Patterns of Gene Duplication and Functional Evolution During the Diversification of the *AGAMOUS* Subfamily of MADS Box Genes in Angiosperms

Elena M. Kramer,[1] M. Alejandra Jaramillo and Verónica S. Di Stilio[2]

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT

Members of the *AGAMOUS* (*AG*) subfamily of MIKC-type MADS-box genes appear to control the development of reproductive organs in both gymnosperms and angiosperms. To understand the evolution of this subfamily in the flowering plants, we have identified 26 new *AG*-like genes from 15 diverse angiosperm species. Phylogenetic analyses of these genes within a large data set of *AG*-like sequences show that ancient gene duplications were critical in shaping the evolution of the subfamily. Before the radiation of extant angiosperms, one event produced the ovule-specific D lineage and the well-characterized C lineage, whose members typically promote stamen and carpel identity as well as floral meristem determinacy. Subsequent duplications in the C lineage resulted in independent instances of paralog subfunctionalization and maintained functional redundancy. Most notably, the functional homologs *AG* from Arabidopsis and *PLENA* (*PLE*) from Antirrhinum are shown to be representatives of separate paralogous lineages rather than simple genetic orthologs. The multiple subfunctionalization events that have occurred in this subfamily highlight the potential for gene duplication to lead to dissociation among genetic modules, thereby allowing an increase in morphological diversity.

THE production of reproductive organs is arguably the most important process in the development of any organism, particularly from an evolutionary standpoint. In the angiosperm model species *Arabidopsis thaliana*, the MADS-box gene *AGAMOUS* (*AG*) is critical to the formation of sex organs in the developing flower (Bowman *et al.* 1989). This function is a component of what is known as the ABC model of organ identity determination. The ABC model describes the combinatorial activities of three classes of genes, termed A, B, and C, which function in overlapping domains to encode the identity of organ primordia that arise from the floral meristem (Coen and Meyerowitz 1991) (Figure 1). In Arabidopsis, *AG* is the primary C class gene, *APETALA1* (*AP1*) and *APETALA2* (*AP2*) are the A class genes, and *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) represent the B class (Bowman *et al.* 1991, 1993). With the exception of *AP2*, all of these genes are representatives of the pan-eukaryotic MADS-box family of transcription factors (reviewed in Shore and Sharrocks 1995; Theissen *et al.* 2000). More specifically, they are classified as type II (Alvarez-Buylla *et al.* 2000) or MIKC-type (Mun-

ster *et al.* 1997), a plant-specific group within the MADS-box gene family. The MIKC abbreviation reflects a conserved structure composed of four domains: the MADS (M) domain, responsible for DNA binding and dimerization (Riechmann *et al.* 1996b); the intervening (I) and keratin-like (K) domains, which mediate dimerization between different MIKC-type proteins (Riechmann *et al.* 1996a); and the variable C-terminal (C) domain, which appears to promote higher-order protein interactions (Egea-Cortines *et al.* 1999). Further investigations into the functions of other florally acting MIKC-type MADS-box genes have led to modifications of the ABC model (Figure 1). On the basis of analysis of the *FBP7* and *FBP11* genes from Petunia, D function was proposed as responsible for the establishment of ovule identity (Colombo *et al.* 1995). Recently, E class genes (Theissen and Saedler 2001), represented in Arabidopsis by *SEPALLATA1–3* (*SEP1–3*), have been identified as critical facilitators of B and C class function (Pelaz *et al.* 2000). Our current understanding of the biochemical interactions among the A, B, C, and E class proteins is that dimerization between individual proteins is mediated by the M, I, and K domains while the interaction of dimers to form higher-order complexes is controlled by the C domain (Egea-Cortines *et al.* 1999; Honma and Goto 2001).

Given their critical roles in controlling floral development, the diversification of the MADS-box gene family has been cited as an important factor in the radiation of the land plants (Theissen *et al.* 2002). This connection
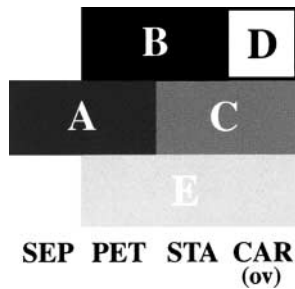
FIGURE 1.—ABC model with modifications as suggested in THEISSEN *et al.* (2002). SEP, sepals; PET, petals; STA, stamens; CAR, carpels; ov, ovules.

between gene duplication, functional diversification, and the evolution of complexity has a relatively long history, having been most notably outlined by OHNO (1970). Early studies of the phenomenon focused on two major pathways of paralog evolution: pseudogene formation *vs.* the acquisition of novel gene function (neofunctionalization). As we have gained a better understanding of the complex nature of gene functions, however, the process of subfunctionalization, as suggested by HUGHES (1999) and elaborated by Force and Lynch (FORCE *et al.* 1999; LYNCH and FORCE 2000), has come to the forefront. Under this model, multiple ancestral functions of a gene lineage may become partitioned between paralogs, causing the duplicates to be selectively maintained without neofunctionalization. Long-term, however, subfunctionalization may result in some degree of divergence as the paralogs specialize or eventually acquire additional functions (HUGHES 1999). It has also become clear that functional redundancy can be maintained for surprisingly long periods (HUGHES and HUGHES 1993), possibly because of an advantage conferred by genetic buffering (ZHANG 2003).

Due to the fairly large amount of functional data that are available for *AG* homologs, this subfamily of MADS-box genes is well suited for an analysis of patterns of functional evolution. In addition to promoting stamen and carpel identity, *AG* function includes repression of *AP1* expression in the third and fourth whorls (GUSTAFSON-BROWN *et al.* 1994) and establishment of the determinate nature of the floral meristem (BOWMAN *et al.* 1989). Analyses of *AG* homologs from both core eudicots and monocots indicate that these functions are broadly conserved, but gene duplications have introduced variation (BRADLEY *et al.* 1993; KEMPIN *et al.* 1993; PNUELI *et al.* 1994; KANG *et al.* 1998; YU *et al.* 1999; KAPOOR *et al.* 2002; KYOZUKA and SHIMAMOTO 2002). For example, the Antirrhinum gene *PLENA* (*PLE*) functions very similarly to *AG* (CARPENTER and COEN 1990), but some aspects of stamen identity are mediated by the closely related gene *FARINELLI* (*FAR*; DAVIES *et al.* 1999). Parsing of function is also thought to have occurred in *Zea mays*, where the paralogs *ZAG1* and

ZMM2 appear to be subfunctionalized into carpel- and stamen-specific paralogs, respectively (MENA *et al.* 1996). In other instances, neofunctionalization has followed gene duplication, as in the case of the *SHATTERPROOF* (*SHP1* and *2*) genes, which are *AG*-like genes from Arabidopsis (LILJEGREN *et al.* 2000). One aspect of their function is to specify tissues that are unique to the silique fruit of the Brassicaceae, indicating that this activity may have been acquired relatively recently (THEISSEN 2000). However, other components of *SHP1/2* function are redundant with both *AG* and the *FBP7*-like gene *SEEDSTICK* (*STK*; FAVARO *et al.* 2003; PINYOPICH *et al.* 2003). *AG* homologs have been identified in all the major gymnosperm lineages (TANDRE *et al.* 1995; RUTLEDGE *et al.* 1998; WINTER *et al.* 1999; JAGER *et al.* 2003) and analyses of expression in Gnetum and Picea suggest that members of the *AG* subfamily play a deeply conserved role in the production of reproductive tissue. In contrast, no clear *AG*-like genes have been recovered in studies of lower vascular plants (MUNSTER *et al.* 1997; HASEBE *et al.* 1998; SVENSSON and ENGSTROM 2002) or mosses (KROGEN and ASHTON 2000; HENSCHEL *et al.* 2002).

Despite these extensive comparative studies, many aspects of the evolution of the *AG* subfamily remain unclear. In particular, the timing of various gene duplication events and the ensuing patterns of molecular and functional evolution are not well defined. In this study, we have sought to obtain better resolution of ortholog/paralog relationships within the phylogeny of *AG*-like genes. To these ends, 26 new *AG* homologs have been identified from 15 angiosperm taxa spanning the core eudicots, magnoliid dicots, and basal ANITA grade (the earliest branching lineages of the angiosperms). Phylogenetic analyses of the expanded *AG* data set have clarified the evolution of the separate C and D gene lineages and revealed both ancient and recent gene duplications. Most notably, we have found that *PLE* and *AG* are not simple genetic orthologs but represent relatively ancient paralogous lineages. This confirms a previous, more limited analysis, which suggested that *AG* and *FAR* are orthologous (DAVIES *et al.* 1999). The implications of these findings for the evolution of gene function within the *AG* subfamily are discussed.

