AMERICAN JOURNAL OF

BOTANY

DECEMBER 2004 VOLUME 91 NUMBER 12

Official Publication of the Botanical Society of America, Inc.

AMERICAN JOURNAL OF BOTANY

"the journal for all plant biologists"

The mission of the society is to promote botany, the field of basic science dealing with the study and inquiry into the form, function, diversity, reproduction, evolution, and uses of plants and their interactions within the biosphere.

December 2004
Volume 91, Number 12

BOTANICAL SOCIETY OF AMERICA

Cover Illustration: Female flowers of *Amborella trichopoda*. *Amborella*, a candidate for the basal-most angiosperm, and B-function MADS-box genes are the focus of a study by Kim et al. (see the summary in the box below and the full article on p. 2102). Photo credit: Sangtae Kim.

In this issue. . .

MADS-box genes and the basal-most angiosperm

Kim et al. examined gene structure and phylogeny of B-function MADS-box genes to provide additional evidence to support the basal-most status of *Amborella* among angiosperms. In particular, the *AP3*- and *PI*-homologues of *Amborella*, *Am.tr.AP3* and *Am.tr.PI*, share several amino acid strings, including a prominent DEAER motif in the C-terminal domain, the most variable transcribed region. Duplication time of two major lineages of B-function genes was estimated to be ca. 260 million years ago, placing the duplication after the split between extant gymnosperms and angiosperms, but well before the oldest angiosperm fossils.

(see p. 2102)

Recovery from drought stress and fungal endophytes

Lolium perenne, a common perennial grass widely used for forage and turf, is often host to a fungal endophyte growing between its leaf cells. In Cheplick's study of endophyte infection on host recovery from severe drought stress, uninfected plants, both drought-stressed and controls produced more tillers, and greater leaf area and mass, than did infected plants. Although the grass–endophyte symbiosis is mutualistic in other well-studied host–fungus pairs, the relationship between *L. perenne* and its endophyte may primarily benefit the fungus, while being detrimental to the host under certain environmental conditions.

(see p. 1960)

AFLP markers to detect a cryptic species in an endangered community

The Little Aguja Pondweed (*Potamogeton clystocarpus*) is restricted to a single drainage in west Texas, where it is nearly morphologically indistinguishable from two more widespread pondweed species. Whittall et al. used AFLPs and DNA sequences from the nuclear and chloroplast genome to examine the degree of genetic distinctiveness, hybridization, and clonality in this endangered pondweed community. Seventy-seven percent of the AFLP variation separates *P. clystocarpus* from its sister species. Hybrids were only detected between the widespread species and clonality remained localized to individual pools. These results indicate *P. clystocarpus* is a genetically distinct cryptic species.

(see p. 1960)

Architectural mimicry of sexual dimorphism is *Solanum*

In the vast majority of taxa with unisexual flowers, male and female flowers differ significantly in size. Most literature on this topic has focused on adaptive explanations for sexual size dimorphism. In *Solanum*, however, Diggle and Miller found that size differences between flower types are due to floral architecture; that is, the different flower types are spatially segregated within inflorescences, and location greatly affects flower size, irrespective of sexual function. Thus, in *Solanum*, architectural effects mimic sexual size dimorphism. To the extent that spatial segregation of flower types is common among taxa with unisexual flowers, such architectural "mimicry" may underlie other cases of sexual dimorphism.

(see p. 2024)

TABLE OF CONTENTS

American Journal of Botany 91(12): 2102–2118. 2004.

PHYLOGENY AND DIVERSIFICATION OF B-FUNCTION MADS-BOX GENES IN ANGIOSPERMS: EVOLUTIONARY AND FUNCTIONAL IMPLICATIONS OF A 260-MILLION-YEAR-OLD DUPLICATION[1]

SANGTAE KIM,[2,6] MI-JEONG YOO,[2] VICTOR A. ALBERT,[3] JAMES S. FARRIS,[4] PAMELA S. SOLTIS,[5] AND DOUGLAS E. SOLTIS[2,6]

[2]Department of Botany, University of Florida, Gainesville, Florida 32611 USA; [3]The Natural History Museums and Botanical Garden, University of Oslo, P.O. Box 1172 Blindern, NO-0318 Oslo, Norway; [4]Molekylärsystematiska laboratoriet, Naturhistoriska riksmuseet, Box 50007, SE-104 05 Stockholm, Sweden; and [5]Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611 USA

B-function MADS-box genes play crucial roles in floral development in model angiosperms. We reconstructed the structural and functional implications of B-function gene phylogeny in the earliest extant flowering plants based on analyses that include 25 new *AP3* and *PI* sequences representing critical lineages of the basalmost angiosperms: *Amborella*, *Nuphar* (Nymphaeaceae), and *Illicium* (Austrobaileyales). The ancestral size of exon 5 in *PI*-homologues is 42 bp, typical of exon 5 in other plant MADS-box genes. This 42-bp length is found in *PI*-homologues from *Amborella* and Nymphaeaceae, successive sisters to all other angiosperms. Following these basalmost branches, a deletion occurred in exon 5, yielding a length of 30 bp, a condition that unites all other angiosperms. Several shared amino acid strings, including a prominent ''DEAER'' motif, are present in the AP3- and PI-homologues of *Amborella*. These may be ancestral motifs that were present before the duplication that yielded the AP3 and PI lineages and subsequently were modified after the divergence of *Amborella*. Other structural features were identified, including a motif that unites the previously described TM6 clade and a deletion in *AP3*-homologues that unites all Magnoliales. Phylogenetic analyses of *AP3*- and *PI*-homologues yielded gene trees that generally track organismal phylogeny as inferred by multigene data sets. With both AP3 and PI amino acid sequences, *Amborella* and Nymphaeaceae are sister to all other angiosperms. Using nonparametric rate smoothing (NPRS), we estimated that the duplication that produced the *AP3* and *PI* lineages occurred approximately 260 mya (231–290). This places the duplication after the split between extant gymnosperms and angiosperms, but well before the oldest angiosperm fossils. A striking similarity in the multimer-signalling C domains of the *Amborella* proteins suggests the potential for the formation of unique transcription-factor complexes. The earliest angiosperms may have been biochemically flexible in their B function and ''tinkered'' with floral organ identity.

Key words: *Amborella*; *AP3*; B-class; basal angiosperms; MADS; *PI*.

MADS (***M****CM1*, ***A****GAMOUS*, ***D****EFICIENS*, and ***S****RF*)-box genes, which encode transcription factors, play important roles in the development of plants, animals, and fungi (Schwarz-Sommer et al., 1990; Shore and Sharrocks, 1995; Theißen and Saedler, 1995; Theißen et al., 1996, 2000; Riechmann and Meyerowitz, 1997; Becker and Theißen, 2003). Almost all known MADS-domain proteins from vascular plants share a conserved structural organization, the so-called MIKC-type domain structure, including MADS (M), intervening (I), keratin-like (K), and C-terminal (C) domains (Theißen et al., 1996, 2000; Münster et al., 1997; Becker and Theißen, 2003; but see Alvarez-Buylla et al., 2000). Phylogenetic analyses have revealed that the MADS-box gene family is composed of several well-defined clades (e.g., Doyle, 1994; Purugganan et al., 1995; Theißen et al., 1996, 2000; Purugganan, 1997; Münster et al., 1997; Johansen et al., 2002; Bodt et al., 2003; Pařenicová et al., 2003; Becker and Theißen, 2003); the members of each clade share similar expression patterns and highly related functions (Theißen and Saedler, 1995; Theißen et al., 1996, 2000; Becker and Theißen, 2003).

Some of the best-known MADS-box genes control the organ identity and initiation of the flower (Theißen et al., 1996, 2000; Theißen and Saedler, 2001). Studies in *Arabidopsis* and *Antirrhinum* have shown that MADS-box genes encode the B function of floral development, which together with A specifies petals, and with C discriminates stamens from carpels (C only) (Coen and Meyerowitz, 1991; Meyerowitz et al., 1991; Coen et al., 1991). In *Arabidopsis*, the B function genes are *AP3* and *PI*. Further complexity in the specification of floral organ identity involves the redundant functions of the *SEPALLATA* MADS-box genes, which are collectively required for petal, stamen, and carpel identities. Protein products of these genes dimerize, and the tetramer (or multimer) configuration is probably the active state in cells (Egea-Cortines et al., 1999; Honma and Goto, 2000; Ferrario et al., 2003).

Two subclades of B-function genes have been recognized, corresponding to *AP3*- and *PI*-homologues, respectively (reviewed in Theißen et al., 2000). Phylogenetic analyses have revealed a complex pattern of gene duplication and divergence in B-class genes. Kramer et al. (1998) and Kramer and Irish (2000) suggested that the ancestral gene of the AP3/PI lineages underwent duplication yielding the AP3 and PI lineages at some time well before the diversification of the angiosperms. Another major gene duplication, which yielded the TM6 and

[1] Manuscript received 9 December 2003; revision accepted 26 August 2004.

The authors thank E. Kramer for sharing an unpublished amino acid sequence of *Illicium* PI and M. Zanis for determining some sequences, M. Buzgo, S. Brockington, G. Theißen, and H. Ma for helpful comments and discussion, and J. Nam for allowing us to use his reverse-translation utility AA2DNA. This work was supported by NSF grant PGR-0115684, the University of Florida Research Foundation, and the Post-doctoral Fellowship Program of Korea Science and Engineering Foundation (KOSEF) for S. Kim.

[6] Authors for reprint requests. (e-mails: dsoltis@botany.ufl.edu and sangtae@botany.ufl.edu).

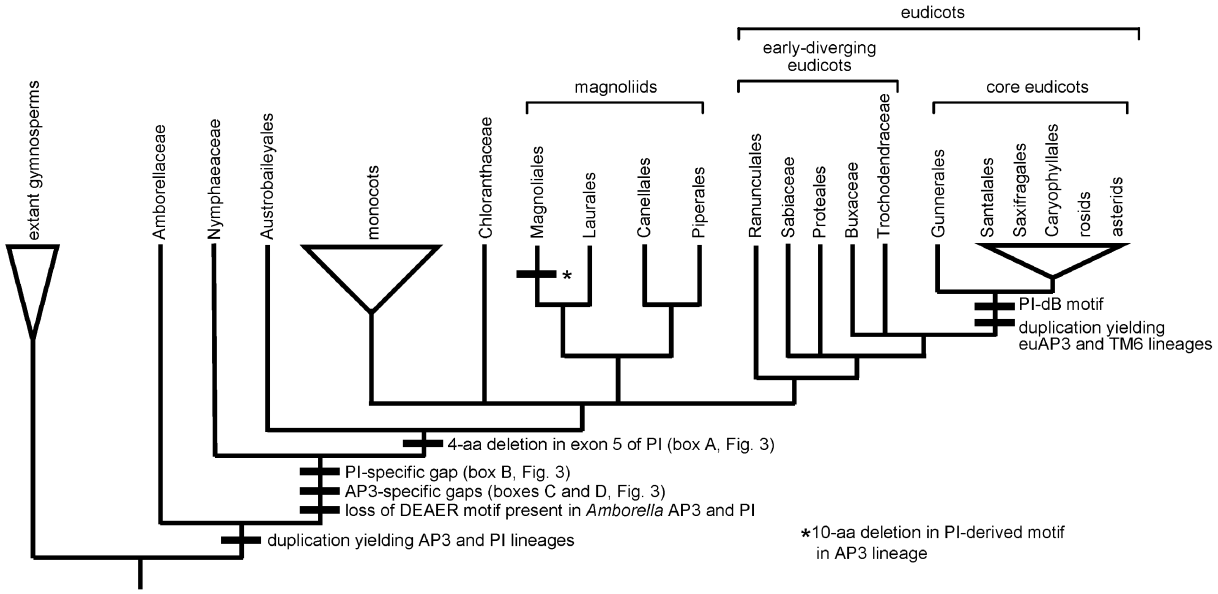

Fig. 1. Summary of angiosperm phylogeny (Qiu et al., 1999; P. Soltis et al., 1999; Zanis et al., 2002; D. Soltis et al., 2000). Some of the structural changes in AP3 and PI identified or localized in this study are plotted on the tree.

euAP3 lineages, is proposed to have occurred in the AP3 lineage just prior to the diversification of the core eudicots (Kramer et al., 1998; Kramer and Irish, 2000), a clade that comprises approximately 75% of all flowering plants (Drinnan et al., 1994).

