Molecular Evolution of Genes Controlling Petal and Stamen Development: Duplication and Divergence Within the *APETALA3* and *PISTILLATA* MADS-Box Gene Lineages

Elena M. Kramer,* Robert L. Dorit† and Vivian F. Irish*

**Departments of Molecular, Cellular and Developmental Biology and †Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520*

Manuscript received December 24, 1997
Accepted for publication March 2, 1998

ABSTRACT

The specification of floral organ identity in the higher dicots depends on the function of a limited set of homeotic genes, many of them members of the MADS-box gene family. Two such genes, *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), are required for petal and stamen identity in Arabidopsis; their orthologs in Antirrhinum exhibit similar functions. To understand how changes in these genes may have influenced the morphological evolution of petals and stamens, we have cloned twenty-six homologs of the *AP3* and *PI* genes from two higher eudicot and eleven lower eudicot and magnolid dicot species. The sequences of these genes reveal the presence of characteristic *PI*- and *AP3*-specific motifs. While the *PI*-specific motif is found in all of the *PI* genes characterized to date, the lower eudicot and magnolid dicot *AP3* homologs contain distinctly different motifs from those seen in the higher eudicots. An analysis of all the available *AP3* and *PI* sequences uncovers multiple duplication events within each of the two gene lineages. A major duplication event in the *AP3* lineage coincides with the base of the higher eudicot radiation and may reflect the evolution of a petal-specific *AP3* function in the higher eudicot lineage.

FLOWERS are a defining characteristic of the angiosperms. The typical hermaphroditic angiosperm flower contains both sterile and reproductive organs. These organs are generally organized into whorls, with a particular organ type arising from a single node on the axis of a determinate floral meristem. The flowers of the model species *Arabidopsis thaliana* display the typical higher eudicot floral organization. The first and second whorls of the flower contain the sterile sepals and petals, respectively. The third whorl contains the stamens, the male reproductive structures which produce pollen. The female reproductive structures, the carpels, arise in the fourth whorl and contain the ovules. Numerous variations on this basic floral architecture exist within the angiosperms. These differences in organ number, structure and phyllotaxy are critical morphological characters in the study of angiosperm systematics.

The evolution of angiosperm floral diversity has been a subject of considerable study. While the stamens and carpels are thought to have each evolved only once, it is widely accepted that the sterile organs have evolved many times within the angiosperms, although the details of these events are unresolved (Cronquist 1988; Takhtajan 1991; Drinnan *et al.* 1994; Endress 1994). Much of the controversy has centered on the number and the nature of petal derivation events within the various angiosperm lineages (Figure 1). Phylogenetic analyses of the angiosperms based on a large rbcL data set have identified two major monophyletic clades (Chase *et al.* 1993; Crane *et al.* 1995; Qiu *et al.* 1993). Both of these groups, the eudicots and the monocots, are rooted within an unresolved basal grade of magnolid dicots. The eudicot clade can be further subdivided into the lower eudicots, comprising the Ranunculidae, basal Hamamelididae and basal Rosidae, and the higher eudicots, made up of the bulk of the flowering plants, including the majority of the model species used for genetic analysis (Drinnan *et al.* 1994). Stamenally-derived petals, called andropetals, have evolved many times within the lower eudicots and at least once at the base of the higher eudicot clade and the monocot clade (Takhtajan 1991). A second type of petals, bracteopetals, are derived from sepals or other sterile subtending organs. The diverse magnolid dicots include species which have been characterized as possessing bracteopetals, as well as a number of species which are considered to have andropetals (Takhtajan 1991). The designations of petals as being andropetallous or bracteopetallous have been based primarily on morphological characters.

The isolation of floral homeotic mutants in Arabidopsis and Antirrhinum has provided an inroad into the dissection of the genetic mechanisms underlying floral diversity and the derivation of floral organs (Komaki *et al.* 1988; Bowman *et al.* 1989; Carpenter and Coen 1990). Genetic analysis of these mutants has led to a model where three classes of genes, known as the

Corresponding author: Vivian Irish, Department of Molecular, Cellular and Developmental Biology, Yale University, P.O. Box 208104, New Haven, CT 06520-8104. E-mail: vivian.irish@yale.edu