MATERIALS AND METHODS

Plant material: A broad developmental range of floral tissue was obtained from the following taxa: *Saxifraga caryana* (Saxifragaceae), *Phytolacca americana* (Phytolaccaceae), *Ranunculus ficaria* (Ranunculaceae), *Helleborus orientalis* (Ranunculaceae), *Clematis integrifolia* (Ranunculaceae), *Aquilegia alpina* (Ranunculaceae), *Thalictrum dioicum* (Ranunculaceae), *Berberis gilgiana* (Berberidaceae), *Akebia quinata* (Lardizabalaceae), *Sanguinaria canadensis* (Papaveraceae), *Meliosma dilleniifolia* (Sabiaceae), *Houttuynia cordata* (Saururaceae), *Chloranthus spicatus* (Chloranthaceae), *Saruma henryii* (Aristolochiaceae), and *Nymphaea* sp. (Nymphaeaceae). Voucher information for all of these species is available in supplemental Table 1 at http://www.genetics.org/supplemental/.

Cloning and characterization of *AG* homologs: Isolation of *AG* homologs was performed using RT-PCR in a manner similar to that described in KRAMER *et al.* (1998). Initial amplification of first-strand cDNA used a degenerate forward primer (5′-GGIMGIGGIAARATIGARATIAARMGIAT) designed to the highly conserved first 10 amino acids of the MADS domain with a poly(T) reverse primer, 5′-CCGGATCTCTAGACGGC CGC(T)$_{17}$. The products of the primary PCR reaction were cleaned with the QIAquick PCR purification kit (QIAGEN, Valencia, CA), diluted 1:100, and used as template in a PCR reaction with a second degenerate primer, 5′-ACIAAYMGI CARGTIACITTYTG, and the same anchored poly(T) reverse primer. This second forward primer is designed to the highly conserved MADS-box sequence TNRQVTFC, in which the C-terminal cysteine represents a synapomorphy for the *AG* subfamily (THEISSEN *et al.* 1996). All PCR amplifications were performed in 100 μl of PCR buffer (200 mM Tris-HCl, pH 8.4; 500 mM KCl; 50 mM MgCl$_2$) containing 50 and 10 pmol of 5′ and 3′ primer, respectively, 25 μmol of each dNTP, and 2 units of PlatinumTaq polymerase (Invitrogen, Carlsbad, CA). The amplification program began with a 12-min activation step at 95°, followed by a 1-min incubation step at 95°, a 30-sec annealing step at temperatures ranging from 50° to 65°, and a 1-min extension at 72°. The program was repeated for 37 cycles and was terminated by a 10-min incubation step at 72°. The amplified PCR products were cloned using the TOPO TA cloning kit (Invitrogen) per manufacturer's instructions. For each taxon, 50–200 clones of >650 bp were characterized by sequencing (BigDye Terminator v3.0, ABI prism 3100, Applied Bioscience, Foster City, CA) and/or restriction analysis. At least 5 independent clones were sequenced for every putative locus. All cDNA sequences have been deposited in GenBank (for accession numbers, see supplemental data available online at http://www.genetics.org/supplemental/). *ScAG*, *CsAG1*, and *CsAG2* were identified in the context of an earlier screen (KRAMER and IRISH 2000), but are being reported here for the first time.

5′ rapid amplification of cDNA ends (RACE) was performed on *MdAG1*, *SrhAG*, and *NymAG1* using the SMART cDNA RACE kit (BD Biosciences Clontech, Palo Alto, CA). Reverse primers for each locus are as follows: *MdAG1*, 5′-ACTATTGTT TGCATATTCATAAAGCCGGCCGCGAGT; *SrhAG*, 5′-TGTGA CATAACCTCATACCCTCCCCCACCTG; and *NymAG1*, 5′-TTC ACTGACACCTTCGCCTAGCATTTGCC.

Phylogenetic analyses: Additional *AG*-like sequences were identified on the basis of previously published analyses and BLAST searches (ALTSCHUL *et al.* 1997; for references and accession numbers, see Table S1 available online at http:// www.genetics.org/supplemental/). In cases in which the database contained nearly identical sequences from the same taxon, only one representative was included. Full-length amino acid and nucleotide alignments of the 26 new *AG* homologs with 66 previously released *AG*-like sequences were initially compiled using ClustalW. ClustalW multiple-alignment parameters were gap penalty 8 and gap extension penalty 2, using the PAM protein weight matrix for the amino acid alignment with transitions weighted for the nucleotide. The alignments were then refined by hand using MacClade 4.0 (MADDISON and MADDISON 2000), and final amino acid and nucleotide alignments were adjusted so that they were identical (for NEXUS files, see supplementary data available at http://www.genetics.org/ supplemental/). The N-terminal extensions present in many *AG*-like genes were excluded from the alignments. The nucleotide alignment was used for phylogenetic analyses while the equivalent amino acid alignment was used only to identify shared sequence characters and generally conserved motifs.

Although the C domain tends to show much lower conservation than the other three regions, alignment is typically possible within subfamilies (KRAMER *et al.* 1998; JOHANSEN *et al.* 2002; TZENG *et al.* 2002). In the case of the *AG* lineage, the generally higher degree of sequence conservation further facilitates the alignment of the C domain. The majority of the indels in this region are due to expansions in repetitive sequences, rather than to a large number of nonsynonymous changes. Several particularly long repetitive stretches (more than five amino acids) present in the C domains of the grass *AG*-like genes were condensed to one or two amino acids to facilitate alignment (see Figure 2). Analyses of a data set lacking the C domain produced phylogenies very similar to those obtained with the full-length alignment, but with less resolution at recent nodes and generally lower bootstrap support (data not shown).

Maximum-parsimony (MP) trees were generated through heuristic searches of 1000 random stepwise additions, with tree bisection-reconnection branch swapping and saving of multiple parsimonious trees (MulTrees on). Gaps were encoded as missing data and third positions were excluded. Bootstrap support was estimated by performing 1000 heuristic searches with 10 additional sequence replicates per bootstrap, using the same criteria as in the original search. Wilcoxon sign-rank (also called a Templeton test; TEMPLETON 1983) and Kishino-Hasegawa (KISHINO and HASEGAWA 1989) tests were conducted on the MP trees to explore topologies that would suggest alternative patterns of gene duplication.

Bayesian phylogenetic analyses were conducted on the nucleotide alignments, including all positions using the program MRBAYES v3.0 (HUELSENBECK and RONQUIST 2001). The best model of evolution was determined using Modeltest v3.06 (POSADA and CRANDALL 1998). The model of DNA substitution selected was GTR + I + Γ, which assumes general time reversibility (GTR), a certain proportion of invariable sites (I), and a gamma approximation of the rate variation among sites (Γ). The option "codon" was used for the nucleotide substitution model, following the probabilistic model of codon evolution by MUSE and GAUT (1994). We ran four chains of the Markov chain Monte Carlo, sampling 1 tree every 100 generations for 1,000,000 generations starting with a random tree. The search reached stationarity after ∼23,000 generations. The first 23,000 generations were considered the "burn-in" period and were not included in generating the consensus phylogeny.

Cloning and characterization of intron 8 region of Nymphaea *AG* homologs: *Nymphaea* sp. genomic DNA was prepared from leaf tissue using the DNeasy plant mini kit (QIAGEN). To obtain fragments of the *NymAG1* genomic locus, the DNA was amplified using a specific forward primer, NymAG1F 5′-CAGCACATCAATCTAATGGAATCCTCCCACCAC with a specific reverse primer, NymAG1R 5′-TGGACCCAACATATT CATGTTACTAATGCTGCTGAT. The primers were designed to regions of the *NymAG1* cDNA predicted to fall within exon 7 for NymAG1F and exon 8 for NymAG1R. PCR amplification was performed using a BD Advantage Genomic PCR kit (BD Biosciences Clontech) per manufacturer's instructions. The amplification program began with a 1-min activation step at 94°, followed by a 15-sec denaturing step at 94°, a 20-sec annealing step at 50°–60°, and a 3-min extension step at 68°, repeated for 30 cycles. The resulting genomic fragments, of ∼1.8 kb in length, were cloned using the TOPO TA cloning kit (Invitrogen). Approximately 30 clones were screened for size and 6 clones were sequenced as described above. The resulting consensus genomic sequence was aligned to the *NymAG1* cDNA to determine exon/intron boundaries. The *NymAG3* genomic fragment was similarly obtained and analyzed using a forward primer, 5′-CTGGAACTACAAAGTGATAATATGTATCTTCGA, designed to fall within exon 6, and a reverse primer, 5′-CAGA CAACACCATAGCATATTGTGCGGTA, designed to bind within the last exon of the cDNA.

RESULTS

Characterization of *AG* homologs and phylogenetic analysis—*AG* homologs show a high degree of conservation: Twenty-six *AG*-like cDNAs were identified in 15 taxa from the core eudicots, magnoliid dicots, and ANITA group. Alignment of the predicted amino acid sequences of the new loci with those of previously identified *AG* homologs reveals a high degree of conservation throughout the M, I, and K regions, with many positions nearly invariant throughout the seed plants (for amino acid alignment, see supplementary data at http://www.genetics.org/supplemental/). Beyond the traditionally defined K domain (Ma *et al.* 1991), positions 95–165 in our alignment, a fairly high level of identity extends through position 185. This includes the K3 region that has been recognized by some researchers as a putative third α-helix (Yang *et al.* 2003). The expected **a** and **d** positions of the predicted (**abcdefg**)$_n$ repeats identified by Yang *et al.* (2003) are all very highly conserved (see online supplemental Figure 1 available at http://www.genetics.org/supplemental/). Although two of the central **a** sites are occupied by charged (165E) or polar (172N) residues rather than by hydrophobic amino acids, buried polar residues have been shown to play important roles in dimerization interactions between AP3 and PI in Arabidopsis (Yang *et al.* 2003), as well as other eukaryotic transcription factors (Zeng *et al.* 1997).