PI-family genes have a highly conserved region of amino acid sequences at the C-terminal end called the ''PI motif.'' Members of the AP3 family have a less well-conserved ''PI derived motif,'' which can be aligned with the PI motif of the PI family. *AP3* genes also have a well-conserved ''euAP3'' or ''paleoAP3'' motif at the C-terminal end. A sequence similar to the paleoAP3 motif is also seen in B-class genes of the fern *Ceratopteris* (Münster et al., 1997). Therefore, the presence of the paleoAP3 motif is a synapomorphy (shared, derived state) of B- and related gene families (*GGM13*-, *GGM2*-, *DAL12*-, and *CJMADS1*-homologues), and the loss of the paleoAP3 motif is believed to be a synapomorphy for the PI lineage (Kramer et al., 1998; Winter et al., 2002a).

The exon lengths of MADS gene families are well conserved in most cases (Johansen et al., 2002). For example, exon 6 in both *AP3*- and *PI*-homologues is 45 bp long (Johansen et al., 2002; Winter et al., 2002a). Additionally, the *PI*-homologues reported to date are all distinguished from the *AP3*-homologues by an exon 5 that is 30 bp rather than 42 bp in length (Johansen et al., 2002; Winter et al., 2002a). In most other MADS gene families, exons 5 and 6 are generally 42 bp (Johansen et al., 2002; Winter et al., 2002a).

The overall framework of angiosperm phylogeny has crystallized (e.g., Qiu et al., 1999; D. Soltis et al., 2000; Zanis et al., 2002) (Fig. 1). A series of studies identified the same early-branching basal angiosperms (Mathews and Donoghue, 1999; Qiu et al., 1999; P. Soltis et al., 1999; Parkinson et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; D. Soltis et al., 2000; Zanis et al., 2002; Borsch et al., 2003): *Amborella* (Amborellaceae) is sister to all other extant angiosperms, either alone, or with the water lilies (Nymphaeaceae). Amborellaceae and Nymphaeaceae are followed by the Austrobaileyales clade (Fig. 1).

To understand the diversification of B-class genes during the evolutionary history of angiosperms, phylogenetic analyses using B-class genes from representatives of all major lineages of angiosperms are needed. However, sequences of the *AP3*- and *PI*-homologues identified to date only represent core- and early-diverging eudicots and some magnoliid taxa (Kramer et al., 1998; Kramer and Irish, 2000). Critical lineages of basal angiosperms, including all of the early-branching taxa (*Amborella*, Nymphaeaceae, and Austrobaileyales), have not yet been analyzed for B-class genes. Furthermore, only a few sequences have so far been obtained for other basal lineages, such as Chloranthaceae (e.g., *Chloranthus*) and the magnoliid clade of Laurales (avocado relatives), Piperales (black pepper relatives), Canellales (mountain pepper relatives), and Magnoliales (magnolia relatives). In the eudicots (Fig. 1), *AP3*- and *PI*-homologues have been isolated from several early-diverging eudicots (Papaveraceae and Buxaceae, the poppy and boxwood families, respectively), but in the core eudicots, only the rosid (rose relatives), Caryophyllales (beet relatives), and asterid (e.g., *Antirrhinum*) clades have so far been sampled. No data are available for Saxifragales (saxifrage relatives), Santalales (mistletoes), or Gunnerales (gunneras). The latter lineage has recently been shown to be sister to all other core eudicots (D. Soltis et al., 2003).

We sought to improve both the taxonomic coverage and our understanding of phylogenetic relationships of *AP3*- and *PI*-homologues among critical lineages of flowering plants. Our goals were to: (1) isolate and sequence both *AP3*- and *PI*-homologues from early-branching angiosperms, as well as from several previously unsampled lineages of core eudicots; (2) trace structural changes in the AP3 and PI lineages, including motif regions and lineage-specific insertions/deletions; (3) conduct phylogenetic analyses of B-class genes to determine how well these gene trees track organismal phylogeny based on results from recent multigene studies; and (4) estimate the timing of the duplication event that yielded the AP3 and PI lineages.

MATERIALS AND METHODS

RNA extraction—RNA extraction kits available commercially (e.g., RNA WIZ Kit from Promega, RNeasy Plant Mini Kit from Qiagen) are designed for use based on investigations of model plants (e.g., *Arabidopsis*). These kits have rarely been tested on other flowering plants, such as lineages of basal angiosperms. In our initial RNA extractions of basal angiosperm taxa (e.g., *Amborella* and *Nuphar*), we employed commercially available kits, and these efforts failed. We therefore developed our own protocol for isolating high yields of RNA from these difficult plant tissues. Our method combined parts of the CTAB DNA extraction protocol (Doyle and Doyle, 1987) with the subsequent use of the RNeasy Plant Mini Kit (Qiagen, Stanford, California, USA) (Detailed method provided at http://www.flmnh.ufl.edu/soltislab/soltis_lab_protocols.htm). This protocol has worked well in all of the basal angiosperm taxa that we have tested. One problem with this method, however, is DNA contamination. We therefore treated all samples with DNase (DNA free; Ambion, Austin, Texas, USA) before subsequent experiments.

RT-PCR and cDNA sequence determination—Reverse transcription was performed following the manufacturer's directions using SuperScript II RNase H-reverse transcriptase (Invitrogen, Carlsbad, California, USA) and polyT primer (5′-CCG GAT CCT CTA GAG CGG CCG $C(T)_{17}$-3′). PCR (polymerase chain reaction) was performed using a B-class gene-specific primer (5′-GGG GTA CCA AYM GIC ARG TIA CIT AYT CIA AGM GIM G-3′) and the polyT primer used in reverse transcription (Kramer et al., 1998). PCR conditions were those employed by Kramer et al. (1998). We also used additional primers for PCR: AP3-PIf (5′-GSK MGI GGI AAG ATC KAG AT-3′) and newAP3-PIf (5′-AAC MGG CAR GTG ACG TAY TC-3′) (designed by Michael Zanis). PCR bands over 800 bp in size were excised from the agarose gel and purified using the Geneclean II Kit (Q·Bio Gene, Carlsbad, California, USA). Purified DNAs were cloned using the TOPO TA Cloning Kit (Invitrogen). Plasmid DNAs were purified from cloned cells using the e.Z.N.A. Plasmid Miniprep Kit (Omega Bio-tek, Doraville, Georgia, USA). Cycle sequencing reactions were performed using the CEQ DTCS-Quick Start Kit (Beckman Coulter, Fullerton, California, USA), and cDNA sequences were determined using a CEQ 8000 sequencing system (Beckman Coulter). Because the positions of all B-class gene-specific primers that we used in this study were approximately 75 bp after the start codon, 5′ end sequences of each gene were not included in this study.

To determine the genomic DNA sequence of *Amborella trichopoda*, we designed specific primers for *Amborella PI*-homologues at the MADS domain (AF1P; 5′-AGC GGA ATA CTG AAG AAG GC-3′) and the 3′ end of the translating region (AF6P; 5′-TGC TGA AGA TTG GGT TGG-3′) for PCR and sequencing. We compared the genomic sequence with the cDNA sequence to determine the sizes of the exons and introns. PCR conditions for genomic DNA analyses followed Kim et al. (2001).

Alignment—A data set was constructed of 103 previously reported B-function genes obtained from GenBank and 24 new sequences that we produced (excluding *Am.tr.AP3-2*; see Results and Discussion) (Table 1). Three different alignments, referred to as alignments I, II, and III, were used for phylogenetic analyses. For all three matrices we first aligned the amino acid sequences and used this amino acid alignment to produce a data set of aligned nucleotide sequences. We then analyzed both nucleotide and amino acid sequences phylogenetically (see next section). Alignment I includes all available *AP3*- and *PI*-homologues from angiosperms. Amino acid sequences of these genes were aligned using CLUSTAL X (Thompson et al., 1997) with the default options and then adjusted manually. In alignment II, we divided sequences into small subgroups based on organismal relationships (e.g., basal angiosperms, Magnoliales, early-diverging eudicots, monocots) or sequence similarity (e.g., euAP3 and TM6). We aligned sequences within each subgroup first. We also placed both *AP3*- and *PI*-homologues of basal angiosperms (*Amborella*, *Nuphar*, and *Illicium*) in a subgroup because we recognized similar shared amino acid sequences between AP3 and PI of *Amborella*, including the ''DEAER'' motif. Based on aligned sequences of this subgroup, other subgroups were subsequently combined into the alignment using the ''profile alignment'' method in CLUSTAL X (file-to-file alignment). We adjusted previously reported motif regions following Kramer et al. (1998). The alignment of the MIK-region in alignment II was the same as that of alignment I; however, alignments I and II differed in the more variable C-domain region. Alignment III includes not only angiosperm sequences, but also three published gymnosperm B-class genes (Mouradov et al., 1998, 1999). These gymnosperm sequences were added to alignment II.

Aligned amino acids were converted to DNA sequences using the program AA2DNA (provided by Jongmin Nam). Separate matrices of DNA and amino acid sequences of alignments I, II, and III, for the entire gene (MIKC) and only the M-, I-, and K-domains were constructed for phylogenetic analyses.

Phylogenetic analyses—Three different methods of phylogenetic inference were used: maximum parsimony (MP), support weighting (Farris, 2001), and Bayesian (Huelsenbeck and Ronquist, 2001). The maximum parsimony analyses for all matrices were conducted using PAUP* 4.0b10 (Swofford, 2001). The search strategy involved 100 random addition replicates with TBR branch swapping saving all optimal trees. To assess support for each node, bootstrap analyses (Felsenstein, 1985) were performed using 100 bootstrap replicate heuristic searches with 10 random taxon addition replicates and TBR branch swapping saving all optimal trees.

We applied the support weighting method (Farris, 2001) to the MIKC-DNA-III and MIK-DNA-III alignments. This method measures the degree to which changes in a character (site) are concentrated in the supported branches of a tree. Jackknife resampling was used to generate randomly selected suites of initial weights in successive support weighting, and this provides a way of assessing the stability of successive weight results (Farris, 2001). Support values for the support weighting tree were generated by parsimony jackknifing (Farris et al., 1996) of the original data matrix using 1000 replicates.

Bayesian analyses were conducted using MrBayes 2.01 (Huelsenbeck and Ronquist, 2001) for MIKC-DNA-III and MIK-DNA-III. We used uniform prior probabilities and the general time-reversible + gamma + I model of molecular evolution. This model of molecular evolution was selected as the optimal model (Akaike information criterion) by ModelTest (Posada and Crandall, 1998). We ran four chains of Markov Chain Monte Carlo (MCMC), sampling every 1000 generations for 1 000 000 generations, starting with a random tree. Stationarity was reached at approximately generation 35 000; thus, the first 35 trees were the ''burn in'' of the chain, and phylogenetic inferences are based on those trees sampled after generation 35 000.

Estimation of divergence time—We calculated MP branch lengths and optimized these using PAUP* 4.0b10 onto a phylogenetic tree based on recent multigene studies (Qiu et al., 1999; P. Soltis et al., 1999; Zanis et al., 2002). Two data sets were analyzed: one contained the euAP3 genes, along with the paleoAP3 and PI genes, and the second contained the TM6, plus the paleoAP3 and PI genes. Separate analyses of the MIK and MIKC regions for each data set were conducted. Trees with branch lengths were transformed into ultrametric trees using nonparametric rate smoothing (NPRS) (Sanderson, 1997) as implemented in TREEEDIT (version 1.0 alpha 10 by A. Rambaut and M. Charleston). The characteristic pollen of the eudicots, combined with their extensive fossil record, places the origin of the eudicots at 125 mya (Hughes, 1994), one of the firmest dates in the paleobotanical record. This minimum age for eudicots was used to calibrate the tree. This calibration point was alternatively applied to the eudicot node in the AP3 and PI clades of each tree. To compute error estimates for the ages, we reapplied the NPRS procedure to 100 bootstrapped matrices obtained by resampling the data irrespective of codon position using PAUP*4.0b10 (cf. P. Soltis et al., 2002).

