of floral meristem identity (IRISH and SUSSEX 1990; HUIJSER *et al.* 1992; BOWMAN *et al.* 1993; FERRANDIZ *et al.* 2000), appear to be unique to angiosperms (Figures 2 and 3). Although the gymnosperm *DAL1*-like genes share some sequence similarity with *AP1/FUL* genes, they are more similar to Arabidopsis *AGL6*, and in all analyses they group outside of the *AP1/FUL* clade (HASEBE and BANKS 1997; PURUGGANAN 1997; HASEBE 1999; WINTER *et al.* 1999; BECKER *et al.* 2000; THEISSEN *et al.* 2000). Thus *AP1/FUL* genes, which play a key role in floral specification in model species, appear to be angiosperm specific. This is not the case for the lineages of other genes that are implicated in the specification of floral meristem identity, such as Arabidopsis *LEAFY* (*LFY*). Members of the *LFY* lineage have been identified also in gymnosperms, where they appear to have a role in the specification of reproductive shoot identity (MELLEROWICZ *et al.* 1998; MOURADOV *et al.* 1998; SHINDO *et al.* 1999; FROHLICH and PARKER 2000). Thus the function of *LFY* in determining reproductive identity predates flowers and is not unique to angiosperms. In contrast, the correlation of the origin of the *AP1/FUL* lineage with the origin of flowers suggests a possible role for these genes in the evolution of this key angiosperm feature.

A duplication at the base of the core eudicots produced the eu*AP1* and eu*FUL* clades: Genes of the eu*AP1* clade are found only in core eudicot species, and these

species also possess eu*FUL* genes, thus providing evidence for a duplication that coincided with the origin of this angiosperm clade (Figures 2–4). Core eudicots, which comprise the majority of extant angiosperm species, have a fixed floral architecture, in contrast to earlier diverging angiosperms, which are more plastic in their floral structure (Endress 1992, 1994; Drinnan *et al.* 1994; Albert *et al.* 1998; Soltis *et al.* 2003). In non-eudicot and non-core eudicot species, floral organs may be arranged in discrete whorls, in continuous spirals, or in a combination of both (*e.g.*, whorled perianth but spiral reproductive organs). Particularly in species with spiral phyllotaxy, the number of organs of each type may be variable from flower to flower. In addition, flowers of non-eudicot and non-core eudicot species may have only one type of sterile perianth organ (tepals), rather than a bipartite perianth of differentiated sepals and petals. In the core eudicots these elements of floral structure become fixed; thus all core eudicot flowers have, as a basic plan, a whorled arrangement of four distinct organ types with a fixed number of organs in each whorl.

The fixation of floral structure in the core eudicots suggests that there may have been changes in floral developmental mechanisms that occurred in conjunction with the origin of this angiosperm group. It is thus notable that the duplication event in the *AP1/FUL* lineage that produced the eu*AP1* gene clade occurred at the base of the core eudicots and furthermore that the predicted euAP1 amino acid sequences contain novel C-terminal motifs that are postulated to confer new functional capabilities on the euAP1 proteins. The correlation of the origin of the eu*AP1* gene clade with the fixation of floral structure in the core eudicots suggests that this new protein structure may have played a role in the evolution of the core eudicot flower.

Similar duplications have been identified in the lineages of other MADS-box floral development genes, suggesting that multiple individual gene duplications or a genome-wide duplication event may have played a role in the evolution of core eudicot floral structure. Kramer *et al.* (1998) studied the phylogeny of the *AP3* (Arabidopsis B-function) gene lineage and identified a duplication event that parallels what is seen in the *AP1/FUL* lineage. They found two gene clades (eu*AP3* and *TM6*) within the core eudicots, whereas outside of the core eudicots they found only one lineage (paleo*AP3*). Furthermore, they identified a conserved motif in the predicted paleoAP3 amino acid sequences that was maintained in the core eudicot TM6 clade, but that was lost in the euAP3 clade and was replaced by a novel motif. Examination of the phylogeny of the *AG* lineage (Arabidopsis C-function) also reveals evidence of a duplication that occurred after the divergence of the monocots (Hasebe and Banks 1997; Davies *et al.* 1999; E. Kramer, personal communication). Thus in the lineages of three key regulators of floral development, *AP1, AP3*, and *AG*, we can identify significant evolutionary events that are correlated with the diversification of the core eudicots. This suggests that there may have been changes in the genetic mechanisms regulating floral development that occurred in conjunction with the origin of the core eudicots. Associating specific sequence motif changes with the evolution of particular morphological novelties will require functional analyses of these genes in core eudicot and noncore eudicot species (*e.g.*, Lamb and Irish 2003).