Following position 185, conservation decreases, with multiple indels due to the expansion of repetitive sequences, particularly in the grass homologs. At the very C-terminal end of the proteins, there are two short, highly conserved regions, which we have termed AG motif I and AG motif II (Figure 2). These motifs primarily contain hydrophobic and polar residues and have no recognizable relation to known functional motifs. They do have some similarity in makeup to the conserved C-terminal sequences of the B lineage, the PI and paleoAP3 motifs (Kramer *et al.* 1998), but no clear positional homology is discernible. The conservation of these regions throughout seed plant *AG*-like sequences defines them as synapomorphies for the subfamily.

Phylogenetic analyses reveal patterns of ancient gene duplication: A full-length nucleotide alignment of 92 *AG*-like sequences was analyzed using MP as implemented by PAUP 4.0b10 (Swofford 2001) and Bayesian analysis using MRBAYES 3.0 (Huelsenbeck and Ronquist 2002). Gymnosperm *AG*-like sequences were used to root the trees on the basis of the findings of earlier studies (Hasebe and Banks 1997; Winter *et al.* 1999; Theissen *et al.* 2000). The resulting MP and Bayesian phylogenies (Figures 3 and 4) are largely in agreement with only minor differences (see below). Overall, the MP analysis shows lower bootstrap support for many nodes, while the Bayesian analysis has relatively high posterior probability values for a majority of nodes. However, posterior probabilities are known to be consider-

ably less stringent than bootstrap values (Suzuki *et al.* 2002; Alfaro *et al.* 2003; Douady *et al.* 2003) and should be considered upper boundaries of confidence for the relationships depicted at these nodes.

Both analyses give strong support to a clade containing all of the angiosperm sequences. Consistent with this, there are many distinct amino acid apomorphies for the angiosperm and gymnosperm clades; however, the lack of an established outgroup for the *AG* subfamily makes it impossible to determine which character states were primitive in the ancestor of all seed plants. Within the angiosperms, the loci are divided into two major clades, which we have termed the C and D lineages. Each lineage contains representatives from throughout the angiosperms, including the basal ANITA group, indicating that they were produced by an ancient gene duplication that predated the diversification of extant angiosperms.

The designation of the "D" lineage is based on the inclusion of the so-called D class genes from Petunia, *FBP7* and *FBP11* (Colombo *et al.* 1995), and follows terminology used in previous publications (Tzeng *et al.* 2002). In the D clade, the position of the Nymphaea representative *NymAG3* differs between the MP and Bayesian analyses. The strongly supported basal placement of *NymAG3* in the Bayesian tree (Figure 4) is more consistent with the position of the Nymphaeales in current angiosperm phylogenies (Qiu *et al.* 1999; Zanis *et al.* 2002). The monocots are represented by the Agapanthus gene *ApMADS2* and a group of grass homologs that are divided into two paralogous lineages, one including Oryza P0408G07.14 and Zea *ZMM25* and the other, Oryza *OsMADS13* and the Zea genes *ZAG2* and *ZMM1*. This indicates that an early gene duplication event that occurred before the common ancestor of rice and maize was followed by a later maize-specific duplication, which yielded the *ZAG2/ZMM1* pair (Figures 3 and 4, solid circles in D lineage). Magnoliid dicot and eudicot sequences are also present in the D clade, demonstrating that representatives of the lineage are widely conserved across the angiosperms. It is notable, however, that no D orthologs were recovered in the RT-PCR survey of the Ranunculales, a finding that is being pursued further with genomic analyses.

The D lineage has a number of distinguishing sequence characteristics, some of which are shown in Figure 5. Overall, members of the clade show higher variability in the AG motif I and II regions than do the gymnosperm *AG*-like genes or C-lineage homologs. Within the D lineage, the core eudicot loci are associated with a loss of conservation in the second residue of AG motif I and with the conversion of positions 6 and 7 in AG motif II to highly conserved lysine residues (Figure 2 and amino acid alignment in online supplemental data available at http://www.genetics.org/supplemental/).

We also investigated whether aspects of genomic structure represent a synapomorphy for the D lineage. *AG*
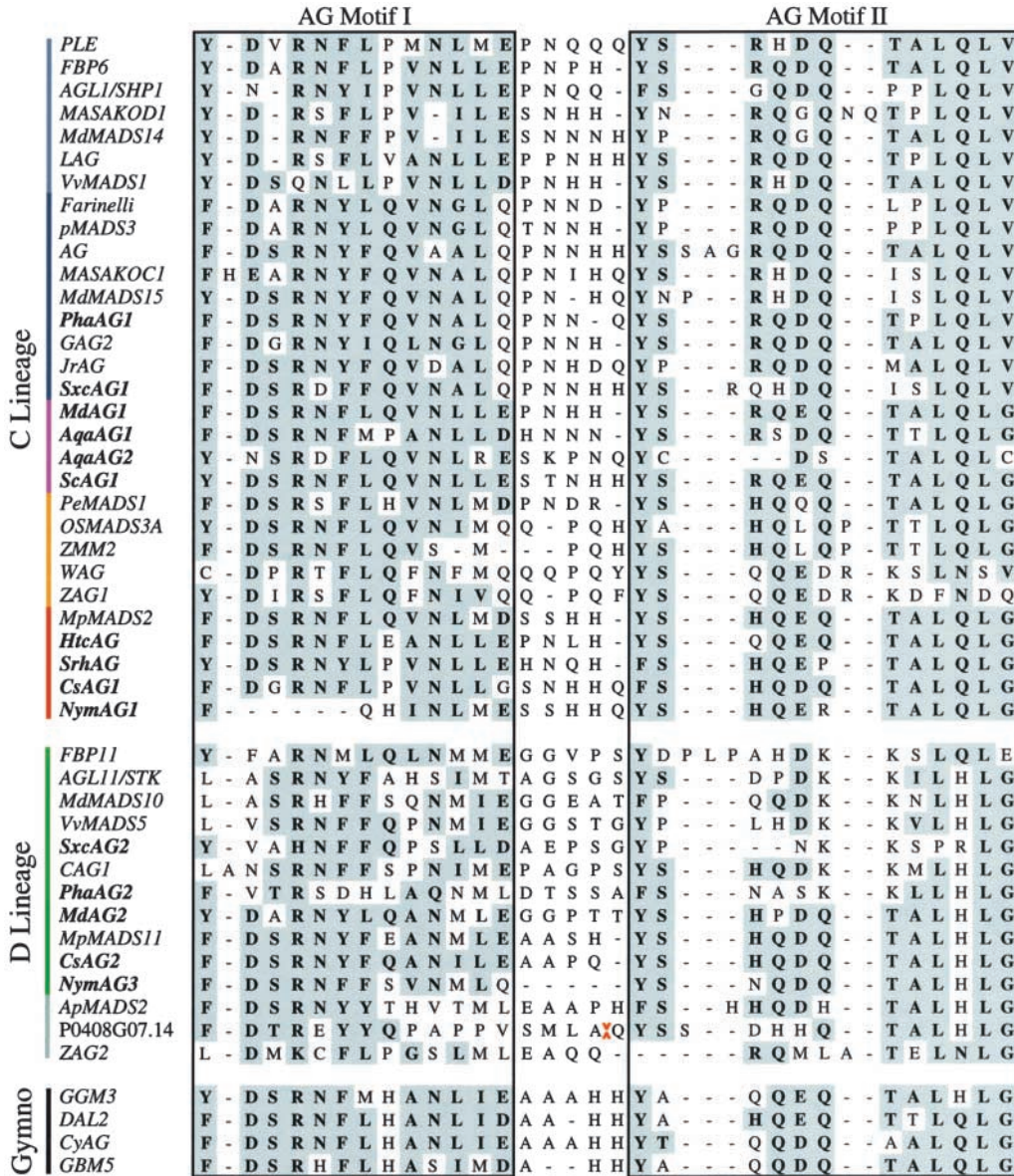
FIGURE 2.—Alignment of C-terminal regions of predicted amino acid sequences for select representatives of the C and D lineages and gymnosperm (Gymno) *AG*-like genes. Colored vertical bars on the left correspond to the phylogenetic positions of the adjacent genes (see Figure 3). Sequences shown in boldface type were identified in this study. Two highly conserved regions, AG motif I and AG motif II, are boxed. Residues that show chemical conservation with the C-lineage consensus sequence are shown in boldface type and shaded. Red arrows in P0408G07.14 indicate the position of a stretch of seven alanines that were removed from the alignment. Consensus sequences for both motifs are from each of the three major lineages in the *AG* subfamily.