RESULTS AND DISCUSSION

We report 25 new B-class gene sequences, which have been cloned from nine taxa including seven basal angiosperms (*Amborella*, *Nuphar*, *Illicium*, *Persea*, *Asimina*, *Eupomatia*, and *Magnolia*) and two eudicots (*Gunnera* and *Ribes*) (Table 1). None of these sequences is closely related to the B-sister gene family (Becker et al., 2002); when we include B-sister

sequences in our analyses, our sequences were placed in the B-class clade instead of the B-sister clade. Because we obtained only an incomplete sequence from one of the *AP3*-homologues of *Amborella* (*Am.tr.AP3-2*), we excluded this sequence from the phylogenetic analyses and used it for sequence comparison only.

The 5′ ends of the translating region of B-class gene sequences have generally not been determined in previous studies because genes were amplified and sequenced using primers designed for the 5′ end of the translating region. Likewise, we did not determine the sequence of this region of most sequences that we report. In our data matrix, which consists of the 24 new full or nearly full-length sequences that we obtained (excluding *Am.tr.AP3-2*) and other sequences from GenBank, only 23 of 54 *PI*-homologues and 34 of 72 *AP3*-homologues include the complete 5′ end of the translating region. We may assume that all B-class genes have the same starting point at the 5′ end of the translating region in the aligned sequence matrix because the MADS domain is so strongly conserved (Ma and dePamphillis, 2000) and all previously published sequences having the complete 5′ end of the translating region have the same starting point in the aligned sequence matrix. The description we provide dealing with gene length is based on this reasonable assumption. The putative size ranges that we estimate for each gene family are 163–239 amino acids (aa) for PI-homologues and 195–261 aa for AP3-homologues. Thus, the average size of AP3-homologues is approximately 20 aa larger than PI-homologues.

Structural features—The AP3-homologues of Magnoliales that we report did not possess the PI-derived motif typical of AP3-homologues (Kramer et al., 1998; Kramer and Irish, 2000), but rather a deletion of 10 aa unique to this clade (Fig. 2). In the remaining AP3-homologues studied, the PI-derived motif was loosely aligned with the PI motif (Fig. 2). We recognized two different groups of sequences in the PI-derived motif. Group A (hereafter referred to as PI-derived A, abbreviated as PI-dA) occurs in all basal angiosperms, with the exception of Magnoliales, and in early-diverging eudicots (Fig. 2). The PI-dA sequences have almost the same number of amino acids as found in the PI motif (Fig. 2) and are easily aligned with the PI motif. The second group of the PI-derived motif, PI-derived B (abbreviated as PI-dB), occurs in all core eudicots (both TM6 and euAP3 lineages). These PI-dB sequences were difficult to align among themselves, as well as with the PI-dA sequences, and with the PI motif (Fig. 2). Less than half of the consensus sequence of the PI-derived motif (FxFRLQPxSQPNLH) (Kramer et al., 1998) is identical to the PI-dB sequences. The PI-dB group typically has many histidine (H) and asparagine (N) residues, which are not present (or are rare) in PI-dA (Fig. 2).

Kramer et al. (1998) described the paleoAP3 motif and its modification following an apparent duplication to form the TM6 and euAP3 clades. Their TM6 clade was recovered in phylogenetic analyses with low bootstrap support and possessed the paleoAP3 motif (Kramer et al., 1998). The TM6 lineage shares the paleoAP3 motif with basal angiosperms and early-diverging eudicots, and the PI-dB motif with the euAP3 gene lineage (Fig. 2). We discovered a region of 10 amino acids that represents a previously undescribed motif that characterizes hypothetical proteins previously referred to as TM6 clade members. This ''TM6 motif'' is located just before the PI-dB region with AVAFANGVxNL, the most frequent amino acids at each position (Fig. 2). The TM6 lineage is therefore easily identified by this newly recognized motif. Although Kramer et al. (1998) placed *CMB2* (*Dianthus*) in their TM6 lineage, the CMB2 aa sequence does not have the newly reported TM6 motif. In our phylogenetic analyses (see below: phylogenetic analyses) *CMB2* fluctuated in position and did not always appear with TM6 genes.

The translated *AP3*- and *PI*-homologues of *Amborella* (Am.tr.AP3 and Am.tr.PI) share a number of amino acid strings in the first half of exon 7 (also a part of the C-domain), which is the most variable region of B-class proteins (Fig. 3). In contrast, conspecific AP3- and PI-homologues from flowering plants other than *Amborella* are not easily alignable in this region. The alignment of AP3- and PI-homologues revealed a PI-specific gap (Fig. 3: box B) and two AP3-specific gaps (Fig. 3: boxes C and D).

Twelve amino acid sites, including a ''DEAER'' motif, which is a string of five amino acids, were identical between Am.tr.AP3 and Am.tr.PI in the 24 aligned amino acids of this variable region of exon 7 (Fig. 3). This shared aa sequence between the AP3- and PI-homologues of *Amborella* was not found in any other angiosperm, including *Nuphar* and *Illicium* (Fig. 3). In the absence of evidence for gene conversion acting among MADS-box genes, we believe that the most parsimonious interpretation of the DEAER motif is that it is an ancestral feature among flowering plant B-class proteins that was subsequently modified in both the AP3 and PI lineages leading to all other angiosperms. This is supported by our failure to locate the DEAER motif in gymnosperm B-class and B-sister protein sequences (Becker et al., 2002). Thus, this feature apparently evolved along the branch leading to flowering plants (its ancestral and unknown stem lineage, now extinct) because there is no indication that it predates the angiosperm-gymnosperm split. QxAL in exon 7 and LExQNK in exon 6 are supporting examples of shared amino acid strings found only in Am.tr.AP3 and Am.tr.PI (Fig. 3).

It is also possible that these shared aa strings are homoplasious, having arisen alternatively via convergent mutations in the two genes or gene conversion. The lack of such similarity between *AP3*- and *PI*-homologues of other species argues against convergent mutations in the *Amborella* genes. To test for the possibility that either ancient or recent gene conversion was responsible for the shared aa strings between *Amborella* AP3- and PI-homologues, we used GENECONV (Sawyer, 1989), with a range of mismatch penalties (gscale = 0, 1, 2). Similarities between the *Amborella* genes were not identified as significant ''fragments.'' We therefore conclude that the shared strings are most likely ancestral motifs. If the shared motifs are indeed ancestral between the AP3- and PI-homologues, our results support *Amborella* as the sister organism to all other extant angiosperms, a placement that has been well supported by plastid, mitochondrial, and nuclear rRNA genes (e.g., Mathews and Donoghue, 1999; Qiu et al., 1999; P. Soltis et al., 1999; Parkinson et al., 1999; Graham and Olmstead, 2000; D. Soltis et al., 2000; Zanis et al., 2002; Borsch et al., 2003; Hilu et al., 2003).

The size of exon 5 in all *PI*-homologues reported to date is 30 bp. However, exon 5 in other MADS genes (e.g., *AP3*-, *AP1*-, and *AG*-homologues) is generally 42 bp in length (Johansen et al., 2002; Winter et al., 2002a). A length of 30 bp for exon 5 was thought to be a general feature of all *PI*-homologues, but genomic DNA sequences of *PI*-homologues have been reported only for *PI* (Tröbner et al., 1992) and *GLO*

TABLE 1. Sequences used in this study. The GenBank accession numbers and the number of colonies that we sequenced are indicated for genes cloned in this study (highlighted in boldface). Angiosperm species are arranged by families and higher groups according to the APG II (2003) system.

Classification	Taxa	*PI*-homologues	*AP3*-homologues	Gymnosperm B-class	Reference
GYMNOSPERMS					
Coniferales					
Pinaceae	*Picea abies*			*DAL13-1* (AF158543)	Sundström et al. (1999) *Dev. Genetics* 25, 253.
				DAL13-2 (AF158544)	Sundström et al. (1999) *Dev. Genetics* 25, 253.
	Pinus radiata			*PrDGL* (AF120097)	Mouradov et al. (1999) *Dev. Genetics* 25, 245.
ANGIOSPERMS					
Amborellaceae	*Amborella trichopoda*	***Am.tr.PI* (AY337760) (29)**			This study
			***Am.tr.AP3-1* (AY337743) (3)**		This study
			***Am.tr.AP3-2* * (AY337744) (1)**		This study
Nymphaeaceae	*Nuphar advena*	***Nu.ad.PI* (AY337736) (5)**			This study
	Nuphar variegatum	***Nu.va.PI* (AY337737) (1)**			This study
			***Nu.va.AP3-1* (AY337745) (3)**		This study
			***Nu.va.AP3-2* (AY337746) (1)**		This study
Chloranthaceae	*Chloranthus spicatus*	*CsPI* (AF230710)			Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			CsAP3 (AF230700)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
Austrobaileyales					
Illiciaceae	*Illicium parviflorum*		***Il.pa.AP3* (AY337747) (1)**		This study
MAGNOLIIDS					
Laurales					
Calycanthaceae	*Calycanthus floridus*	*CfPI-1*(AF230708)			Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
		CfPI-2 (AF230709)			Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			CfAP3-1 (AF230699)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			CfAP3-2 (AF230701)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
Lauraceae	*Persea americana*	***Pe.am.PI* (AY337738) (3)**			This study
			***Pe.am.AP3* (AY337748) (3)**		This study
Magnoliales					
Annonaceae	*Asimina longifolia*	***As.lo.PI* (AY337739) (14)**			This study
			***As.lo.AP3* (AY337749) (11)**		This study
Eupomatiaceae	*Eupomatia bennettii*	***Eu.be.PI* (AY337740) (3)**			This study
			***Eu.be.AP3-1* (AY337750) (12)**		This study
			***Eu.be.AP3-2* (AY337751) (4)**		This study
	Eupomatia laurina	***Eu.la.PI* (AY337741) (8)**			This study

TABLE 1. Continued.

Classification	Taxa	*PI*-homologues	*AP3*-homologues	Gymnosperm B-class	Reference
Magnoliaceae	*Liriodendron tulipifera*	*LtPI* (AF052864)			Kramer et al. (1998) *Genetics* 149, 765.
			LtAP3 (AF052878)		Kramer et al. (1998) *Genetics* 149, 765.
	Magnolia grandiflora		***Ma.gr.AP3* (AY337752) (3)**		This study
	Magnolia kobus (=*praecocissima*)	*MpMADS8* (AB050650)	*MpMADS7* (AB050649)		Unpublished Unpublished
	Michelia figo	*MfPI* (AF052863)			Kramer et al. (1998) *Genetics* 149, 765.
			MfAP3 (AF052877)		Kramer et al. (1998) *Genetics* 149, 765.
Piperales					
Aristolochiaceae	*Asarum europaeum*	*AePI* (AF230707)			Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			AeAP3-1 (AF230697)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
Piperaceae	*Peperomia hirta*	*PhPI*(AF052865)			Kramer et al. (1998) *Genetics* 149, 765.
			PhAP3 (AF052879)		Kramer et al. (1998) *Genetics* 149, 765.
	Piper magnificum	*PmPI-1* (AF052866)			Kramer et al. (1998) *Genetics* 149, 765.
		PmPI-2 (AF052867)			Kramer et al. (1998) *Genetics* 149, 765.
Saururaceae	*Saururus chinensis*		*MADS651* (AY057378)		Unpublished
MONOCOTS					
Alismatales					
Alismataceae	*Sagittaria montevidensis*	*SmPI* (AF230712)			Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			SmAP3 (AF230705)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
Asparagales					
Hyacinthaceae	*Hyacinthus orientalis*	*HPI1* (AF134114)			Unpublished
		HPI2 (AF134115)			Unpublished
Orchidaceae	*Orchis italica*	*OrcPI* (AB094985)			Unpublished
Hemerocallidaceae	*Hemerocallis* hybrid		*HhMADS1* (AF209729)		Lange (1999). Dissertation. University of California, Davis.
Dioscoreales					
Taccaceae	*Tacca chantieri*	*TcPI* (AF230713)			Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			TcAP3 (AF230706)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
Liliales					
Liliaceae	*Lilium regale*	*LRGLOA* (AB071379)			Winter et al. (2002) *Mol. Biol. Evol.* 19, 587.
		LRGLOB (AB071380)			Winter et al. (2002) *Mol. Biol. Evol.* 19, 587.
			LRDEF (AB071378)		Winter et al. (2002) *Mol. Biol. Evol.* 19, 587.
	Lilium longiflorum		*LMADS1* (AF503913)		Tzeng and Yang (2001) *Plant Cell Physiol.* 42, 1156.
Poales					
Poaceae	*Oryza sativa*	*OSMADS2* (L37526)			Chung et al. (1995) *Plant Science* 109, 45.
		OSMADS4 (L37527)			Chung et al. (1995) *Plant Science* 109, 45.
		nmads1 (AF095645)			Yuan et al. (2000) *Prog. Nat. Sci.* 10, 357.
			OSMADS16 (AF077760)		Moon et al. (1999) *Plant Mol. Biol.* 40, 167.
	Triticum aestivum		*TaMADS51* (AB007506)		Murai et al. (1998) *Plant Phisol.* 118, 330.
	Zea mays	*ZMM16* (AJ292959)			Münster et al. (2001) *Gene* 262, 1.
		ZMM18 (AJ292960)			Münster et al. (2001) *Gene* 262, 1.
		ZMM29 (AJ292961)			Münster et al. (2001) *Gene* 262, 1.
			SILKY1 (AF181479)		Ambrose et al. (2000) *Molecular Cell* 5, 569.
EUDICOTS					
Buxaceae	*Pachysandra terminalis*		*PtAP3-1* (AF052870)		Kramer et al. (1998) *Genetics* 149, 765.
			PtAP3-2 (AF052871)		Kramer et al. (1998) *Genetics* 149, 765.