Additional duplications occurred during the evolution of the *AP1/FUL* lineage: The topology of the most parsimonious trees found in this analysis indicates that there have been at least three other duplications within the *AP1/FUL* lineage. In addition to the eu*AP1* and eu*FUL* core eudicot gene clades, the phylogeny shows a third clade of core-eudicot sequences, the core eudicot *FUL*-like genes. The presence of three clades of core eudicot genes suggests that there were two *AP1/FUL* lineage duplication events within the core eudicots (Figures 2–4). However, bootstrap support for this position of the core eudicot *FUL*-like gene clade is weak (<50%; Figure 2). Preliminary analyses based on a slightly smaller data set suggested that the additional duplication occurred at the base of the eudicots, rather than within the core eudicots (results not shown); not surprisingly, bootstrap support for that topology was also low (<50%).

The presence of two clades of monocot *FUL*-like genes is evidence of another duplication. The observed topology (Figures 2 and 3) of successively branching monocot FUL-like gene clades implies that the duplication occurred prior to the origin of the monocots. This requires that both resulting paralogous lineages were maintained in the monocots, but that one of the lineages was lost in later branching angiosperm groups. An alternative explanation is suggested by the uneven taxon representation in the two clades. The smaller, earlier branching clade is composed of genes from three species (Tradescantia, Oryza, and Hordeum), whereas the larger clade is composed of genes from these three species and seven more. Tradescantia, Oryza, and Hordeum are all members of one lineage of monocots, the commelinoids, suggesting that the duplication may have occurred within the commelinoid monocots. Under this scenario, the successive branching pattern of the two monocot *FUL*-like gene clades would most likely be due to a higher rate of divergence in the smaller clade. *AP1/FUL* genes from earlier branching angiosperm lineages are needed to clarify the position of this duplication.

The results of our analysis also indicate a duplication within the ranunculids. The ranunculid *FUL*-like genes group in two subclades, one of which is moderately well supported (82%) but one of which has weak (<50%) bootstrap support (Figure 2). The two clades together form a weakly supported (<50%) monophyletic group, with most species being represented in both clades. This topology suggests a single duplication at the base of the ranunculids. Evidence of duplication events within the ranunculids was also seen in phylogenetic analyses of

the *AP3* and *PI* gene family (Kramer and Irish 1999). The results of that study point to several separate duplication events within different ranunculid lineages. In contrast, our analysis of the *AP1/FUL* lineage suggests a single duplication at the base of the ranunculid clade.

Phylogenetic analysis clarifies orthology and paralogy: The phylogenetic analysis presented here provides a framework for the assessment of the orthology and paralogy of *AP1/FUL* genes by identifying duplication events in the history of this gene lineage and by defining the resulting paralogous gene clades. Previous studies have not had a basis for determining orthology or paralogy of newly identified *AP1/FUL* genes, although differences between eu*AP1* and eu*FUL* genes have been noted (Elo *et al.* 2001; Gocal *et al.* 2001; Jang *et al.* 2002). As a result, comparisons have been made among paralogous genes (*e.g.*, Kyozuka *et al.* 1997; Pouteau *et al.* 1997; Immink *et al.* 1999; Jia *et al.* 2000; Gocal *et al.* 2001). Our analysis clearly demonstrates that the eu*AP1* sequences form a derived clade within the *AP1/FUL* family (Figures 2, 3, and 5A). Thus non-core eudicot and non-eudicot *AP1/FUL* genes, such as those of monocots, magnoliids, and ranunculids, are most appropriately compared with other *FUL*-like and with eu*FUL* genes, and not with eu*AP1* genes. Likewise, comparisons between core eudicot eu*AP1* and eu*FUL* genes should be done in the context of their belonging to paralogous gene lineages and possessing different sequence motifs.