homologs are unusual for MIKC-type genes in that they often possess eight introns rather than the typical six (BRUNNER *et al.* 2000; JOHANSEN *et al.* 2002). The additional two introns are positioned 5′ of the MADS domain and in the last codon of AG motif II, which is commonly the last codon of the protein (YANOFSKY *et al.* 1990; BRADLEY *et al.* 1993). This organization is observed in several C-lineage members and in one gymnosperm (RUTLEDGE *et al.* 1998; BRUNNER *et al.* 2000), suggesting that the presence of eight introns is likely to be primitive in the *AG* subfamily. However, all of the D-lineage genes for which genomic structure is available (*AGL11/STK, OsMADS13, ZAG2,* and *ZMM11*) are missing intron 8 at the 3′-end of AG motif II (THEISSEN *et al.* 1995; ARABIDOPSIS GENOME INITIATIVE 2000; CHOISNE *et al.* 2002). Although this sampling is quite limited, it does include both core eudicot and grass species and could indicate

that the loss of intron 8 is a shared character of the D lineage. To explore this possibility, we cloned and sequenced genomic fragments corresponding to the intron 8 region of C- and D-lineage representatives from *Nymphaea, NymAG1* and *NymAG3*, respectively. Alignment of the genomic and cDNA sequences clearly shows that both *NymAG1* and *NymAG3* have introns at the expected position for intron 8 (see online supplemental Figures 2 and 3 available at http://www.genetics.org/supplemental/). These findings indicate that the D lineage did not lose intron 8 before the radiation of extant angiosperms. It remains possible that the D lineage lost intron 8 after the early divergence of the Nymphaeales, but it may also be that the lack of intron 8 in *STK* and the grass D genes arose independently. This potential clearly exists since it is also known to have occurred independently at least once in the C lineage, as evi-
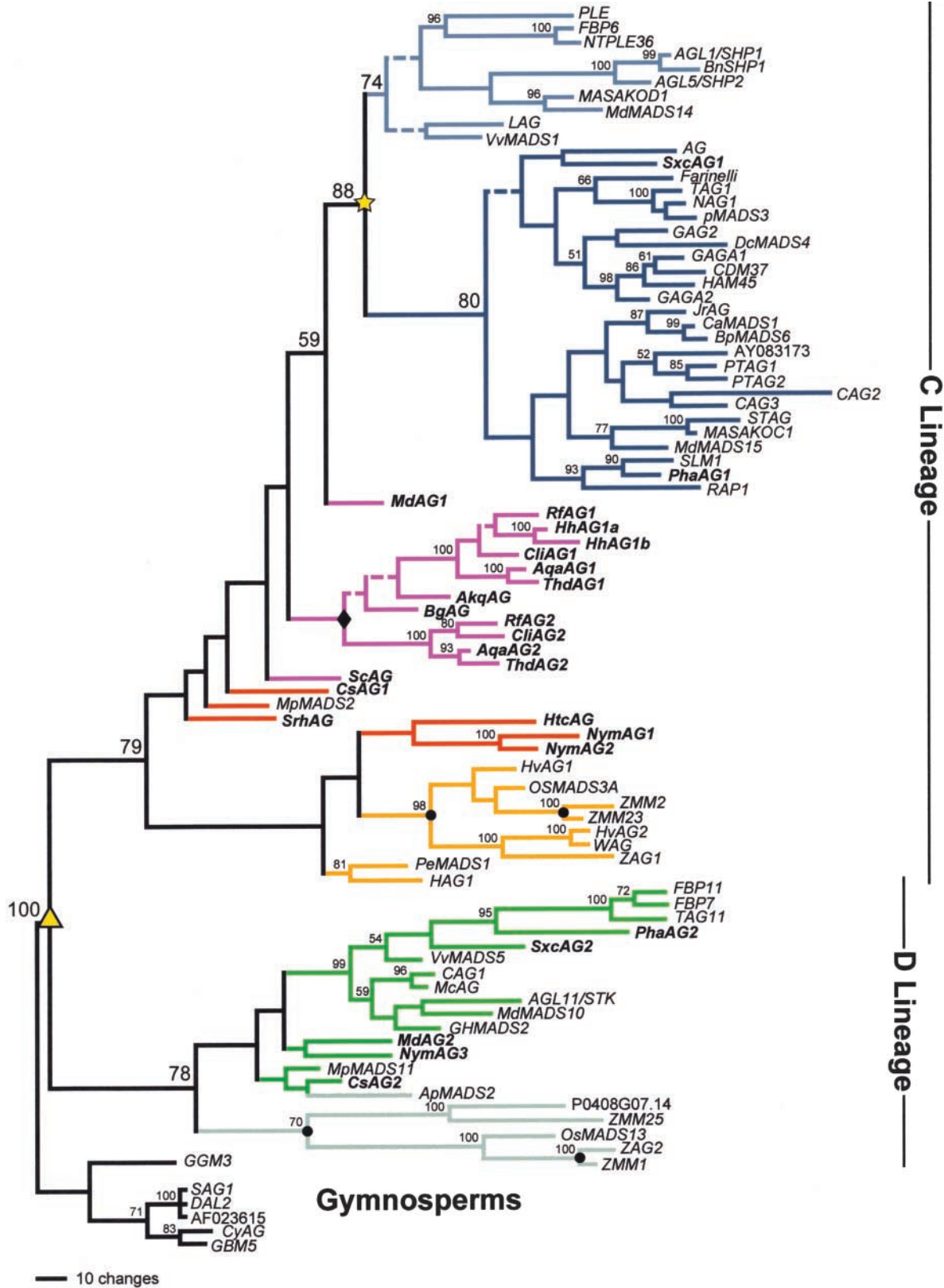
FIGURE 3.—One randomly chosen tree from 20 equally parsimonious trees of 3228 steps. The numbers next to each node give bootstrap support from 1000 replicates. Gene names shown in boldface type were identified in this study. Dashed branches collapse in the strict consensus. Branch coloring is as follows: black, gymnosperm *AG*-like genes; gray, monocot D lineage; dark green, magnoliid dicot, ANITA grade, and lower eudicot D lineage; light green, core eudicot D lineage; orange, monocot C lineage; red, magnoliid dicot and ANITA grade C lineage; purple, lower eudicot C lineage; dark blue, eu*AG* core eudicot C lineage; and light blue, *PLE* core eudicot C lineage. The yellow triangle indicates the C/D gene duplication; the black circles, gene duplications in the grass C and D lineages; the black diamond, a gene duplication in the Ranunculales C lineage; and the yellow star, the eu*AG/PLE* gene duplication.
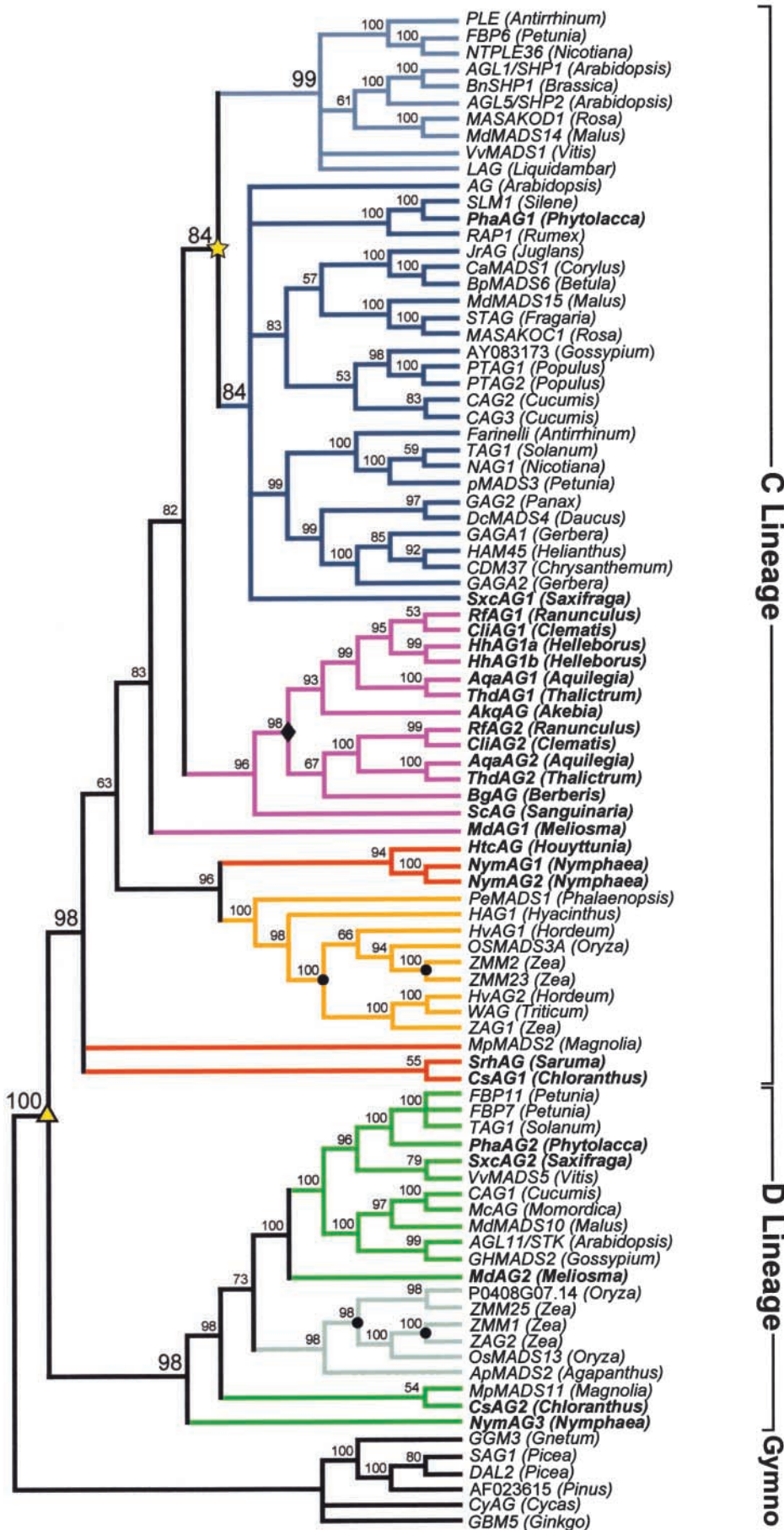
FIGURE 4.—A 50% majority rule tree derived from those trees sampled after "burn-in." The numbers next to each node indicate the posterior probabilities for those branches. Branch coloration and symbols have the same significance as in Figure 3. The taxon of origin is shown in parentheses after each gene name.
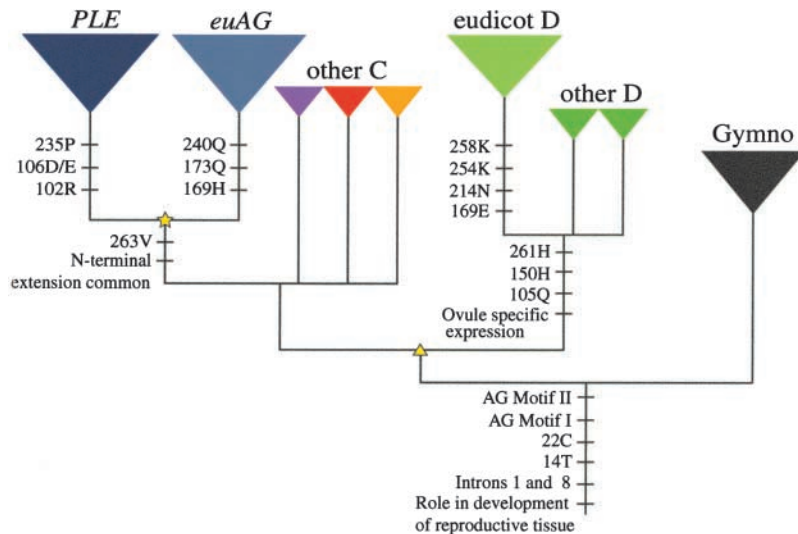
FIGURE 5.—Simplified phylogeny of the *AG* subfamily with diagnostic character states mapped onto branches. Sequence character states refer to the amino acid alignment (see online supplemental data at http://www.genetics.org/supplemental/). Other character states were inferred on the basis of 5′ RACE, comparison of genomic and cDNA sequences, and published reports of expression patterns (see text). The yellow triangle indicates the C/D duplication event while the star represents the eu*AG*/*PLE* duplication.