TABLE 1. Continued.

Classification	Taxa	*PI*-homologues	*AP3*-homologues	Gymnosperm B-class	Reference
Ranunculales					
Papaveraceae	*Dicentra eximia*	*DePI* (AF052857)			Kramer et al. (1998) *Genetics* 149, 765.
			DeAP3 (AF052875)		Kramer et al. (1998) *Genetics* 149, 765.
	Papaver californicum		*PcAP3* (AF052872)		Kramer et al. (1998) *Genetics* 149, 765.
	Papaver nudicaule	*PnPI-1* (AF052855)			Kramer et al. (1998) *Genetics* 149, 765.
		PnPI-2 (AF052856)			Kramer et al. (1998) *Genetics* 149, 765.
			PnAP3-1 (AF052873)		Kramer et al. (1998) *Genetics* 149, 765.
			PnAP3-2 (AF052874)		Kramer et al. (1998) *Genetics* 149, 765.
	Sanguinaria canadensis	*ScPI* (AF130871)			Kramer and Irish (1999) *Nature* 399, 144.
			ScAP3 (AF130868)		Kramer and Irish (1999) *Nature* 399, 144.
Ranunculaceae	*Caltha palustris*	*CpPI* (AF052858)			Kramer et al. (1998) *Genetics* 149, 765.
	Delphinium ajacis	*DaPI* (AF052862)			Kramer et al. (1998) *Genetics* 149, 765.
	Ranunculus bulbosus	*RbPI-1* (AF052859)			Kramer et al. (1998) *Genetics* 149, 765.
		RbPI-2 (AF052860)			Kramer et al. (1998) *Genetics* 149, 765.
			RbAP3-1 (AF052876)		Kramer et al. (1998) *Genetics* 149, 765.
			RbAP3-2 (AF130869)		Kramer and Irish (1999) *Nature* 399, 144.
	Ranunculus ficaria	*RfPI-1* (AF052858)			Kramer et al. (1998) *Genetics* 149, 765.
		RfPI-2 (AF130872)			Kramer and Irish (1999) *Nature* 399, 144.
			RfAP3-1 (AF052854)		Kramer et al. (1998) *Genetics* 149, 765.
			RfAP3-2 (AF130870)		Kramer and Irish (1999) *Nature* 399, 144.
Core eudicots					
Gunnerales					
Gunneraceae	*Gunnera tinctoria*		***Gu.ti.AP3-1* (AY337753) (2)**		This study
			***Gu.ti.AP3-2* (AY337754) (3)**		This study
			***Gu.ti.AP3-3* (AY337755) (3)**		This study
			***Gu.ti.AP3-4* (AY337756) (3)**		This study
			***Gu.ti.AP3-5* (AY337757) (3)**		This study
Caryophyllales					
Caryophyllaceae	*Dianthus caryophyllus*		*CMB2* (L40405)		Baudinette et al. (2000) *Plant Science* 155, 123.
	Silene latifolia	*SLM2* (X80489)			Hardenack et al. (1994) *Plant Cell* 6, 1775.
			SLM3 (X80490)		Hardenack et al. (1994) *Plant Cell* 6, 1775.
Polygonaceae	*Rumex acetosa*		*RAD1* (X89113)		Ainsworth et al. (1995) *Plant Cell 7*, 1583.
			RAD2 (X89108)		Ainsworth et al. (1995) *Plant Cell 7*, 1583.
Saxifragales					
Saxifragaceae	*Ribes sanguineum*	***Ri.sa.PI* (AY337742) (6)**			This study
			***Ri.sa.AP3-1* (AY337758) (2)**		This study
			***Ri.sa. AP3-2* (AY337759) (1)**		This study
ROSIDS					
Myrtales					
Myrtaceae	*Eucalyptus grandis*	*EGM2* (AF029976)			Southerton et al. (1998) *Plant Physiol.* 118, 365.
Eurosids I					
Curcurbitales					
Cucurbitaceae	*Cucumis sativus*	*CUM26* (AF043255)			Unpublished
Fabales					
Fabaceae	*Medicago sativa*		*NMH7* (L41727)		Heard and Dunn (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 5273.

TABLE 1. Continued.

Classification	Taxa	*PI*-homologues	*AP3*-homologues	Gymnosperm B-class	Reference
Fagales					
Juglandaceae	*Juglans regia*		*JrAP3* (AJ313089)		Unpublished
Malpighiales					
Salicaceae	*Populus trichocarpa*		*PTD* (AF057708)		Sheppard et al. (2000) *Plant Phys.* 124, 627.
Rosales					
Rosaceae	*Malus domestica*	*MdPI* (AJ291490)			Yao et al. (2001) *Proc. Natl. Acad. Sci. U.S.A.* 98, 1306.
			MdTM6 (AB081093)		Unpublished
			MdMADS13 (AJ251116)		Unpublished
	Rosa rugosa	*MASAKO BP* (AB038462)			Kitahara et al. (2001) *Plant Science* 161, 549.
			MASAKO B3 (AB055966)		Kitahara et al. (2001) *Plant Science* 161, 549.
Eurosids II Brassicales					
Brassicaceae	*Arabidopsis thaliana*	*PI* (D30807)			Goto and Meyerowitz (1994) *Genes Develop.* 8, 1548.
			AP3 (AF115814)		Purugganan & Suddith (1999) *EMBO J.* 151, 839.
	Brassica napus		*BnAP3* (AF124814)		Unpublished
	Brassica oleracea		*BobAP3*(U67456)		Carr and Irish (1997) *Planta* 201, 179.
			Boi1AP3 (U67453)		Carr and Irish (1997) *Planta* 201, 179.
			Boi2AP3 (U67455)		Carr and Irish (1997) *Planta* 201, 179.
ASTERIDS Cornales					
Hydrangeaceae	*Hydrangea macrophylla*	*HmPI* (AF230711)			Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			HmAP3 (AF230702)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
			HmTM6 (AF230703)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
Euasterids I Lamiales					
Oleaceae	*Syringa vulgaris*	*SvPI* (AF052861)			Kramer et al. (1998) *Genetics* 149, 765.
			SvAP3 (AF052869)		Kramer et al. (1998) *Genetics* 149, 765.
Plantaginaceae	*Antirrhinum majus*	*GLO* (X68831)			Tröbner et al. (1992) *EMBO J.* 11, 4693.
			DEF (X52023)		Sommer et al. (1990) *EMBO J.* 9, 605.
Solanales					
Solanaceae	*Lycopersicon esculentum*		*LeAP3* (AF052868)		Kramer et al. (1998) *Genetics* 149, 765.
			TM6 (X60759)		Pnueli et al. (1991) *Plant J.* 1, 255.
	Nicotiana tabacum	*NTGLO* (X67959)			Hansen et al. (1993) *Mol. Gen. Genet.* 239, 310.
			NTDEF (X96428)		Davies et al. (1996) *Plant J.* 10, 663.
	Petunia hybrida	*FBP1* (M91190)			Angenent et al. (1992) *Plant Cell* 4, 983.
		PMADS2 (X69947)			Kush et al. (1993) *Plant Physiol.* 102, 1051.
			PMADS1 (X69946)		Kush et al. (1993) *Plant Physiol.* 102, 1051.
			PhTM6 (AF230704)		Kramer and Irish (2000) *Int. J. Plant. Sci.* 161, S29.
	Solanum tuberosum		*STDEF* (X67511)		Garcia-Maroto et al. (1993) *Plant J.* 4, 771.
Euasterids II Apiales					
Apiaceae	*Daucus carota*		*DcMADS3* (AJ271149)		Unpublished
Asterales					
Asteraceae	*Gerbera hybrida*	*GGLO1* (AJ009726)			Yu et al. (1999) *Plant J.* 17, 51.
			GDEF1 (AJ009724)		Yu et al. (1999) *Plant J.* 17, 51.
			GDEF2 (AJ009725)		Yu et al. (1999) *Plant J.* 17, 51.
	Hieracium piloselloides		*HPDEF1* (AF180364)		Guerin et al. (2000) *Planta* 210, 914.
			HPDEF2 (AF180365)		Guerin et al. (2000) *Planta* 210, 914.

* Only M-, I-, and K-domain regions were determined. This sequence was excluded from phylogenetic analyses.

Fig. 2. A part of the alignment of the C domain of AP3- and PI-homologues showing various motifs. Gray boxes in AP3 indicate a deletion found only in Magnoliales. The PI-derived A (PI-dA) motif and PI-derived B (PI-dB) motif in the previously recognized PI-derived motif and the TM6 motif (thick-lined box) are newly recognized in this study. The most frequent amino acid at each site in the TM6 motif is highlighted in boldface.

(Goto and Meyerowitz, 1994), each of which has an exon 5 of 30 bp. However, by sequencing genomic DNA, we determined the size of exon 5 in *Amborella* to be 42 instead of 30 bp.

Using the confirmed splicing sites from *Am.tr.PI* derived in this study and previously reported for *PI*, *GLO*, *AP3*, and *DEF*, we determined the putative splicing sites for all other B-class genes (Fig. 3). We recognized the length of exon 5 in each sequence because the splicing sites of exon 5 of these aligned five genes matched exactly. This comparison revealed that *Amborella* and *Nuphar* both have an exon 5 of 14 aa (42 bp), whereas all other angiosperms analyzed have an exon 5 that is 10 aa (30 bp) in length (Fig. 3). *Illicium* (star anise), representing Austrobaileyales (the sister group to all other angiosperms after *Amborella* and Nymphaeaceae; the sequence of *Illicium* PI was kindly provided by Elena Kramer), also has an exon 5 of 10 aa. Because all AP3-homologues, as well as most other MADS families, have an exon 5 of 14 aa, a deletion of four amino acids must have occurred in exon 5 just after the branches leading to *Amborella* and Nymphaeaceae, providing additional support for their basal placement in angiosperms.

Exon 5 is located in the K3 portion of the K domain, which plays an important role in the specificity and strength of dimerization between AP3 and PI proteins in both *Arabidopsis* and *Antirrhinum* (Zachgo et al., 1995; Riechmann et al., 1996). In *Arabidopsis*, deletion of K3 affects the strength of AP3/PI dimerization more than does deletion of the I or C domains (Yang et al., 2003). In *Amborella* and *Nuphar*, the K3 amphipathic-helix motif has a different structure than the $(abcdefg)_n$ repeat (a and d being hydrophobic residues) known from core eudicots. Thus, the large deletion in exon 5 that occurred in angiosperms after the branch to Nymphaeaceae may have generated different AP3/PI heterodimerization capacity after the branches to *Amborella* and Nymphaeaceae.