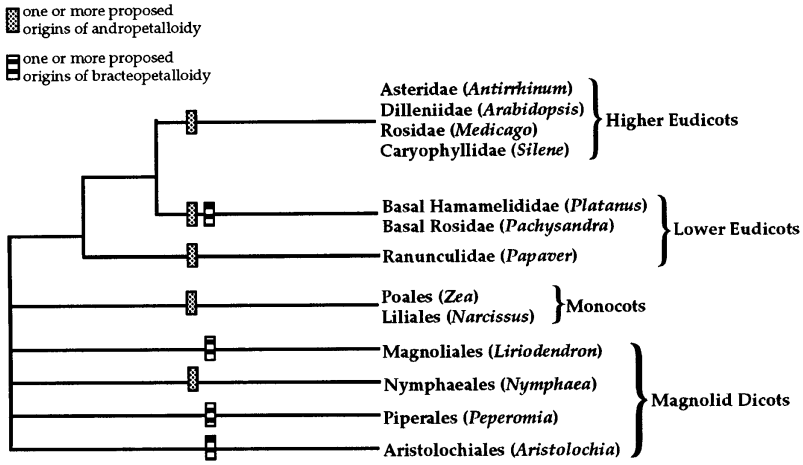

Figure 1.—Simplified phylogeny showing the relationships between major angiosperm clades (Chase *et al.* 1993; Crane *et al.* 1995; Friis and Endress 1990; Qiu *et al.* 1993). Lineages which have been proposed to represent independent derivations of andropetals and bracteopetals are indicated (Dahlgren *et al.* 1984; Drinnan *et al.* 1994; Takhtajan 1991).

A, B, and C classes, interact in a combinatorial manner to specify particular organ identities (Bowman *et al.* 1991; Coen and Meyerowitz 1991; Meyerowitz *et al.* 1991). Mutations in the B group genes display transformations of the petals into sepals and the stamens into carpels, indicating that the corresponding wild-type gene products are required for specifying petal and stamen identity (Jack *et al.* 1992; Goto and Meyerowitz 1994). Like many of the organ-identity genes, the Arabidopsis B group genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) encode products which contain a MADS domain, a highly conserved region of approximately 57 amino acids which has been shown to play a role in DNA binding and protein dimerization (Norman *et al.* 1988; Pollock and Treisman 1991). The plant representatives of the MADS-box family are further distinguished by the presence of a 70-amino acid region called the K domain (Figure 2; Ma *et al.* 1991; Davies and Schwarz-Sommer 1994). This portion of the protein is predicted to form two to three amphipathic helices which may facilitate protein-protein interactions (Pnueli *et al.* 1991). Separating the MADS and K domains is a short intervening region (I) of approximately thirty amino acids which, along with the K domain, has been shown to play a role in dimerization specificity (Krizek and Meyerowitz 1996; Reichmann *et al.* 1996a). The C-terminal portions of the proteins vary considerably in size and sequence within the family and have yet to be assigned any specific function.

Figure 2.—Domains of the plant MADS-box gene as defined by sequence conservation and protein structure. The MADS-box displays the highest level of conservation within the family and has been shown to play a role in DNA binding and protein dimerization. The I and K regions also appear to play roles in dimerization but show lower levels of sequence conservation. The C-terminal region is the most divergent in sequence but does contain some highly conserved lineage-specific motifs.

While many of the currently characterized MADS domain proteins appear to function as homodimers, the *AP3* and *PI* gene products are thought to act together as a heterodimeric transcription factor. This conclusion is supported by the finding that both the *AP3* and *PI* gene products are required for DNA binding and nuclear localization (McGonigle *et al.* 1996; Riechmann *et al.* 1996a; Hill *et al.* 1998). Furthermore, the AP3 and PI proteins have been shown to bind to each other in immunoprecipitation experiments (Goto and Meyerowitz 1994; Riechmann *et al.* 1996a). The maintenance of *AP3* and *PI* expression depends upon the presence of both gene products, suggesting that the AP3/PI heterodimer promotes the transcription of *AP3* and *PI*, perhaps directly (Jack *et al.* 1992; Goto and Meyerowitz 1994). Interestingly, phylogenetic analysis of the entire plant MADS box gene family has shown that the *AP3* and *PI* lineages are the products of a duplication event which makes them more closely related to each other than to any of the other MADS-box genes (Doyle 1994; Purugganan *et al.* 1995; Purugganan 1997; Theissen *et al.* 1996).