The presence of two distinct clades of core eudicot genes with *FUL*-like sequence characteristics (eu*FUL* and core eudicot *FUL*-like; Figures 2 and 3) suggests that designation of genes as *FUL* orthologs (members of the eu*FUL* clade) may be difficult in the absence of a phylogenetic analysis. On the basis of sequence examination alone it is difficult to determine if a given core eudicot gene belongs to the eu*FUL* or *FUL*-like clade. For instance, Müller *et al.* (2001) suggested *DEFH28* as the probable Antirrhinum *FUL* ortholog on the basis of analysis of the coding sequence and promoter region, the similarities of expression pattern, and the overexpression phenotype in Arabidopsis. In our analysis, however, *DEFH28* groups with the core eudicot *FUL*-like genes, not with the eu*FUL* clade. We cloned, in addition to *SQUA* and *DEFH28*, a third Antirrhinum gene (here designated *AmFUL*) that groups with the eu*FUL* clade and thus can be considered the ortholog of *FUL*. This obviously does not preclude the possibility that *DEFH28* has functional roles similar to those of *FUL*, as suggested by Müller *et al.* (2001). *DEFH28* and *AmFUL* may have redundant functions, a suggestion supported by the close sequence similarity of the eu*FUL* and core eudicot *FUL*-like genes.

Diversification of C-terminal domains: The C-terminal domain of AP1/FUL predicted proteins is highly variable, as is characteristic of plant MADS-domain-containing proteins. Nonetheless, there is a strongly conserved hydrophobic six-amino-acid motif at the end of all FUL-like and euFUL proteins. This FUL-like motif can be seen in the outgroup (SEP, DAL1-like, and AGL6) sequences (Figure 5A), although the exact residue composition is not strictly conserved. The high degree of conservation of this motif is a strong indication that it is functionally important and suggests that its loss and replacement with a different motif in euAP1 proteins may result in altered functional capabilities of the euAP1 proteins.

Several studies have investigated the significance of the C-terminal domain and the conserved motifs of euAP1 proteins. Krizek and Meyerowitz (1996) made constructs consisting of the MI region of AP1 and the KC region of AGAMOUS and found that under the control of the CaMV 35S promoter the construct was able to partially rescue the strong *ap1-1* mutant. Constructs consisting only of the AP1 MI domains failed to produce an overexpression phenotype in wild-type Arabidopsis. Krizek and Meyerowitz (1996) concluded that although K and C domains are required for proper AP1 protein function, the specific motifs present in the AP1 K and C domains are not completely required and can be partially replaced by those of AG.

In contrast, other studies have localized specific functions of Arabidopsis AP1 and Antirrhinum SQUA to the C-terminal domain (Cho *et al.* 1999; Egea-Cortines *et al.* 1999; Yalovsky *et al.* 2000). A domain rich in acidic residues (Figure 5A) has been identified in Arabidopsis AP1 and the putative euAP1 orthologs from *Raphanus sativus*, *Nicotiana sylvestris*, and *N. tabacum* as a transcription activation domain (Cho *et al.* 1999). In Arabidopsis AP1 and its ortholog from the closely related Raphanus this domain was found to function synergistically with an upstream glutamine-rich region, the two regions together producing a higher activation level than the sum of that produced by the two regions separately. This upstream region is lacking from the Nicotiana sequences, which correspondingly showed lower levels of activation. The euFUL and FUL-like predicted proteins lack an acidic domain, but most have regions rich in glutamine or proline. Glutamine- and proline-rich regions have been shown to confer transcription activation activity (Gerber *et al.* 1994), but this function has not been tested in these proteins. The high degree of conservation of the acidic domain in euAP1 sequences, however, suggests that these proteins will show strong transcription activation activity.

The final four amino acids of the predicted proteins of eu*AP1* genes conform to a farnesylation signal (CaaX), which has been shown to be functional in AP1 and to be required to produce an AP1 overexpression phenotype in Arabidopsis (Yalovsky *et al.* 2000). Research in other species, however, is equivocal as to whether this farnesylation motif is required for proper functioning of these proteins. The predicted product of *PEAM4*, the eu*AP1* gene from *P. sativum*, ends after the second amino acid of the farnesylation motif. Nonetheless,

when overexpressed in Arabidopsis in the strong *ap1-1* mutant background, *PEAM4* is able largely to rescue the mutant phenotype (Berbel *et al.* 2001). *NtMADS11*, from *N. tabacum* (Jang *et al.* 2002), and *LtMADS2*, from *Lolium temulentum* (Gocal *et al.* 2001), are *FUL*-like genes, the predicted products of which lack the farnesylation motif entirely, yet, when overexpressed in Arabidopsis, they also show partial rescue of strong *ap1* mutants. In all these cases the heterologous gene is able partially to replace *AP1* function, but not completely. One can therefore interpret this as meaning either that the farnesylation motif is unimportant, because proteins that lack it can partly substitute for AP1 functionally, or that the farnesylation motif is required, because proteins that lack it are not able to completely substitute for AP1. However, the high degree of conservation suggests that this motif is likely to play an important role in modulating eu*AP1* gene function.