Certain amino acid residues are crucial for AP3/PI heterodimerization in *Arabidopsis*: 97-E and 98-N in PI and 98-N and 102-R in AP3 (Yang et al., 2003). We confirmed that these residues are conserved in most AP3- and PI-homologues of angiosperms. However, some of these residues are not present in AP3- and PI-homologues of *Amborella*: Am.tr.PI has 97-D, Am.tr.AP3-1 has 98-S, and Am.tr.AP3-2 has 102-Q (Fig. 4). The PI-homologue of *Amborella* (Am.tr.PI) has 98-N and 102-R, which are conserved residues in most AP3-homologues. One of the AP3-homologues of *Amborella* (Am.tr.AP3-2) has 97-E and 98-N, which are conserved residues in most PI-homologues, and another *Amborella* AP3-homologue (Am.tr.AP3-1) has 97-E and 102-R, which are conserved residues in PI- and AP3-homologues, respectively. An AP3/PI homologue from the gymnosperm *Gnetum gnemon*, GGM2, has 97-E, 98-N, and 102-R (Winter et al., 2002b; Yang et al., 2003) (Fig. 4).

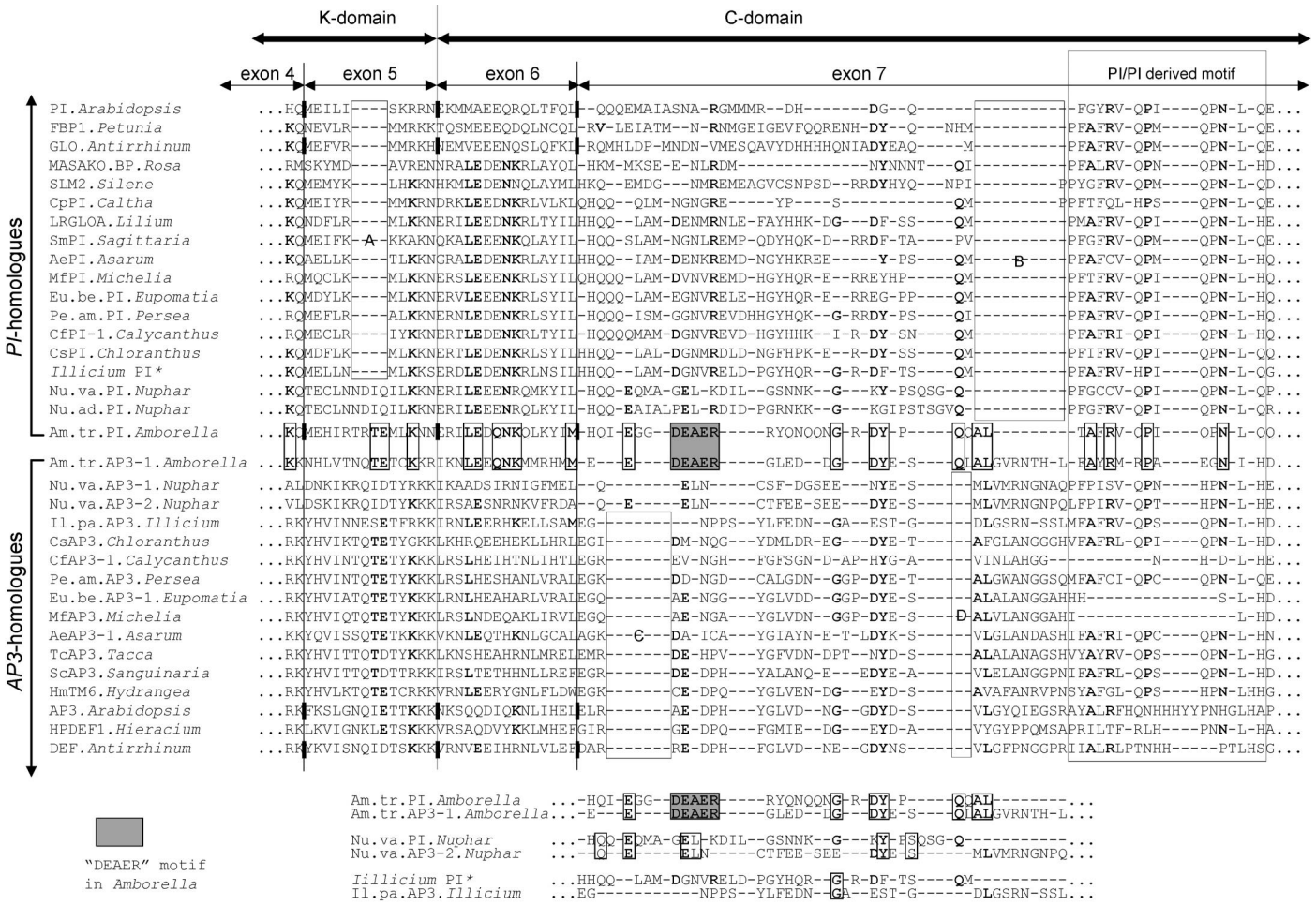

Fig. 3. Translations of the *AP3*- and *PI*-homologues of *Amborella* (Am.tr.AP3 and Am.tr.PI) show several shared amino acid strings, including a prominent ''DEAER motif'' (gray box) in exon 7, the most variable part of the gene. PI-specific (box B) and AP3-specific (boxes C and D) gaps were recognized. Shared amino acid residues between Am.tr.AP3 and Am.tr.PI (hollow boxes) are highlighted in bold. Vertical lines indicate putative splicing sites recognized from the comparison of confirmed splicing sites (solid bars) in *Am.tr.PI, PI* (Goto and Meyerowitz, 1994), *GLO* (Tröbner et al., 1992), *AP3* (Purugganan and Suddith, 1999), and *DEF* (Schwarz-Sommer et al., 1992). Box A indicates a four-amino acid deletion found in exon 5 of all PI-homologues except *Amborella* and *Nuphar*. AP3- and PI-homologues of *Amborella*, *Nuphar* (Nymphaeaceae), and *Illicium* (Austrobaileyales), respectively, are compared in detail at the bottom of the figure. *Provided by E. Kramer; unpublished data.

Differences in *Amborella* of amino acid residues required for heterodimerization in *Arabidopsis* suggest that *Amborella* B-function proteins may not have the same capacity to heterodimerize as do those of *Arabidopsis* (and perhaps other angiosperms). Furthermore, because hypothetical *Amborella* AP3 and PI proteins share residues at these sites with GGM2, which forms a homodimer (Winter et al., 2002b), both *Amborella* AP3- and PI-homologues may have the ability to homodimerize, which has only been demonstrated so far for PI-homologues of some monocots (Winter et al., 2002b). Recent studies using transgenic *Arabidopsis* plants indicate that the C terminus of AP3 is sufficient to confer AP3 functionality on the paralogous PI protein (Lamb and Irish, 2003). This finding, when extended to *Amborella* and its indistinct AP3 and PI C domains, further supports the possibility that Am.tr.AP3-1/Am.tr.AP3-1, Am.tr.AP3-2/Am.tr.AP3-2, and Am.tr.AP3-1/Am.tr.AP3-2 might be capable of homo- and heterodimerization.

The C domains apparently signal for assembly of ternary protein complexes for several MADS proteins in core eudicots (Egea-Cortines et al., 1999; Ferrario et al., 2003). Indeed, higher-order multimers are probably the active state of B-function MADS-box proteins (Egea-Cortines et al., 1999; Honma and Goto, 2000; Theißen, 2001; Ferrario et al., 2003), and the composition of possible active multimers may only be limited by (i) available MADS translation products in cells and (ii) the capacity of different AP3 and PI dimers to recognize these translation products for effective multimerization. The strong similarity between *Amborella* AP3 and PI C-domain amino acid sequences, if it extends functionally to indistinct AP3-/PI-homologue recognition *in vivo,* could suggest that the range of possible multimer combinations in *Amborella* may be substantially greater than that of eudicots or monocots.

Our data suggest that the evolution of B-function MADS-box gene controls in the development of the earliest flowers was rather dynamic, with different ''experiments'' tried. *Amborella trichopoda* may have been the most biochemically flexible angiosperm in terms of B-protein multimers that could operate during its reproductive development. However, the amino acid structural evidence we present suggests that this

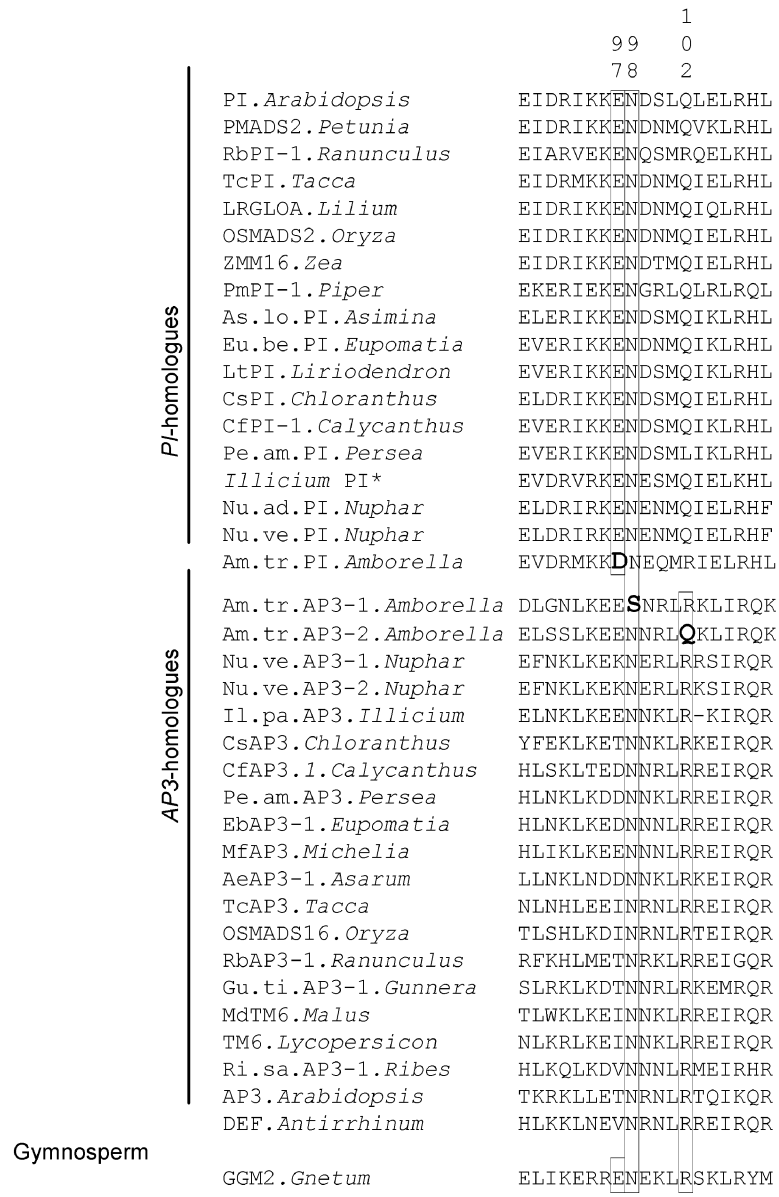

Fig. 4. Alignment of K1 subdomain. The motifs containing 97-E and 98-N in PI and 98-N and 102-R in AP3 are important for dimerization in *Arabidopsis* (Yang et al., 2003). These residues are substituted in *Amborella*: 97-D in Am.tr.PI, 98-S in Am.tr.AP3–1, and 102-Q in Am.tr.AP3–2.

inferred flexibility was rapidly lost before the bulk of the angiosperm radiation occurred. The unique phylogenetic position of *Amborella* coupled with its apparently ancestral and flexible mode of B-gene function make it a model organism that should be studied more intensively.

Structural evolution: summary—We reconstructed a hypothesis of the structural evolution of B-class genes in angiosperms using a simplified summary phylogeny (Fig. 1). The common ancestor of *AP3* and *PI* possessed an exon 6 of 14 aa (42 bp). An insertion occurred in this ancestral B-class gene to produce an exon 6 of 15 aa (45 bp) (Winter et al., 2002a). The ancestor of *AP3* and *PI* also possessed an exon 5 of 14 aa (42 bp) (Winter et al., 2002a). After the duplication and divergence of *AP3* and *PI*, a deletion of four amino acids occurred in exon 5 of *PI* during the early diversification of the angiosperms. This deletion occurred after the nodes leading to *Amborella* and Nymphaeaceae, which retain the ancestral state. All other angiosperms have the derived condition—an exon 5 of 10 aa (30 bp).