denced by the *SHP1/2* genes of Arabidopsis (MA *et al.* 1991).

The C lineage contains both of the originally described C-function genes, *AG* from Arabidopsis and *PLE* from Antirrhinum. The *NymAG-1* and *-2* loci do not fall at the base of the C clade in either analysis, which could indicate ancient patterns of gene duplication and extinction, but also may be an artifact due to the limited sampling from magnoliid dicots and the ANITA group. Parsimony analyses in which the Nymphaea loci are constrained to the base of the C lineage produced 30 trees only 9 steps longer than the original MP tree, a difference that is not significant by either the KH or the Templeton test. Monocot representatives include loci from the Orchidaceae, Amaryllidaceae, and Poaceae. The topology of the grass C-lineage genes suggests a pattern of gene duplication similar to what is observed in the D lineage: an early gene duplication was apparently followed by a later event in the Zea lineage. The *AG* homologs from the Ranunculales form a well-supported, single clade in the Bayesian analysis, but they are paraphyletic in the MP tree. In both phylogenies, the Ranunculaceae loci are separated into two paralogous lineages, indicating that they were produced by a gene duplication that at least predated the last common ancestor of the family (solid diamonds in Figures 3 and 4). The position of the lower eudicot Meliosma, represented by *MdAG1*, differs somewhat between the MP and Bayesian analyses, with the MP position being more consistent with the most recent phylogeny of the eudicots (SOLTIS *et al.* 2003).

All of the core eudicot C-lineage loci fall into a single clade with strong support in both analyses; however, this group is deeply split into two separate lineages. *PLE* and other *AG*-like genes from Petunia, Nicotiana (tobacco), Arabidopsis, Malus (apple), Rosa, Vitis (grapevine), and Liquidambar (sweetgum) form one clade, which we refer to as the *PLE* lineage. Sister to this lineage

is what we call the eu*AG* lineage, which includes *AG*, the Antirrhinum gene *FAR*, and an array of *AG* homologs from across the core eudicots. The *PLE* and eu*AG* lineages include six paralog pairs, such as *FBP6* and *pMADS3* from Petunia, which comprise taxa from both the Rosids and the Asterids, the two major core eudicot groups. Furthermore, loci from the Vitaceae, Caryophyllales, and Saxifragales are clearly placed in one lineage or the other. This topology indicates that the paralogous *PLE* and eu*AG* lineages were produced by a gene duplication that occurred before the diversification of the core eudicots, meaning that *AG* and *PLE* are not simple genetic orthologs but relatively ancient paralogs.

To test this finding, we reanalyzed the data set using MP under a series of topological constraints. If all core eudicot loci are constrained by superorder (Rosids, Asterids, etc.), the analysis recovers 31 trees, each 35 steps longer than the MP tree, which are significantly different by both tests at $P < 0.001$. In these trees, the eu*AG* and *PLE* lineage members still sort out into two corresponding clades within each constrained superorder group (data not shown). The use of backbone constraints that would accept the pre-core eudicot duplication but force *AG* and *PLE* to be genetic orthologs resulted in 24 trees, 20 steps longer than the original MP tree, a difference that is significant at $P < 0.05$. Consistent with these results, the *PLE* and eu*AG* lineages each possess a number of diagnostic amino acid character states (Figure 5).

One characteristic commonly found in C-lineage members is the presence of a N-terminal extension preceding the MADS domain (JAGER *et al.* 2003), which is not typical of MIKC-type MADS-box genes (PURUGGANAN *et al.* 1995). These regions are variable in sequence and length, ranging from 13 to 52+ amino acids (see online supplemental Figure 4 at http://www.genetics.org/supplemental/; JAGER *et al.* 2003). Of the 31 complete core eudicot mRNA sequences, 29 show extensions

but they are found only in three of the eight complete monocot sequences (see supplemental online Figure 4 at http://www.genetics.org/supplemental/). N-terminal extensions have not been seen in any D-lineage members or gymnosperm *AG*-like genes characterized to date (JAGER *et al.* 2003). Analysis of AG function in Arabidopsis indicates that the large N-terminal extension found in this protein is not essential to any major aspect of gene function (MIZUKAMI *et al.* 1996). The most likely scenario for the appearance of N-terminal extensions in C-lineage members seems to be that in-frame ATG codons have evolved several times independently within the large 5′-untranslated region that is common to the *AG* subfamily (JAGER *et al.* 2003). To explore the evolution of this novel domain, we performed 5′ RACE on *MdAG1*, *ThdAG1*, *SrhAG*, and *Nym AG1*, C-lineage loci representing the lower eudicots, magnoliid dicots, and ANITA grade. None of these cDNAs display N-terminal extensions and have the first in-frame ATG immediately preceding the MADS domain. This suggests that the frequent presence of an N-terminal extension is primarily a characteristic of the core eudicot C-lineage members, with the domain having evolved independently at least one other time in the monocots.

DISCUSSION

Implications of sequence conservation: It is perhaps not surprising to find that members of the *AG* subfamily exhibit a high degree of sequence conservation, given their critical role in producing reproductive organs. Consistent with this pattern, several studies have shown that constitutive expression of heterologous *AG*-like genes in Arabidopsis (RUTLEDGE *et al.* 1998; TANDRE *et al.* 1998) or in Nicotiana (MANDEL *et al.* 1992; KANG *et al.* 1995) produces phenotypes similar to that of 35S:*AG* (MIZUKAMI and MA 1992). These results suggest that the sequence conservation of *AG* homologs reflects a similar conservation of biochemical interactions. While the M, I, and K domains have been clearly shown to be involved in DNA binding and protein dimerization (RIECHMANN *et al.* 1996a,b), the function of the C domain is poorly understood. Ectopic expression experiments have demonstrated that deletion of the entire C domain, including most of the putative K3 α-helix (see online supplemental Figure 1 at http://www.genetics. org/supplemental/), produces a dominant negative form of AG (MIZUKAMI *et al.* 1996). This indicates that although the C domain is not required for DNA binding or dimerization, it is essential for full protein function. Furthermore, conserved C-terminal motifs have been identified in many lineages of MIKC-type MADS-box genes (KRAMER *et al.* 1998; JOHANSEN *et al.* 2002; LITT and IRISH 2003; VANDENBUSSCHE *et al.* 2003) and several lines of evidence indicate that these motifs are functionally important (KRIZEK and MEYEROWITZ 1996; LAMB

et al. 2002). It remains to be determined, however, how components of the C domain, such as the K3 or C-terminal motifs, might contribute to higher-order protein interactions or other aspects of *AG* function.