New motifs in euAP1 proteins may have arisen via translational frameshift: Inspection of different possible translation frames of ranunculid *FUL*-like and various core eudicot eu*AP1* sequences shows that the change in C-terminal motifs may have arisen at least in part via a simple translational frameshift. For instance, the translation of *CmFL2*, a noncore eudicot *FUL*-like gene identified from Chelidonium (Figure 1), in one frame terminates with RLCPPGCFIT, the final four amino acids of which form a canonical farnesylation motif. However, in the correct frame the translation is LMP GWMLHH, which lacks the farnesylation motif but has the expected FUL-like motif (Figure 5B). Evidence for this frameshift can be seen in the different translation frames of some genes from non-core eudicot species, which have only *FUL*-like genes (as in the Chelidonium example above), as well as in genes from core eudicot species, which possess eu*AP1*, eu*FUL*, and *FUL*-like genes. For example, the correct translation of the Betula eu*AP1* gene *BpMADS3* terminates with CHLGCFAT, whereas one of the two alternative translations terminates with MSPWMLCH, which contains five of the six residues characteristic of the FUL-like motif, including the strictly conserved tryptophan (Figure 5B). Thus the farnesylation motif characteristic of the predicted protein products of the eu*AP1* genes may have been derived by insertion of a single nucleotide or by loss of two nucleotides upstream of the FUL-like motif of an ancestral *FUL*-like gene.

Implications for the ABC model of floral organ specification: Arabidopsis *AP1* and Antirrhinum *SQUA* are members of the eu*AP1* clade, and likewise Arabidopsis *AP3* and its Antirrhinum ortholog *DEF* are members of the eu*AP3* clade (Kramer *et al.* 1998); these clades are core-eudicot specific, and genes with those sequence properties are not found in species outside of the core eudicots. *AP1/SQUA* and *AP3/DEF* are among the key regulatory genes upon which the ABC model is based; however, non-core eudicot species such as ranunculids, monocots, and magnoliids do not have eu*AP1* or eu*AP3* genes. By extension, these species lack proteins with the conserved motifs characteristic of euAP1 and euAP3 proteins. The correlation between the origin of these derived genes and the origin of the core eudicots suggests that at that point in angiosperm evolution there were changes in the mechanisms regulating floral development. Thus we must consider the possibility that outside of the core eudicots the molecular mechanisms underlying floral development may differ from what is seen in Arabidopsis and Antirrhinum.

Arabidopsis remains the only species identified so far in which a mutant for a gene belonging to the *AP1/FUL* lineage results in a misspecification of organ identity (Irish and Sussex 1990; Mandel *et al.* 1992; Bowman *et al.* 1993). However, Arabidopsis has three closely related genes, *AP1*, *FUL*, and *CAL*, with redundant roles in floral meristem specification (Ferrandiz *et al.* 2000); thus it is likely that even strong *ap1* mutants show only a partial loss of function for this role. In strong *ap1* mutants, the meristem shows some floral identity, particularly in the inner two whorls, in which the reproductive organs develop nearly normally. The outer two whorls show some transition to floral character in the whorled arrangement of the organs, but their leaf-like nature and the presence of secondary flowers demonstrate a persistent inflorescence character. Thus the loss of proper organ identity may be a consequence of the incomplete nature of the transition from inflorescence to floral meristem, particularly in the outer two whorls.

According to this interpretation of the role of the *AP1* gene, there is no discrete A-function; rather, the apparent misspecification of floral organ identity in the Arabidopsis *ap1* mutant is a consequence of the incomplete specification of floral meristem identity. If sepal production represents the ground state function of a florally determined meristem, floral organ identity can be adequately specified with only the equivalent of the B- and C-function of the ABC model, as articulated by Schwarz-Sommer *et al.* (1990; see also Egea-Cortines and Davies 2000; Theissen *et al.* 2000). Thus the lack of examples of A-function mutants, coupled with the data presented here regarding the restriction of the eu*AP1* genes to the core eudicots, suggests that the universality and perhaps the concept of A-function should be reevaluated.