Prior to the duplication yielding the PI and AP3 lineages, several characteristic motifs must have been present in the ancestral protein. These ancestral motifs include the PI and paleoAP3 motifs (Kramer et al., 1998; Kramer and Irish, 2000). Several modifications to the PI-derived motif occurred in the angiosperms. Following the divergence of the AP3 and PI lineages, the PI motif was retained in the PI lineage and was modified in the AP3 lineage to form the PI-derived motif. All angiosperms except the core eudicots have what we term the PI-dA motif. A deletion of 10 aa in this motif unites all members of Magnoliales. The PI-derived motif was also extensively modified in the early evolution of the core eudicots

(both in the TM6 and euAP3 lineages) to form what we term the PI-dB motif.

Other structural changes also occurred in the early diversification of the AP3 lineage. For example, a portion of exon 7 was modified to yield the TM6 motif, which characterizes all members of the TM6 clade.

All motifs that are newly reported in this study were consistent in different alignments (see Materials and Methods) that we used in this study.

Phylogenetic analyses—Details of the trees obtained through phylogenetic analyses were sensitive to the alignment used. This is not surprising given the small number of characters (312–348 aligned amino acids) and large number of genes (129 sequences). However, in general, the gene trees for both *AP3* and *PI* roughly follow organismal phylogeny as inferred using multiple plastid, mitochondrial, and nuclear rRNA genes (e.g., Mathews and Donoghue, 1999; Qiu et al., 1999; P. Soltis et al., 1999; Parkinson et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; D. Soltis et al., 2000; Zanis et al., 2002; Borsch et al., 2003). Some local relationships among genes from closely related taxa were very similar regardless of the alignment, nature of the sequence (DNA or amino acid), or regions of the gene (MIKC or MIK) used, or the lineage investigated (*AP3*- or *PI*-homologues). The abbreviation for each matrix is indicated in Table 2; hereafter, we follow these abbreviations.

The basalmost sequences of the AP3 and PI lineages were nearly the same in every maximum parsimony analysis (Table 2; Fig. 5), although the internal support for relationships was low (<50% bootstrap support) based on analyses of both amino acid and DNA sequences. In the trees based on amino acids of the MIKC region, Am.tr.PI, followed by two PI-homologues of *Nuphar* (Nymphaeaceae), Nu.ad.PI and Nu.va.PI, are typically the subsequent sisters to all remaining PI-homologues. Similarly, considering the AP3-homologues, Am.tr.AP3 plus two AP3-homologues of *Nuphar* (Nu.va.AP3-1 and Nu.va.AP3-2) are the sister group to all remaining AP3-homologues in all amino acid analyses except MIK-AA-III (Table 2). For amino acid sequences of the MIKC regions, trees from the three different alignments agreed that *Amborella* and *Amborella + Nuphar* are the sisters to all other sequences in the PI and AP3 lineages, respectively. However, *Illicium* is sister to all other *AP3* sequences in all analyses based on nucleotide sequences using both MP and support weighting (Table 2; Fig. 6), although support for this relationship is weak: <50% bootstrap value in all DNA MP analyses and <50% jackknife value on the support weighted tree (Fig. 6).

Several clades of *AP3*- and *PI*-homologues that correspond to well-supported organismal clades were consistently recognized, including Magnoliales and monocots. The euAP3 clade, which was previously described (Kramer et al., 1998), was recovered in most analyses, but not all (Table 2). The Magnoliales clade in the PI lineage and the euAP3 clade in the AP3 lineage were recognized in all analyses (Table 2).

Most of the clades of sequences recovered within orders (e.g., Magnoliales, Laurales, and Poales) were consistently obtained with high bootstrap support regardless of the alignment used and with both amino acid and nucleotide sequences. For example, the clade of four AP3-homologues of Magnoliaceae, MpMADS7 (*Magnolia kobus*), Ma.gr.AP3 (*Magnolia grandiflora*), MfAP3 (*Michelia*), and LtAP3 (*Liriodendron*), received 85%, 75%, and 87% bootstrap support in analyses of the

TABLE 2. Summary of results of parsimony analyses for 10 different data matrices. Matrix names are abbreviated by their meaning. For example, MIKC-AA-I refers to M-, I-, K-, and C-domain regions, amino acid sequence, and alignment I. Dash means <50% bootstrap support. The "X" indicates the node was not recognized in the strict consensus tree.

Matrix	No. of aligned characters	No. of MP trees	CI	Basal taxa		Node support								
				AP3 lineage	PI lineage	AP3/PI	PI lineage				AP3 lineage			
							Mag.*	Mon.†	C-eu.‡	Mag.*	Mon.†	TM6	euAP3	C-eu.‡
MIKC-AA-I	312	14	0.411	*Amborella + Nuphar*	*Amborella*	100[b]	62	—	—	76	—	X(—)[d]	86	—
MIKC-DNA-I	936	2	0.164	*Illicium*	*Amborella*	93[b]	69	94	—	79	84	—(—)[d]	78	52
MIK-AA-I	170	>5000	0.355	*Amborella + Nuphar*	*Amborella*	100[b]	—	X	—	—	—	X(X)[d]	—	X
MIK-DNA-I	510	36	0.151	*Illicium*	*Amborella* or *Nuphar*	92[b]	—	83	—	—	76	—(X)[d]	—	—
MIKC-AA-II	334	17	0.414	*Amborella + Nuphar*	*Amborella*	100[b]	—	—	—	72	30	X(—)[d]	71	X
MIKC-DNA-II	1002	>5000	0.144	*Illicium*	*Peperomia + Piper*	100[b]	53	60	—	55	70	X(X)[d]	—	X
MIKC-AA-III	348	142	0.420	*Amborella + Nuphar*	*Amborella*	98/73[c]	63	—	—	86	—	X(—)[d]	93	—
MIKC-DNA-III[a]	1044	3	0.176	*Illicium*	*Asarum*	57/77[c]	—	64	X	75	67	X(—)[d]	62	—
MIK-AA-III	177	4532	0.355	*Amborella*	*Peperomia + Piper*	96/66[c]	—	—	—	X	—	X(X)[d]	—	X
MIK-DNA-III[a]	531	22	0.148	*Illicium*	*Peperomia + Piper*	—/69[c]	—	60	X	—	64	X(X)[d]	—	X

* Magnoliales.
† Monocots.
‡ Core-eudicots.
[a] Matrix used for Bayesian analysis, support weighting, and estimating divergence time.
[b] Bootstrap support for separating AP3- and PI-lineage. Each lineage was outgroup of each other.
[c] Bootstrap support for AP3- and PI-lineage. Three gymnosperm sequences were used as outgroup.
[d] Excluding *CMB2* (including *CMB2*).

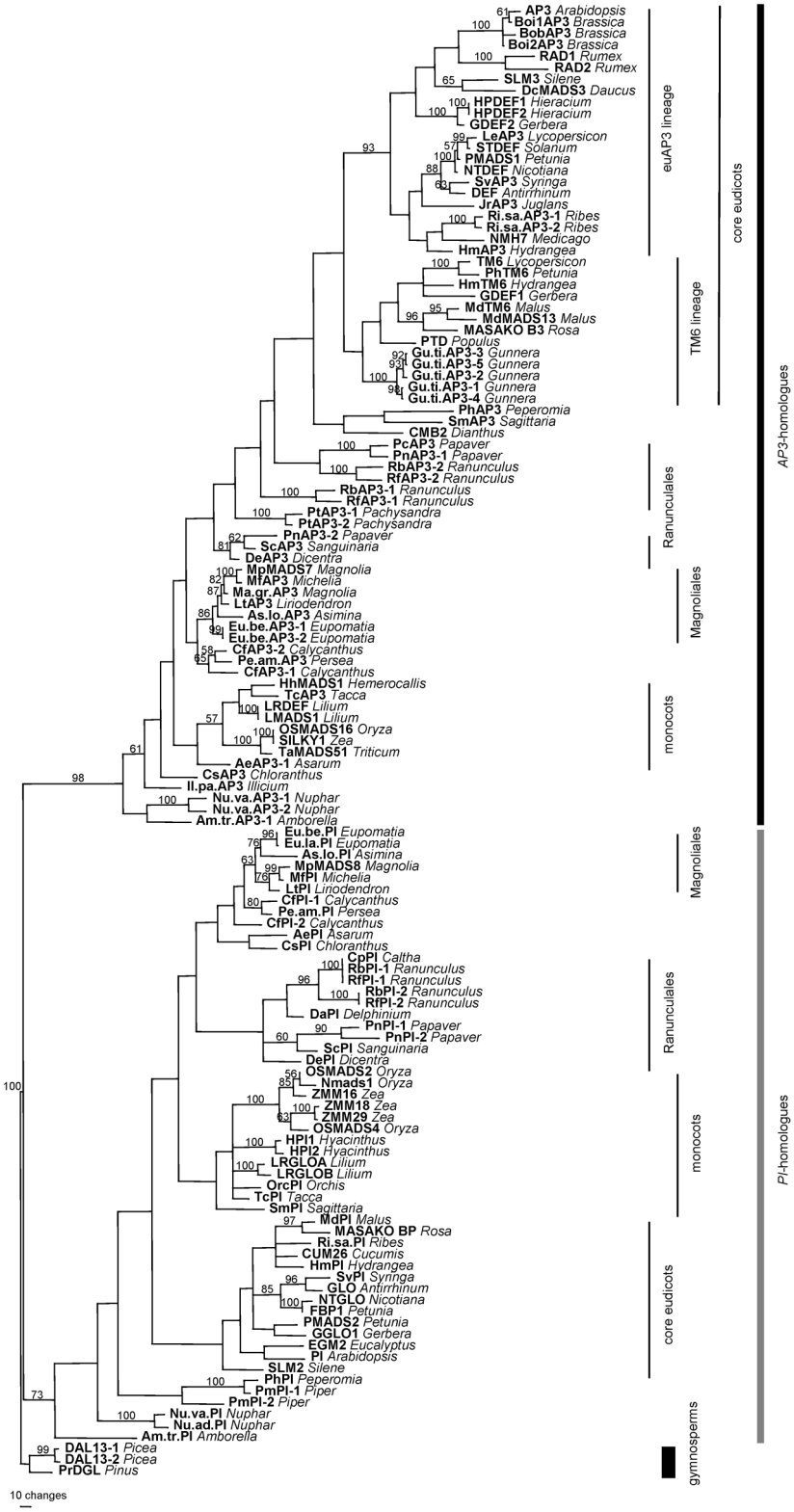

Fig. 5. Strict consensus of 72 MP trees (CI = 0.355) using M-, I-, K-, and C-domain regions of amino acid sequences (MIKC-AA-III), shown as a phylogram. Numbers above the branches are bootstrap values. Only values over 50% are indicated.