Gene duplications in the C lineage have led to subfunctionalization and maintained redundancy: *AGAMOUS and PLENA are not simple genetic orthologs:* Phylogenetic analyses of the large *AG* homolog data set show that *PLE* and *AG* actually represent paralogous lineages derived from a gene duplication that occurred within the lower eudicots. This confirms similar results obtained in much more limited analyses (DAVIES *et al.* 1999; KROGEN and ASHTON 2000; SVENSSON *et al.* 2000). Representatives of both the *PLE* and eu*AG* lineages have been identified in six taxa but loss-of-function data are available only for the paralogs from Arabidopsis, Antirrhinum, and Petunia (BOWMAN *et al.* 1989; CARPENTER and COEN 1990; DAVIES *et al.* 1999; LILJEGREN *et al.* 2000; KAPOOR *et al.* 2002). In Arabidopsis, *AG* and *SHP1/2* exhibit a mix of redundant and distinct functions. While *AG* fulfills the primary aspects of C function (BOWMAN *et al.* 1989), *SHP1* and *-2* play both a unique role in the differentiation of the replum margin (LILJEGREN *et al.* 2000) and a redundant one in promoting carpel and ovule identity (WESTERN and HAUGHN 1999; PINYOPICH *et al.* 2003). Similar to what has been found with other types of paralogs (LEE and SCHIEFELBEIN 2001; SKAER *et al.* 2002), *SHP1/2* can substitute for aspects of *AG*'s stamen identity function, although they do not usually perform this role (PINYOPICH *et al.* 2003). The *SHP1/2* genes are thought to be genetically downstream of *AG*, possibly directly (SAVIDGE *et al.* 1995). In Antirrhinum, functional evolution has taken an alternate route, leaving *PLE* the primary C-function gene and the eu*AG* ortholog *FAR* a largely redundant paralog that contributes to stamen differentiation (DAVIES *et al.* 1999). In contrast to what is observed in Arabidopsis, *PLE* and *FAR* are not functionally interchangeable and it is *FAR* that appears to be genetically downstream of *PLE*. Loss-of-function analysis of *pMADS3* in Petunia suggests that *pMADS3* and *FBP6* are neither fully redundant nor completely separate in function, with both contributing to aspects of organ identity and meristem determinacy (KAPOOR *et al.* 2002).

Given that the combined functions of the paralog pairs in each species are roughly equivalent, the most parsimonious explanation is that most of these functions were present in the common ancestral repertoire. Following their formative gene duplication event, ∼100–120 million years ago (MYA; MAGALLON *et al.* 1999), it appears that subfunctionalization was the primary trend, although various degrees of maintained redundancy have also been observed. While it may be common for *AG* homologs to control late aspects of carpel development, the role of *SHP1/2* in the replum margin could be considered a kind of neofunctionalization. It remains possible, if not likely, that alternate scenarios such as paralog loss or more dramatic neofunctionalization

have occurred in other core eudicots. The phylogenetic findings do not undermine our understanding of the functional homology of *PLE* and *AG* since their highly similar functions were clearly inherited from a common ancestor. Their paralogous relationship does underscore the fluid nature of functional evolution following gene duplication and demonstrates the importance of evaluating genetic orthology and functional homology as separate entities (THEISSEN 2002).

Interestingly, gene duplication events have also been identified in the *AP3* and *AP1* gene lineages close to the base of the core eudicots (KRAMER *et al.* 1998; LITT and IRISH 2003). In the case of *AP3*, a gene duplication gave rise to the eu*AP3* and *TM6* paralogous lineages while in *AP1* an event produced the eu*AP1* and eu*FUL* lineages. Further sampling in lower eudicot taxa will be necessary to determine whether the *AG*, *AP3*, and *AP1* duplications were coincident. Unlike *AP3* and *AP1*, which underwent dramatic changes in otherwise conserved motifs following their lower eudicot duplications (KRAMER *et al.* 1998; LITT and IRISH 2003), there are comparatively few fixed differences between the eu*AG* and *PLE* lineages. It does appear, however, that the base of the core eudicots was a critical period in angiosperm evolution with many significant changes in both floral morphology (ENDRESS 1990) and the gene lineages that control floral organ identity.

Gene duplications have also shaped the evolution of the C lineage in the grass family: Approximately 50–70 MYA (GAUT 2002), a gene duplication predating the last common ancestor of Zea, Hordeum (rye), Triticum (wheat), and Oryza gave rise to the paralogous lineages defined by the Zea genes *ZAG1* and *ZMM2* (SCHMIDT and AMBROSE 1998). This was followed by a segmental allo-tetraploidization event in the Zea lineage (GAUT and DOEBLEY 1997), which produced the *ZMM2/ZMM23* paralog pair (MUNSTER *et al.* 2002). The expression patterns of *ZAG1* and *ZMM2* indicate that the paralogs have become subfunctionalized, with *ZAG1* more strongly expressed in carpels and *ZMM2* in stamens (MENA *et al.* 1996). However, the phenotype of plants with insertional mutations in *ZAG1* (MENA *et al.* 1996) indicates that carpel identity is redundantly controlled, possibly by other *AG*-like genes or novel factors similar to the *DROOPING LEAF* locus identified in rice (NAGA-SAWA *et al.* 2003). In Oryza, it remains unclear as to whether the *ZMM2* ortholog *OsMADS3* participates in all aspects of C function, as suggested by antisense transgenic lines (KANG *et al.* 1998), or primarily promotes stamen identity, as indicated by ectopic expression of the gene (KYOZUKA and SHIMAMOTO 2002). An Oryza *ZAG1* ortholog has not yet been annotated in the genome, but in the closely related Hordeum, orthologs of both *ZAG1* and *ZMM2* have been identified. It will be interesting to learn whether these genes are subfunctionalized in a manner similar to the Zea genes or show a different pattern of functional evolution. In general,

it is notable that subfunctionalization appears to be the trend for C-lineage paralogs in both the core eudicots and grasses.

The D lineage is defined by distinct aspects of protein sequence and expression pattern: *Can a distinct function be defined for the D lineage?* The concept of D function was first proposed on the basis of functional studies of the *FBP7* and *FBP11* genes in Petunia. The elimination of *FBP7/11* expression results in the transformation of ovules into pistil-like structures, while ectopic expression of *FBP7* results in the production of ovules on the sepals and, occasionally, the petals (ANGENENT *et al.* 1995; COLOMBO *et al.* 1995). These results were taken to indicate that the genes could promote ovule identity in disassociation from carpel identity, thereby requiring a fourth class of gene activity (COLOMBO *et al.* 1995). In contrast, analysis of the D-lineage member from Arabidopsis, *STK*, has shown that ovule identity is promoted by the combined activity of both C and D orthologs (FAVARO *et al.* 2003; PINYOPICH *et al.* 2003). These contrasting results may be due, in part, to different derivations of the placenta, which initiates as free central in Petunia (ANGENENT *et al.* 1995) but is marginal in Arabidopsis (GASSER and ROBINSON-BEERS 1993). In any case, the Arabidopsis results indicate that an absolute separation of C- and D-lineage functions is not universally applicable. Although the concept of a D function has been embraced in the literature (THEISSEN *et al.* 2000, 2002; FAVARO *et al.* 2002; TZENG *et al.* 2002), which we recognize by our designation of the D lineage, the control of ovule development might also be considered a component of C function *sensu lato*. Consistent with this argument, several C-lineage members have independently acquired primarily ovule-specific expression patterns, including *SHP1/2* in Arabidopsis (MA *et al.* 1991), *CAG2* in Cucumis (PERL-TREVES *et al.* 1998), and *ThdAG2* in Thalictrum (V. S. DI STILIO and E. M. KRAMER, unpublished results). At the same time, it must be acknowledged that almost all of the D-lineage members characterized to date, including core eudicot and grass orthologs, exhibit ovule-specific expression (SCHMIDT *et al.* 1993; ANGENENT *et al.* 1995; LOPEZ-DEE *et al.* 1999; BOSS *et al.* 2002; TZENG *et al.* 2002; PINYOPICH *et al.* 2003), the one exception being *CAG1/CUM10* from Cucumis (KATER *et al.* 1998). Therefore, although C-lineage members have retained the potential to contribute to ovule identity, the ancient C/D duplication event does appear to have been followed by a restriction of expression in the D lineage such that the genes generally do not function in male sporogenic tissue. Further studies of the expression patterns and functions of D orthologs will be necessary to clarify the conservation of their role in ovule development and to determine whether they typically function in an exclusive manner, as in Petunia, or in a redundant one, as in Arabidopsis.