The authors thank Akira Kanno, Francisco Madueño, and Elena Kramer for permission to use unpublished sequences, Jongmin Nam for use of the aa2dna program, and our colleagues, in particular Lena Hileman, for comments on the manuscript. This work was supported by grant no. 2001-35304-09901 from the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service to A.L.

LITERATURE CITED

Albert, V. A., M. H. G. Gustafsson and L. DiLaurenzio, 1998 Ontogenetic systematics, molecular developmental genetics, and the

angiosperm petal, pp. 349–374 in *Molecular Systematics of Plants II*, edited by D. E. Soltis, P. S. Soltis and J. J. Doyle. Kluwer Academic Publishers, New York.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Alvarez-Buylla, E. R., S. J. Liljegren, S. Pelaz, S. E. Gold, C. Burgeff *et al.*, 2000a MADS box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. Plant J. **24:** 457–466.

Alvarez-Buylla, E. R., S. Pelaz, S. J. Liljegren, S. E. Gold, C. Burgeff *et al.*, 2000b An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA **97:** 5328–5333.

Ampomah-Dwamena, C., B. A. Morris, P. Sutherland, B. Veit and J.-L. Yao, 2002 Down-regulation of *TM29*, a tomato *SEPALLATA* homolog, causes parthenocarpic fruit development and floral reversion. Plant Physiol. **130:** 605–617.

Angenent, G. C., J. Franken, M. Busscher, A. Van Dijken, J. L. Van Went *et al.*, 1995 A novel class of MADS box genes is involved in ovule development in Petunia. Plant Cell **7:** 1569–1582.

Angiosperm Phylogeny Group, 1998 An ordinal classification for the families of flowering plants. Ann. Mo. Bot. Gard. **85:** 531–553.

Becker, A., K.-U. Winter, H. Saedler and G. Theissen, 2000 MADS-box gene diversity in seed plants 300 million years ago. Mol. Biol. Evol. **17:** 1425–1434.

Berbel, A., C. Navarro, C. Ferrandiz, L. A. Canas, F. Madueno *et al.*, 2001 Analysis of PEAM4, the pea AP1 functional homologue, supports a model for AP1-like genes controlling both floral meristem and floral organ identity in different plant species. Plant J. **25:** 441–451.

Bowman, J. L., J. Alvarez, D. Weigel, E. M. Meyerowitz and D. R. Smyth, 1993 Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes. Development **119:** 721–743.

Buchner, P., and J.-P. Boutin, 1998 A MADS box transcription factor of the AP1/AGL9 subfamily is also expressed in the seed coat of pea (Pisum sativum) during development. Plant Mol. Biol. **38:** 1253–1255.

Cho, S., S. Jang, S. Chae, K. M. Chung, Y.-H. Moon *et al.*, 1999 Analysis of the C-terminal region of Arabidopsis thaliana APETALA1 as a transcription activation domain. Plant Mol. Biol. **40:** 419–429.

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

Davies, B., P. Motte, E. Keck, H. Saedler, H. Sommer *et al.*, 1999 PLENA and FARINELLI: redundancy and regulatory interactions between two antirrhinum MADS-box factors controlling flower development. EMBO J. **18:** 4023–4034.

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

Drinnan, A. N., P. R. Crane and S. B. Hoot, 1994 Patterns of floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots). Plant Syst. Evol. **8:** 93–122.

Egea-Cortines, M., and B. Davies, 2000 Beyond the ABCs: ternary complex formation in the control of floral organ identity. Trends Plant Sci. **5:** 471–476.

Egea-Cortines, M., H. Saedler and H. Sommer, 1999 Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. EMBO J. **18:** 5370–5379.

Elo, A., J. Lemmetyinen, M. L. Turunen, L. Tikka and T. Sopanen, 2001 Three MADS-box genes similar to APETALA1 and FRUITFULL from silver birch (Betula pendula). Physiol. Plant **112:** 95–103.

Endress, P. K., 1992 Evolution and floral diversity: the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum*. Int. J. Plant Sci. **153:** S106–S122.