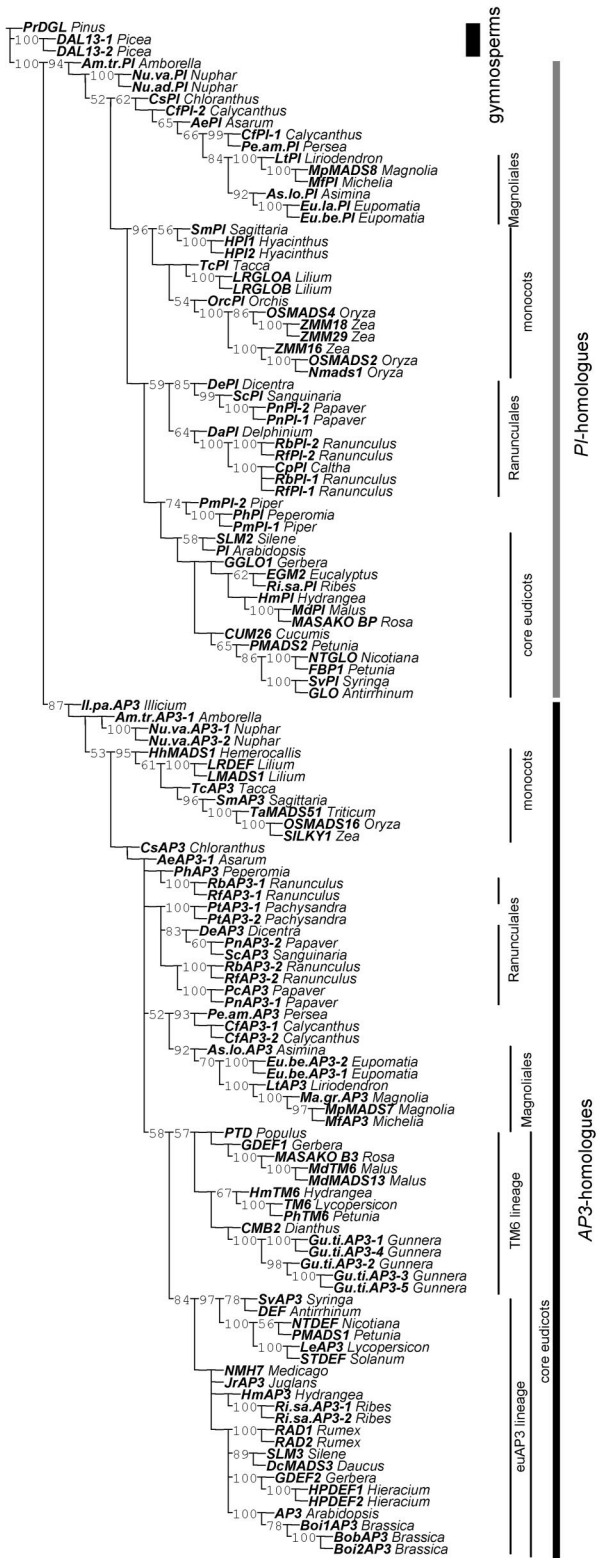

Fig. 6. Phylogenetic tree resulting from analysis of nucleotides using Support Weighting with jackknife values from non-weighted analysis.

MIKC-AA-I, MIKC-AA-II, and MIKC-AA-III alignments, respectively. The clade of three PI-homologues of Magnoliaceae, MpMADS8 (*Magnolia kobus*), MfPI (*Michelia*), and LtPI (*Liriodendron*), received 74%, 70%, and 76% support in the analyses of MIKC-AA-I, MIKC-AA-II, and MIKC-AA-III alignments, respectively.

A well-defined clade of Ranunculales was observed only in the PI lineage in our analyses of the MIK-DNA-III alignment. In other analyses of other data sets, a subset of Ranunculales sequences typically was recovered, but not all Ranunculales sequences formed a clade. Many duplication events have occurred in Ranunculaceae (Kramer et al., 2003). Hence, the evolutionary history of *AP3* and *PI* in Ranunculaceae is apparently complex. If all orthologues have not been sampled for this clade, it is reasonable to expect that the placement of some sequences in the gene tree may not reflect the organismal tree because of extensive duplication and subsequent diversification.

Genes having relatively short sequences such as *PhPI* (*Peperomia*), *nmads1* (*Oryza*), *PnPI-1* (*Papaver*), *CMB2* (*Dianthus*), and *RAD2* (*Rumex*) were especially sensitive to the alignment and the method of analysis used. For example, *PhPI* was sister to the clade of euAP3 sequences in our analyses of alignment MIKC-AA-I. However, it was sister to the TM6 clade in analyses of alignment MIKC-AA-II. In contrast, *PhPI* was embedded in the TM6 clade in analyses of the MIK-AA-I alignment and was grouped together with some AP3 genes of Ranunculales (*RfAP3-1* and *RbAP3-1*) in analyses of alignments MIKC-DNA-I and MIK-DNA-I. Lastly, *PhPI* was embedded in the euAP3 clade in analyses of MIK-DNA-III.

The TM6 lineage (Kramer et al., 1998) was not recovered in all analyses. In our analyses of data sets MIKC-AA-I, MIKC-DNA-I, MIKC-AA-II, and MIKC-DNA-III, a TM6 clade was observed, but without bootstrap support >50%. Furthermore, the TM6 clade does not include *CMB2*, an unusual aa sequence that was placed in the TM6 clade by Kramer et al. (1998). However, as noted, CMB2 also lacks the TM6 motif and therefore does not appear to belong to the TM6 lineage.

The topology of the support weighted tree (Farris, 2001) based on nucleotide sequences is very similar to that of the MP trees based on translated amino acids (Fig. 6). Magnoliales, monocots, and core eudicots were recognized in both the PI and AP3 lineages. *Amborella* is the sister to all other angiosperms in the PI lineage, but *Illicium* is sister to all other angiosperms in the AP3 lineage; jackknife support for both relationships is <50%. The TM6 and euAP3 lineages were found using support weighting, and *CMB2* was included in the TM6 lineage.

Bayesian inference using nucleotide sequences produced a tree similar to those obtained with parsimony (Fig. 7). Many clades received relatively high posterior probabilities, much higher than the corresponding bootstrap values. The tendency of Bayesian analyses to yield posterior probability values higher than bootstrap values has been discussed (e.g., Huelsenbeck and Ronquist, 2001; Suzuki et al., 2002). Clades for which we obtained posterior probability values (×100) of 95–100 in the MIKC analysis include (Fig. 8): Magnoliales (100), monocots (100), the euAP3 lineage (100), the TM6 lineage (100, including *CMB2*), and core eudicots (100) for the AP3 lineage; Magnoliales (100), Ranunculales (100), monocots (100), and the core eudicots (99) for the PI lineage.

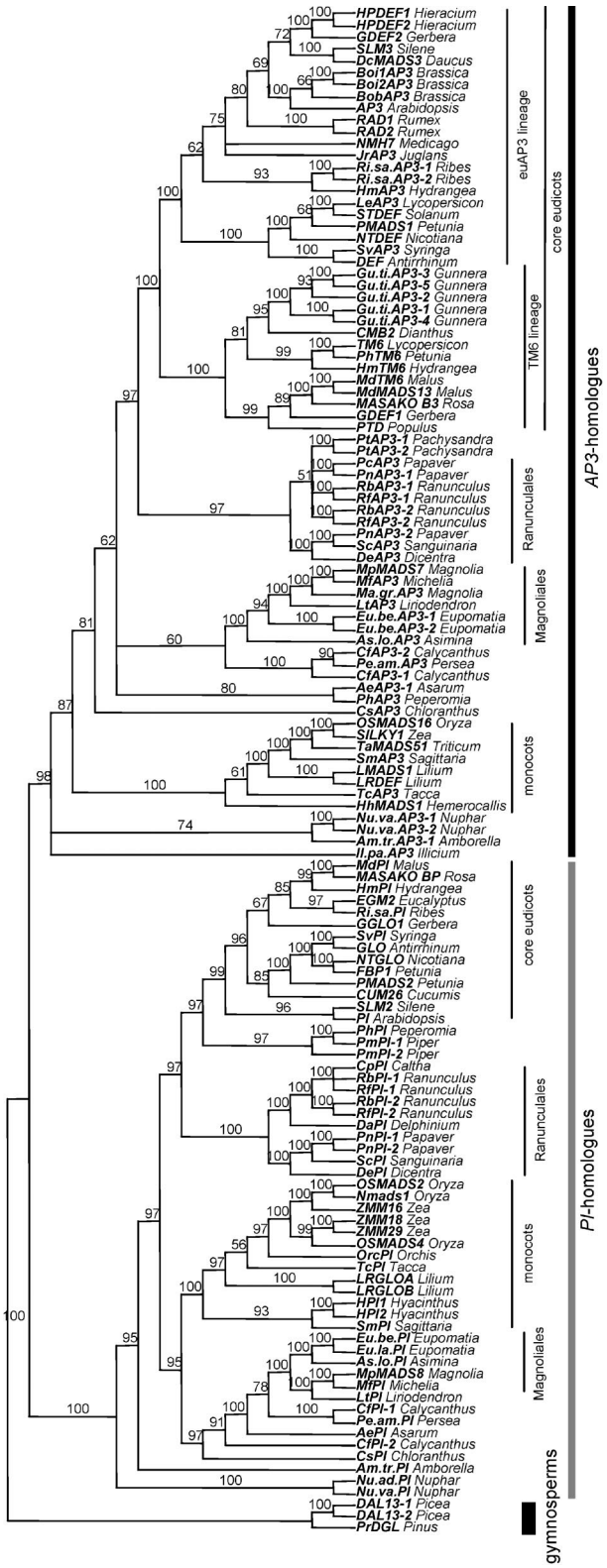

Fig. 7. Phylogenetic tree from Bayesian analysis. Numbers above branches are posterior probabilities multiplied by 100.

Estimation of divergence time—The evolution of MADS-box genes has involved a series of gene duplications and subsequent diversification (e.g., Purugganan, 1997; Kramer et al., 1998, 2003; Theißen et al., 2000; Nam et al., 2003). A duplication yielding the *AP1*-like and *SEPALLATA*-like plus *AGL6*-like genes occurred approximately 374 million years ago (Nam et al., 2003), and the ages of several other prominent MADS-box gene duplications have also been estimated (e.g., Purugganan et al., 1995; Purugganan, 1997; Nam et al., 2003). However, the age of the AP3/PI duplication could not be estimated in these earlier studies because of the accelerated rates of *AP3* and *PI* sequence evolution relative to other MADS-box genes (Purugganan et al., 1995; Purugganan, 1997; Nam et al., 2003). Here we focus only on B-class genes and use nonparametric rate smoothing (NPRS), which accommodates rate heterogeneity by permitting rates to vary among branches, thereby allowing us to estimate the timing of the *AP3*/*PI* duplication. Although the effectiveness of NPRS in accommodating rate heterogeneity has not been adequately tested and NPRS may overcompensate for rate inconstancy (Sanderson and Doyle, 2001), estimates in other studies based on NPRS, penalized likelihood (Sanderson, 2002), and Bayesian methods (Thorne and Kishino, 2002) are similar (e.g., S. Renner, University of Missouri, St. Louis, personal communication; C. Bell et al., University of Florida, unpublished data). We conclude that NPRS provides reasonable estimates for data sets with heterogeneous rates.

Using our *AP3*- and *PI*-homologue data sets and NPRS (Sanderson, 1997), we estimated that the duplication that produced the AP3 and PI lineages occurred approximately 260 mya (range of 230–290 mya) (Table 3). This date places the duplication shortly after the split between extant gymnosperms and angiosperms and on the ''stem'' (ancestral) lineage of extant flowering plants. Extant seed plants originated approximately 290–309.2 mya (Mapes and Rothwell, 1984, 1991), and most evidence indicates a very early split between the living gymnosperms and the line leading to angiosperms (P. Soltis et al., 2002). The absence of *AP3*- or *PI*-specific homologues in extant gymnosperms was previously used to suggest that the *AP3*/*PI* duplication arose prior to the origin of the angiosperms (Doyle, 1994; Purugganan et al., 1995; Kramer et al., 1998; Becker and Theißen, 2003). However, few gymnosperms and none of the basalmost angiosperms had been examined, and it was therefore unclear whether the duplication occurred early in angiosperm history, just prior to the origin of angiosperms, early along the stem lineage leading to angiosperms, or even early in seed plant evolution. Although two classes of *AP3*/*PI* homologues have been suggested for gymnosperms based on comparison of motifs (Kramer and Irish, 2000), phylogenetic analyses to date have not supported this distinction (Hasebe, 1999; Kramer and Irish, 2000).

In summary, our results suggest that the *AP3*/*PI* duplication occurred shortly after the divergence of extant gymnosperms and angiosperms, well before the oldest flowering plant fossils (generally placed in the range of 125–131.8 mya; Hughes, 1994; P. Soltis et al., 2002). Importantly, this implies that the joint expression of *AP3* and *PI* may not have immediately resulted in the formation of petals, structures for which they control the development in extant angiosperms. Recognizable flowers did not appear for perhaps another 100 million years after the estimated timing of the *AP3*/*PI* gene duplication. The co-expression of *AP3*-and *PI*-homologues could nevertheless reflect an evolutionary innovation of animal-attractive, petal-

TABLE 3. Divergence times between AP3 and PI lineages estimated directly from original matrices and from bootstrapped matrices using NPRS. The ''TM6 data set'' includes the TM6 lineage (plus paleoAP3 and PI genes) but excludes the euAP3 lineage; the ''euAP3 data set'' indicates the converse.