Was the C/D gene duplication significant for the evolution of the angiosperms? Given that all gymnosperm

AG-like genes examined to date are expressed in microsporophylls, megasporophylls, and ovules (RUTLEDGE *et al.* 1998; TANDRE *et al.* 1998; WINTER *et al.* 1999), the ovule-specific expression of the D lineage can be described as a subfunctionalization (JAGER *et al.* 2003). It remains unclear as to whether or not substantial redundancy in the ovule identity program has always existed between C and D orthologs. If a large degree of redundancy has always been present, the D lineage may have been retained primarily to provide genetic buffering in the crucial ovule development pathway. In either case, however, the presence of ovule-specific D-lineage genes could also have increased the degree of dissociation between the sporophyll and ovule genetic pathways. Consider the apparent gymnosperm identity programs: microsporophylls are encoded by two independent gene lineages, *AP3/PI*-like and *AG*-like, while both megasporophylls and ovules are controlled by *AG*-like alone. Within this context, the evolution of the D paralogs created an alternative genetic source for elaboration of ovule morphology. This remains true even if the C and D genes remained redundant to some degree, analogous to the *SHP* genes promoting novel aspects of Arabidopsis carpel morphology. A complete subfunctionalization could have had more profound effects on both megasporophyll and ovule evolution. It has been recognized that subfunctionalization can free paralogs to adapt specifically to narrow functional repertoires (HUGHES 1994, 1999; ZHANG 2003). Similarly, the process can increase the modularity of whole genetic pathways, which allows characters to evolve without pleiotropic effects (RAFF 1996; WAGNER and ALTENBERG 1996). The suggestion has been made that greater modularity, at any hierarchical level, is a kind of key innovation that may be associated with radiations in diversity (YANG 2001). Therefore, it is possible that subfunctionalization between the C and D lineages decoupled megasporophyll and ovule development, facilitating evolutionary modifications of both structures.

Overall, this analysis of the *AG* subfamily has demonstrated the dynamic nature of functional evolution following gene duplication and underscores the importance of conducting both phylogenetic and functional analyses of gene lineages. It is also quite clear that the current extent of our knowledge regarding the functions of *AG*-like genes is entirely restricted to the core eudicots and grasses. To achieve a more thorough understanding of the evolution of the *AG* subfamily, it is critical to obtain functional data for C- and D-lineage members from intervening angiosperm lineages.

LITERATURE CITED

ALFARO, M. E., S. ZOLLER and F. LUTZONI, 2003 Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. **20:** 255–266.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. H. 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. PELAZ, S. J. LILJEGREN, S. E. GOLD, C. BURGEFF *et al.*, 2000 An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA **97:** 5328–5333.

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.

ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

BOSS, P. K., E. SENSI, C. HUA, C. DAVIES and M. R. THOMAS, 2002 Cloning and characterization of grapevine (*Vitis vinifera* L.) MADS-box genes expressed during inflorescence and berry development. Plant Sci. **162:** 887–895.

BOWMAN, J. L., D. R. SMYTH and E. M. MEYEROWITZ, 1989 Genes directing flower development in *Arabidopsis*. Plant Cell **1:** 37–52.

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

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.

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

BRUNNER, A. M., W. H. ROTTMANN, L. A. SHEPPARD, K. KRUTOVSKII, S. P. DIFAZIO *et al.*, 2000 Structure and expression of duplicate *AGAMOUS* orthologues in poplar. Plant Mol. Biol. **44:** 619–634.

CARPENTER, R., and E. S. COEN, 1990 Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. Genes Dev. **4:** 1483–1493.

CHOISNE, N., G. ORJEDA, L. CATTOLICO, N. DEMANGE, P. WINCKER *et al.*, 2002 *Oryza sativa* chromosome 12 sequencing. Genoscope.

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

COLOMBO, L., J. FRANKEN, E. KOETJE, J. VAN WENT, H. J. M. DONS *et al.*, 1995 The *Petunia* MADS box gene FBP11 determines ovule identity. Plant Cell **7:** 1859–1868.

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.

DOUADY, C. J., F. DELSUC, Y. BOUCHER, W. F. DOOLITTLE and E. J. P. DOUZERY, 2003 Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. **20:** 248–254.

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.

ENDRESS, P. K., 1990 Patterns of floral construction in ontogeny and phylogeny. Biol. J. Linn. Soc. **39:** 153–175.

FAVARO, R., R. G. H. IMMINK, V. FERIOLI, B. BERNASCONI, M. BYZOVA *et al.*, 2002 Ovule-specific MADS-box proteins have conserved protein-protein interactions in monocot and dicot plants. Mol. Genet. Genomics **268:** 152–159.

FAVARO, R., A. PINYOPICH, R. BATTAGLIA, M. KOOIKER, L. BORGHI *et al.*, 2003 MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. Plant Cell **15:** 2603–2611.

FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y.-L. YAN *et al.*, 1999 Preservation of duplicate genes by complementary, degenerate mutations. Genetics **151:** 1531–1545.

Gasser, C. S., and K. Robinson-Beers, 1993 Pistil development. Plant Cell **5:** 1231–1239.

Gaut, B. S., 2002 Evolutionary dynamics of grass genomes. New Phytol. **154:** 15–28.

Gaut, B. S., and J. Doebley, 1997 DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA **94:** 6809–6814.

Gustafson-Brown, C., B. Savidge and M. F. Yanofsky, 1994 Regulation of the *Arabidopsis* homeotic gene APETALA1. Cell **76:** 131–143.

Hasebe, M., and J. A. Banks, 1997 Evolution of MADS gene family in plants, pp. 179–197 in *Evolution and Diversification of Land Plants*, edited by K. Iwatsuki and P. H. Raven. Springer, Tokyo.

Hasebe, M., C.-K. Wen, M. Kato and J. A. Banks, 1998 Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. Proc. Natl. Acad. Sci. USA **95:** 6222–6227.

Henschel, K., R. Kofuji, M. Hasebe, H. Saedler, T. Munster *et al.*, 2002 Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. Mol. Biol. Evol. **19:** 801–814.

Honma, T., and K. Goto, 2001 Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409:** 525–529.

Huelsenbeck, J. P., and F. Ronquist, 2001 MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:** 754–755.

Huelsenbeck, J., and F. Ronquist, 2002 MrBayes v3. Uppsala University, Uppsala, Sweden.

Hughes, A. L., 1994 The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. Ser. B Biol. Sci. **256:** 119–124.

Hughes, A. L., 1999 *Adaptive Evolution of Genes and Genomes*. Oxford University Press, Oxford.

Hughes, M. L., and A. L. Hughes, 1993 Evolution of duplicate genes in a tetraploid animal, Xenopus laevis. Mol. Biol. Evol. **10:** 1360–1369.

Jager, M., A. Hassanin, M. Manuel, H. Le Guyader and J. Deutsch, 2003 MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. Mol. Biol. Evol. **20:** 842–854.

Johansen, B., L. B. Pedersen, M. Skipper and S. Frederikson, 2002 MADS-box gene evolution: structure and transcription patterns. Mol. Phylogenet. Evol. **23:** 458–480.

Kang, H.-G., Y.-S. Noh, Y.-Y. Chung, M. A. Costa, K. An *et al.*, 1995 Phenotypic alterations of petal and sepal by ectopic expression of a rice MADS box gene in tobacco. Plant Mol. Biol. **29:** 1–10.

Kang, H.-G., J.-S. Jeon, S. Lee and G. An, 1998 Identification of class B and class C floral organ identity genes from rice plants. Plant Mol. Biol. **38:** 1021–1029.

Kapoor, M., S. Tsuda, Y. Tanaka, T. Mayama, Y. Okuyama *et al.*, 2002 Role of *Petunia pMADS3* in determination of floral organ meristem identity, as revealed by its loss of function. Plant J. **32:** 115–127.

Kater, M. M., L. Colombo, J. Franken, M. Busscher, S. Masiero *et al.*, 1998 Multiple *AGAMOUS* homologs from cucumber and *Petunia* differ in their ability to induce reproductive organ fate. Plant Cell **10:** 171–182.

Kempin, S. A., M. A. Mandel and M. F. Yanofsky, 1993 Conversion of perianth into reproductive organs by ectopic expression of the tobacco floral homeotic gene *NAG1*. Plant Physiol. **103:** 1041–1046.

Kishino, H., and M. Hasegawa, 1989 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29:** 170–179.

Kramer, E. M., and V. F. Irish, 2000 Evolution of the petal and stamen developmental programs: evidence from comparative studies of the lower eudicots and basal angiosperms. Int. J. Plant Sci. **161:** S29–S40.

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

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

Kyozuka, J., and K. Shimamoto, 2002 Ectopic expression of *Os-MADS3*, a rice ortholog of *AGAMOUS*, caused a homeotic transformation of lodicules to stamens in transgenic rice plants. Plant Cell Physiol. **43:** 130–135.