Endress, P. K., 1994 Floral structure and evolution of primitive angiosperms: recent advances. Plant Syst. Evol. **192:** 79–97.

Endress, P. K., 2001 Origins of flower morphology. J. Exp. Zool. **291:** 105–115.

Felsenstein, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

Ferrandiz, C., Q. Gu, R. Martienssen and M. F. Yanofsky, 2000 Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. Development **127:** 725–734.

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

Gerber, H. P., K. Seipel, O. Georgiev, M. Hofferer, M. Hug *et al.*, 1994 Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science **263:** 808–811.

Gocal, G. F., R. W. King, C. A. Blundell, O. M. Schwartz, C. H. Andersen *et al.*, 2001 Evolution of floral meristem identity genes. Analysis of Lolium temulentum genes related to APETALA1 and LEAFY of Arabidopsis. Plant Physiol. **125:** 1788–1801.

Gu, Q., C. Ferrandiz, M. F. Yanofsky and R. Martienssen, 1998 The *FRUITFULL MADS*-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development **125:** 1509–1517.

Hasebe, M., 1999 Evolution of reproductive organs in land plants. J. Plant Res. **112:** 463–474.

Hasebe, M., and J. A. Banks, 1997 Evolution of MADS gene family in plants, pp. 179–198 in *Evolution and Diversification of Land Plants*, edited by K. Iwatsuki and P. H. Raven. Springer, Tokyo.

Huijser, P., J. Klein, W.-E. Lonnig, H. Meijer, H. Saedler *et al.*, 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.

Immink, R. G. H., D. J. Hannapel, S. Ferrario, M. Busscher, J. Franken *et al.*, 1999 A petunia MADS box gene involved in the transition from vegetative to reproductive development. Development **126:** 5117–5126.

Irish, V. F., and I. M. Sussex, 1990 Function of the apetala-1 gene during Arabidopsis floral development. Plant Cell **2:** 741–753.

Jang, S., K. An, S. Lee and G. An, 2002 Characterization of tobacco MADS-box genes involved in floral initiation. Plant Cell Physiol. **43:** 230–238.

Jia, H., R. Chen, B. Cong, K. Cao, C. Sun *et al.*, 2000 Characterization and transcriptional profiles of two rice MADS-box genes. Plant Sci. **155:** 115–122.

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

Kramer, E. M., and V. F. Irish, 1999 Evolution of genetic mechanisms controlling petal development. Nature **399:** 144–148.

Kramer, E. M., R. L. Dorit and V. F. Irish, 1998 Molecular evolution of petal and stamen development: gene duplication and divergence within the APETALA3 and PISTILLATA 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.

Krogan, N. T., and N. W. Ashton, 2000 Ancestry of plant MADS-box genes revealed by Bryophyte (Physcomitrella patens) homologues. New Phytol. **147:** 505–517.

Kumar, S., K. Tamura, I. Jakobsen and M. Nei, 2001 MEGA2: molecular evolutionary genetics analysis software, Arizona State University, Tempe, AZ.

Kyozuka, J., R. Harcourt, W. J. Peacock and E. S. Dennis, 1997 Eucalyptus has functional equivalents of the Arabidopsis AP1 gene. Plant Mol. Biol. **35:** 573–584.

Lamb, R. S., and V. F. Irish, 2003 Functional divergence within the *PETALA3/PISTILLATA* floral homeotic gene lineages. Proc. Natl. Acad. Sci. USA **100:** 6558–6563.

Lawton-Rauh, A. L., E. R. Alvarez-Buylla and M. D. Purugganan, 2000 Molecular evolution of flower development. Trends Ecol. Evol. **15:** 144–149.

Lowman, A. C., and M. D. Purugganan, 1999 Duplication of the *Brassica oleracea APETALA1* floral homeotic gene and the evolution of domesticated cauliflower. Genetics **90:** 514–520.

Mandel, M. A., C. Gustafson-Brown and M. F. Yanofsky, 1992 Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature **360:** 273–277.

Mellerowicz, E. J., K. Horgan, A. Walden, A. Coker and C. Walter, 1998 PRFLL—a Pinus radiata homologue of FLORICAULA and LEAFY is expressed in buds containing vegetative

shoot and undifferentiated male cone primordia. Planta **206:** 619–629.