Data set	Region	Calibration point: origin of eudicots in	Divergence time (mya)	
			Direct	Bootstrap ± 1 SD
TM6	MIKC	PI lineage	244	244.75 ± 10.13
		AP3 lineage	290	289.96 ± 7.85
	MIK	PI lineage	231	226.59 ± 13.07
		AP3 lineage	282	277.96 ± 13.81
euAP3	MIKC	PI lineage	241	239.78 ± 7.42
		AP3 lineage	264	262.59 ± 12.04
	MIK	PI lineage	230	227.58 ± 9.55
		AP3 lineage	247	245.68 ± 7.35

like organs prior to the recognition of flowering plants in the fossil record. Transference of B-class gene function from control over gymnosperm sex determination to petal-like and true petal organ identity (Albert et al., 1998) is an attractive hypothesis meriting further research.

LITERATURE CITED

APG II. 2003. An updated classification of the angiosperms. *Botanical Journal of the Linnean Society* 141: 399–436.

ALBERT, V. A., M. H. G. GUSTAFSSON, AND L. DI LAURENZIO. 1998. Ontogenetic systematics, molecular developmental genetics, and the angiosperm petal. *In* P. Soltis, D. Soltis, and J. Doyle [eds.], Molecular systematics of plants II, 349–374. Kluwer, Boston, Massachusetts, USA.

ALVAREZ-BUYLLA, E. R., S. PELAZ, S. J. LILJEGREN, S. E. GOLD, C. BURGEFF, G. S. DITTA, L. R. DE POUPLANA, L. MARTINEZ-CASTILLA, AND M. F. YANOFSKY. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plant and animals. *Proceedings of the National Academy of Sciences, USA* 97: 5328–5333.

BARKMAN, T. J., G. GHENERY, J. R. MCNEAL, J. LYONS-WEILER, AND C. W. DEPAMPHILIS. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proceedings of the National Academy of Sciences, USA* 97: 13166–13171.

BECKER, A., K. KAUFMANN, A. FREIALDENHOVEN, C. VINCENT, M.-A. LI, H. SAEDLER, AND G. THEIßEN. 2002. A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Molecular Genetics and Genomics* 266: 942–950.

BECKER, A., AND G. THEIßEN. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution* 29: 464–489.

BODT, S. D., J. RAES, Y. V. DE PEER, AND G. THEIßEN. 2003. And then there were many: MADS goes genomic. *Trends in Plant Science* 8: 475–483.

BORSCH, T., K. W. HILU, D. QUANDT, V. WILDE, C. NEINHUIS, AND W. BARTHLOTT. 2003. Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558–576.

COEN, E. S., J. S. DOYLE, M. ROMERO, R. ELLIOT, R. MAGRATH, AND R. CARPENTER. 1991. Homeotic genes controlling flower development in *Antirrhinum*. *Development (Supplement)* 1: 149–156.

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

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 Systematics and Evolution supplement* 8: 93–122.

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

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation from small amount of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

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 Journal* 18: 5370–5379.

FARRIS, J. S. 2001. Support weighting. *Cladistics* 17: 389–394.

FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, AND A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.

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

FERRARIO, S., R. G. IMMINK, A. SHCHENNIKOVA, J. BUSSCHER-LANGE, AND G. C. ANGENENT. 2003. The MADS box gene *FBP2* is required for the *SEPALLATA* function in petunia. *Plant Cell* 15: 914–925.

GRAHAM, S. W., AND R. G. OLMSTEAD. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

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

HASEBE, M. 1999. Evolution of reproductive organs in land plants. *Journal of Plant Research* 112: 463–474.

HILU, K. W., T. BORSCH, K. MÜLLER, D. E. SOLTIS, P. S. SOLTIS, V. SAVOLAINEN, M. W. CHASE, M. POWELL, L. A. ALICE, R. EVANS, H. SAUQUET, C. NEINHUIS, T. A. SOTTA, J. G. ROHWER, C. S. CHAMPBELL, AND L. CHATROU. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.

HONMA, T., AND K. GOTO. 2000. 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.

HUGHES, N. F. 1994. The enigma of angiosperm origins. Cambridge University Press, Cambridge, UK.

JOHANSEN, B., L. B. PEDERSEN, M. SKIPPER, AND S. FREDERIKSEN. 2002. MADS-box gene evolution-structure and transcription patterns. *Molecular Phylogenetics and Evolution* 23: 458–480.

KIM, S., C.-W. PARK, Y.-D. KIM, AND Y. SUH. 2001. Phylogenetic relationships in family Magnoliaceae inferred from *ndhF* sequences. *American Journal of Botany* 88: 717–728.

KRAMER, E. M., V. S. DI STILIO, AND P. M. SCHLUTER. 2003. Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. *International Journal of Plant Sciences* 164: 1–11.

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.

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. *International Journal of Plant Sciences* 161: S29–S40.

LAMB, R. S., AND V. F. IRISH. 2003. Functional divergence within the *APETALA3/PISTILLATA* floral homeotic gene lineages. *Proceedings of the National Academy of Sciences, USA* 100: 6558–6563.

MA, H., AND C. DEPAMPHILLIS. 2000. The ABCs of floral evolution. *Cell* 101: 5–8.

MAPES, G., AND G. W. ROTHWELL. 1984. Permineralized ovulate cones of *Lebachia* from the late Palaeozoic limestones of Kansas. *Paleontology* 27: 69–94.

MAPES, G., AND G. W. ROTHWELL. 1991. Structure and relationships of primitive conifers. *Neues Jahrbuch fur Geologie and Paläontologie-Abhandlungen* 183: 269–287.

MATHEWS, S., AND M. J. DONOGHUE. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.

MEYEROWITZ, E. M., J. L. BOWMAN, L. L. BROCKMAN, G. N. DREWS, T. JACK, L. E. SIEBURTH, AND D. WEIGEL. 1991. A genetic and molecular model for floral development in *Arabidopsis thaliana*. *Development (Supplement)* 1: 157–167.

MOURADOV, A., T. V. GLASSIK, B. A. HAMDORF, L. C. MURPHY, S. S. MARLA, Y. YANG, AND R. TEASDALE. 1998. Family of MADS-box genes expressed early in male and female reproductive structures of monterey pine. *Plant Physiology* 117: 55–61.

MOURADOV, A., B. HAMDORF, R. D. TEASDALE, J. KIM, K.-U. WINTER, AND G. THEIßEN. 1999. A DEF/GLO-like MADS-box gene from a gymnosperm: *Pinus radiata* contains an ortholog of angiosperm B class floral homeotic genes. *Developmental Genetics* 25: 245–252.

MÜNSTER, T., J. PAHNKE, A. DI ROSA, J. T. KIM, W. MARTIN, H. L. SAEDLER, AND G. THEIßEN. 1997. Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proceedings of the National Academy of Sciences, USA* 94: 2415–2420.

NAM, J., C. W. DEPAMPHILIS, H. MA, AND M. NEI. 2003. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Molecular Biology and Evolution* 20: 1435–1447.

PAŘENICOVÁ, L., S. DE FOLTER, M. KIEFFER, D. S. HORNER, C. FAVALLI, J. BUSSCHER, H. E. COOK, R. M. INGRAM, M. M. KATER, B. DAVIES, G. C. ANGENENT, AND L. COLOMBO. 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*. *Plant Cell* 15: 1538–1551.

PARKINSON, C. L., K. L. ADAMS, AND J. D. PALMER. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* 9: 1485–1488.

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

PURUGGANAN, M. D. 1997. The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *Journal of Molecular Evolution* 45: 392–396.

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

PURUGGANAN, M. D., AND J. I. SUDDITH. 1999. Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILATA* genes of *Arabidopsis thaliana*. *Genetics* 151: 839–848.

QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN, AND M. W. CHASE. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404–407.

RIECHMANN, J. L., B. A. KRIZEK, AND E. M. MEYEROWITZ. 1996. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences, USA* 93: 4793–4798.

RIECHMANN, J. L., AND E. M. MEYEROWITZ. 1997. MADS domain proteins in plant development. *Biological Chemistry* 378: 1079–1101.

SANDERSON, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* 14: 1218–1231.

SANDERSON, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–109.

SANDERSON, M. J., AND J. A. DOYLE. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. *American Journal of Botany* 88: 1499–1516.

SAWYER, S. A. 1989. Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* 6: 526–538.

SCHWARZ-SOMMER, Z., I. HUE, P. HUIJSER, P. J. FLOR, R. HANSEN, F. TETENS, W. E. LONNING, H. SAEDLER, AND H. SOMMER. 1992. Characterization of *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO Journal* 11: 251–263.

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.

SHORE, P., AND A. D. SHARROCKS. 1995. The MADS-box family of transcription factors. *European Journal of Biochemistry* 229: 1–13.

SOLTIS, D. E., A. E. SENTERS, M. J. ZANIS, S. KIM, J. D. THOMPSON, P. S. SOLTIS, L. P. R. DE CARENE, P. K. ENDRESS, AND J. S. FARRIS. 2003. Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. *American Journal of Botany* 90: 461–470.

SOLTIS, D. E., P. S. SOLTIS, M. W. CHASE, M. E. MORT, D. C. ALBACH, M. ZANIS, V. SAVOLAINEN, W. H. HAHN, S. B. HOOT, M. F. FAY, M. AXTELL, S. M. SWENSEN, L. M. PRINCE, W. J. KRESS, K. C. NIXON, AND J. S. FARRIS. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

SOLTIS, P. S., D. E. SOLTIS, AND M. W. CHASE. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.

SOLTIS, P. S., D. E. SOLTIS, V. SAVOLAINEN, P. R. CRANE, AND T. G. BARRACLOUGH. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Sciences, USA* 99: 4430–4435.

SUZUKI, Y., G. V. GLAZKO, AND M. NEI. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences, USA* 99: 16138–16143.

SWOFFORD, D. L. 2001. PAUP* 4.0b10: phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, Massachusetts, USA.

THEIßEN, G. 2001. Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology* 4: 75–85.

THEIßEN, G., A. BECKER, A. DI ROSA, A. KANNO, J. T. KIM, T. MUNSTER, K.-U. WINTER, AND H. SAEDLER. 2000. A short history of MADS-box genes in plants. *Plant Molecular Biology* 42: 115–149.

THEIßEN, G., J. KIM, AND H. SEADLER. 1996. Classification and phylogeny of the MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution* 43: 484–516.

THEIßEN, G., AND H. SAEDLER. 1995. MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. *Current Opinion in Genetics and Development* 3: 628–639.

THEIßEN, G., AND H. SAEDLER. 2001. Floral quartets. *Nature* 409: 469–471.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

THORNE, J. L., AND H. KISHINO. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* 51: 689–702.

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

WINTER, K.-U., C. WEISER, K. KAUFMANN, A. BOHNE, C. KIRCHNER, A. KANNO, H. SAEDLER, AND G. THEIßEN. 2002a. On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm *Gnetum*. *The Plant Journal* 31: 457–475.

WINTER, K.-U., C. WEISER, K. KAUFMANN, A. BOHNE, C. KIRCHNER, A. KANNO, H. SAEDLER, AND G. THEIßEN. 2002b. Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. *Molecular Biology and Evolution* 19: 587–596.

YANG, Y., L. FANNING, AND T. JACK. 2003. The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APTEALA3 and PISTILLATA. *The Plant Journal* 33: 47–59.

ZACHGO, S., ED. SILVA, P. MOTTE, W. TRÖBNER, H. SAEDLER, AND Z. SCHWARZ-SOMMER. 1995. Functional analysis of the *Antirrhinum* floral homeotic *DEFICIENS* gene in vivo and in vitro by using a temperature-sensitive mutant. *Development* 121: 2861–2875.

ZANIS, M. J., D. E. SOLTIS, P. E. SOLTIS, S. MATHEWS, AND M. J. DONOGHUE. 2002. The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences, USA* 99: 6848–6853.