Lamb, R. S., T. A. Hill, Q. K. Tan and V. F. Irish, 2002 Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. Development **129:** 2079–2086.

Lee, M. M., and J. Schiefelbein, 2001 Developmentally distinct MYB genes encode functionally equivalent proteins in *Arabidopsis*. Development **128:** 1539–1546.

Liljegren, S. J., G. S. Ditta, Y. Eshed, B. Savidge, J. L. Bowman *et al.*, 2000 *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature **404:** 766–770.

Litt, A., and V. F. Irish, 2003 Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. Genetics **165:** 821–833.

Lopez-Dee, Z. P., P. Wittich, M. E. Pe, D. Rigola, I. del Buono *et al.*, 1999 *OsMADS13*, a novel rice MADS-box gene expressed during ovule development. Dev. Genet. **25:** 237–244.

Lynch, M., and A. Force, 2000 The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459–473.

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.

Maddison, D. R., and W. P. Maddison, 2000 *MacClade 4.0: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, MA.

Magallon, S., P. R. Crane and P. S. Herendeen, 1999 Phylogenetic pattern, diversity, and diversification of eudicots. Ann. Mo. Bot. Gard. **86:** 297–372.

Mandel, M. A., J. L. Bowman, S. A. Kempin, H. Ma, E. M. Meyerowitz *et al.*, 1992 Manipulation of flower structure in transgenic tobacco. Cell **71:** 133–143.

Mena, M., B. A. Ambrose, R. B. Meeley, S. P. Briggs, M. F. Yanofsky *et al.*, 1996 Diversification of C-function activity in maize flower development. Science **274:** 1537–1540.

Mizukami, Y., and H. Ma, 1992 Ectopic expression of the floral homeotic gene *agamous* in transgenic *Arabidopsis* plants alters floral organ identity. Cell **71:** 119–131.

Mizukami, Y., H. Huang, M. Tudor, Y. Hu and H. Ma, 1996 Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations. Plant Cell **8:** 831–845.

Munster, T., J. Pahnke, A. Di Rosa, J. T. Kim, W. Martin *et al.*, 1997 Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. Proc. Natl. Acad. Sci. USA **94:** 2415–2420.

Munster, T., W. Deleu, L. U. Wingen, M. Ouzunova, J. Cacharron *et al.*, 2002 Maize MADS-box genes galore. Maydica **47:** 287–301.

Muse, S. V., and B. S. Gaut, 1994 A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11:** 715–724.

Nagasawa, N., M. Miyoshi, Y. Sano, H. Satoh, H. Hirano *et al.*, 2003 SUPERWOMAN1 and DROOPING LEAF genes control floral organ identity in rice. Development **130:** 705–718.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Heidelberg, Germany.

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

Perl-Treves, R., A. Kahana, N. Rosenman, Y. Xiang and L. Silberstein, 1998 Expression of multiple *AGAMOUS*-like genes in male and female flowers of cucumber (*Cucumis sativus* L.). Plant Cell Physiol. **39:** 701–710.

Pinyopich, A., G. S. Ditta, B. Savidge, S. J. Liljegren, E. Baumann *et al.*, 2003 Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature **424:** 85–88.

Pnueli, L., D. Harevan, S. Rounsley, M. F. Yanofsky and E. Lifschitz, 1994 Isolation of the tomato *AGAMOUS* gene *TAG1* and analysis of its homeotic role in transgenic plants. Plant Cell **6:** 163–173.

POSADA, D., and K. A. CRANDALL, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

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.

QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. A. SOLTIS *et al.*, 1999 The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature **402:** 404–407.

RAFF, R. A., 1996 *The Shape of Life.* The University of Chicago Press, Chicago.

RIECHMANN, J. L., B. A. KRIZEK and E. M. MEYEROWITZ, 1996a Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc. Natl. Acad. Sci. USA **93:** 4793–4798.

RIECHMANN, J. L., M. WANG and E. M. MEYEROWITZ, 1996b DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res. **24:** 3134–3141.

RUTLEDGE, R., S. REGAN, O. NICOLAS, P. FOBERT, C. COTE *et al.*, 1998 Characterization of an *AGAMOUS* homologue from the conifer black spruce (Picea mariana) that produces floral homeotic conversions when expressed in *Arabidopsis*. Plant J. **15:** 625–634.

SAVIDGE, B., S. D. ROUNSLEY and M. F. YANOFSKY, 1995 Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. Plant Cell **7:** 721–733.

SCHMIDT, R. J., and B. A. AMBROSE, 1998 The blooming of grass flower development. Curr. Opin. Plant Biol. **1:** 60–67.

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.

SHORE, P., and A. D. SHARROCKS, 1995 The MADS-box family of transcription factors. Eur. J. Biochem. **229:** 1–13.

SKAER, N., D. PISTILLO, J.-M. GIBERT, P. LIO, C. WULBECK *et al.*, 2002 Gene duplication at the achaete-scute complex and morphological complexity of the peripheral nervous system in Diptera. Trends Genet. **18:** 399–405.

SOLTIS, D. E., A. E. SENTERS, M. ZANIS, S. KIM, J. D. THOMPSON *et al.*, 2003 Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. Am. J. Bot. **90:** 461–470.

SUZUKI, Y., G. V. GLAZKO and M. NEI, 2002 Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA **99:** 16138–16143.

SVENSSON, M., and P. ENGSTROM, 2002 Closely related MADS-box genes in club moss (*Lycopodium*) show broad expession patterns and are structurally similar to, but phylogenetically distinct from typical seed plant MADS-box genes. New Phytol. **154:** 439–450.

SVENSSON, M., H. JOHANNESSON and P. ENGSTROM, 2000 The *LAMB1* gene from the clubmoss, *Lycopodium annotinum*, is a divergent MADS-box gene, expressed specifically in sporogenic structures. Gene **253:** 31–43.

SWOFFORD, D. L., 2001 *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.

TANDRE, K., M. SVENSON, M. SVENSSON and P. ENGSTROM, 1998 Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. Plant J. **15:** 615–623.

TEMPLETON, A. R., 1983 Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution **37:** 221–244.

THEISSEN, G., 2000 Shattering developments. Nature **404:** 711–713.

THEISSEN, G., 2002 Secret life of genes. Nature **415:** 741.

THEISSEN, G., and H. SAEDLER, 2001 Floral quartets. Nature **409:** 469–471.

THEISSEN, G., T. STRATER, A. FISCHER and H. SAEDLER, 1995 Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of *AGAMOUS*-like MADS-box genes from maize. Gene **156:** 155–166.

THEISSEN, G., J. T. KIM and H. SAEDLER, 1996 Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J. Mol. Evol. **43:** 484–516.

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.

THEISSEN, G., A. BECKER, K. U. WINTER, T. MUNSTER, C. KIRCHNER *et al.*, 2002 How the land plants learned their floral ABCs: the role of MADS-box genes in the evolutionary origin of flowers, pp. 173–205 in *Developmental Genetics and Plant Evolution*, edited by Q. C. B. CRONK, R. M. BATEMAN and J. A. HAWKINS. Taylor & Francis, London.

TZENG, T. Y., H.-Y. CHEN and C.-H. YANG, 2002 Ectopic expession of carpel-specific MADS box genes from *Lily* and *Lisianthus* causes similar homeotic conversion of sepal and petal in *Arabidopsis*. Plant Physiol. **130:** 1827–1836.

VANDENBUSSCHE, M., G. THEISSEN, Y. VAN DE PEER and T. GERATS, 2003 Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. Nucleic Acids Res. **31:** 4401–4409.

WAGNER, G. P., and L. ALTENBERG, 1996 Complex adaptations and the evolution of evolvability. Evolution **50:** 967–976.

WESTERN, T. L., and G. W. HAUGHN, 1999 BELL1 and AGAMOUS genes promote ovule identity. Plant Cell **7:** 1859–1868.

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.

YANG, A. S., 2001 Modularity, evolvability, and adaptive radiations: a comparison of the hemi- and holometabolous insects. Evol. Dev. **3:** 59–72.

YANG, Y., L. FANNING and T. JACK, 2003 The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. Plant J. **33:** 47–59.

YANOFSKY, M. F., H. MA, J. L. BOWMAN, G. N. DREWS, K. A. FELDMANN *et al.*, 1990 The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature **346:** 35–39.

YU, D., M. KOTILAINEN, E. POLLANEN, M. MEHTO, P. ELOMAA *et al.*, 1999 Organ identity genes and modified patterns of flower development in *Gerbera hybrida* (Asteraceae). Plant J. **17:** 51–62.

ZANIS, M., D. E. SOLTIS, P. S. SOLTIS, S. MATHEWS and M. J. DONOGHUE, 2002 The root of the angiosperms revisited. Proc. Natl. Acad. Sci. USA **99:** 6848–6853.

ZENG, A., A. M. HERNDON and J. C. HU, 1997 Buried asparagines determine the dimerization specificities of leucine zipper mutants. Proc. Natl. Acad. Sci. USA **94:** 3673–3678.

ZHANG, J., 2003 Evolution by gene duplication: an update. Trends Ecol. Evol. **18:** 292–298.