Moon, Y. H., H. G. Kang, J. Y. Jung, J. S. Jeon, S. K. Sung *et al.*, 1999 Determination of the motif responsible for interaction between the rice APETALA1/AGAMOUS-LIKE9 family proteins using a yeast two-hybrid system. Plant Physiol. **120:** 1193–1204.

Mouradov, A., T. Glassick, B. Hamdorf, L. Murphy, B. Fowler *et al.*, 1998 NEEDLY, a Pinus radiata ortholog of FLORICAULA/LEAFY genes, expressed in both reproductive and vegetative meristems. Proc. Natl. Acad. Sci. USA **95:** 6537–6542.

Müller, B. M., H. Saedler and S. Zachgo, 2001 The MADS-box gene DEFH28 from Antirrhinum is involved in the regulation of floral meristem identity and fruit development. Plant J. **28:** 169–179.

Pelucchi, N., F. Fornara, C. Favalli, S. Masiero, C. Lago *et al.*, 2002 Comparative analysis of rice MADS-box genes expressed during flower development. Sex. Plant Reprod. **15:** 113–122.

Pouteau, S., D. Nicholls, F. Tooke, E. S. Coen and N. Battey, 1997 The induction and maintenance of flowering in *Impatiens*. Development **124:** 3343–3351.

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. **45:** 392–396.

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

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.

Rounsley, S. D., G. S. Ditta and M. F. Yanofsky, 1995 Diverse roles for MADS box genes in Arabidopsis development. Plant Cell **7:** 1259–1269.

Savolainen, V., M. W. Chase, S. B. Hoot, C. M. Morton, D. E. Soltis *et al.*, 2000 Phylogenetics of flowering plants based on combined analysis of plastid atpB and rbcL gene sequences. Syst. Biol. **49:** 306–362.

Schwarz-Sommer, Z., P. Huijser, W. Nacken, H. Saedler and H. Sommer, 1990 Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science **250:** 931–936.

Shindo, S., M. Ito, K. Ueda, M. Kato and M. Hasebe, 1999 Characterization of MADS genes in the gymnosperm Gnetum parvifolium and its implication on the evolution of reproductive organs in seed plants. Evol. Dev. **1:** 180–190.

Soltis, E. D., and P. S. Soltis, 2000 Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol. Biol. **42:** 45–75.

Soltis, D. E., A. E. Senters, M. J. Zanis, S. Kim, J. D. Thompson *et al.*, 2003 Gunnerales are sister to other core eudicots and exhibit some floral features of early-diverging eudicots. Am. J. Bot. **90:** 461–470.

Southerton, S., H. Marshall, M. Aidyn and R. D. Teasdale, 1998 Eucalyptus MADS-box genes expressed in developing flowers. Plant Physiol. **118:** 365–372.

Sundström, J., A. Carlsbecker, M. E. Svensson, M. Svenson, U. Johanson *et al.*, 1999 MADS-box genes active in developing pollen cones of Norway spruce (Picea abies) are homologous to the B-class floral homeotic genes in angiosperms. Dev. Genet. **25:** 253–266.

Swofford, D. L., 2000 *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sinauer Associates, Sunderland, MA.

Tandre, K., V. A. Albert, A. Sundas and P. Engstrom, 1995 Conifer homologues to genes that control floral development in angiosperms. Plant Mol. Biol. **27:** 69–78.

Taylor, S. A., J. M. I. Hofer, I. C. Murfet, J. D. Sollinger, S. R. Singer *et al.*, 2002 *PROLIFERATING INFLORESCENCE MERISTEM*, a MADS-box gene that regulates floral meristem identity in Pea. Plant Physiol. **129:** 1150–1159.

Theissen, G., A. Becker, A. Di Rosa, A. Kanno, J. T. Kim *et al.*, 2000 A short history of MADS box genes in plants. Plant Mol. Biol. **42:** 115–149.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Weigel, D., and E. M. Meyerowitz, 1994 The ABCs of floral homeotic genes. Cell **78:** 203–209.

Winter, K. U., A. Becker, T. Munster, J. T. Kim, H. Saedler *et al.*, 1999 MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc. Natl. Acad. Sci. USA **96:** 7342–7347.

Yalovsky, S., M. Rodriguez-Concepcion, K. Bracha, G. Toledo-Ortiz and W. Gruissem, 2000 Prenylation of the floral transcription factor APETALA1 modulates its function. Plant Cell **12:** 1257–1266.

Communicating editor: D. Weigel