The conservation of B group functions across the angiosperms is being addressed through studies of *AP3* and *PI* homologs in several higher eudicot species. In general, the expression patterns of these genes are all quite similar to those seen in Arabidopsis (Theissen and Saedler 1995; Irish and Kramer 1998). The mutant phenotypes, in those species where they have been analyzed, are generally consistent with a conserved role for *AP3* and *PI* in promoting the establishment of petal and stamen identity (Sommer *et al.* 1990; Trobner *et al.* 1992; Angenent *et al.* 1993; van der Krol and Chua 1993). Furthermore, the *AP3* ortholog from Antirrhinum, *DEFICIENS* (*DEF*), is able to largely replace endogenous Arabidopsis *AP3* function (Irish and Yamamoto 1995; Samach *et al.* 1997). Similarly, both *DEF* and the Antirrhinum *PI* ortholog, *GLOBOSA* (*GLO*), have been

shown to promote stamen and petal identity when ectopically expressed in Nicotiana (Davies *et al.* 1996). Taken together, these results support the conclusion that *AP3* and *PI* orthologs are responsible for determining petal and stamen identity within the higher eudicots.

Since the petals of the higher eudicots are thought to be homologous, it is not surprising that all the higher eudicots studied to date appear to have a conserved petal developmental pathway. In order to examine how independent petal derivation events may be reflected in the petal developmental program, it is necessary to examine species whose petals are not homologous to those of the higher eudicots. Accordingly, we have cloned *AP3* and *PI* homologs from eleven lower eudicot and magnolid dicot species in an effort to understand the evolution both of these gene lineages and of the pathways of petal specification. We present a phylogenetic analysis of the B group genes which indicates that the path of B group gene evolution is more complex than previously thought. Our analysis suggests that there are, in fact, two paralogous *AP3* lineages in the higher eudicots: one represented by the well-studied *AP3* ortholog group and the other containing the tomato *AP3* paralog *TM6* and several related genes. The data suggest that these two lineages are the result of a gene duplication event which occurred after the divergence of the Buxaceae in the lower eudicots but before the diversification of the higher eudicots. Sequence analysis reveals that the *AP3*-like genes of the lower eudicots and magnolid dicots are actually more similar to the members of the *TM6* lineage than they are to the higher eudicot *AP3* lineage. Although we have also identified duplication events in the *PI* lineage, none appear to date back to the base of the higher eudicots. In addition, the *PI* homologs we have isolated display a greater overall conservation of sequence than do the *AP3* lineage members. Based on these observations, we present a model for the evolution of the B group gene lineage and discuss how duplication and divergence in this gene lineage may have influenced the evolution of petals.

MATERIALS AND METHODS

Species sampled and sources of plant material: The species included in this analysis are given in Tables 1–3, along with family membership, general collection information and GenBank accession numbers for the gene sequences. Throughout the text we have followed the taxonomic designations of Cronquist (1981) for dicots (the only exception to this being the designation of the Ranunculidae as an independent subclass) and Dahlgren *et al.* (1984) for monocots. The choice of taxa was influenced by both phylogenetic position and specimen availability.

Cloning and analysis: For each species, total RNA was prepared using Trizol (GIBCO BRL, Gaithersburg, MD) from whole flower buds collected across a range of developmental stages. Poly-A mRNA was extracted from total RNA using Magnetight Oligo (dT) particles (Novagen, Madison, WI). Single-stranded cDNA was synthesized by priming with the oligonucleotide 5′-CCGGATCCTCTAGAGCGGCCGC$(T)_{17}$ from 500 ng of poly-A RNA. This poly-T primer was used with a second primer with the sequence 5′-GGGGTACCAA(C/T)(A/C)GI CA(A/G)GTIACITA(T/C)TCIAAG(A/C)GI(A/C)G-3′ in a polymerase chain reaction (PCR) to amplify MADS-box-containing cDNAs (primary PCR reaction). PCR analysis was performed in 100 μl of PCR buffer (10 mm Tris pH 8.3, 50 mm KCl, 1.5 mm $MgCl_2$) containing 50 pmol and 20 pmol of 5′ and 3′ primer, respectively, 200 μmol of each deoxyribonucleotide triphosphate, and 2.5 units of AmpliTaq Gold polymerase (Perkin Elmer, Foster City, CA). Amplification began with a Taq-activation step of 12 min at 95°C, followed by 10 cycles of 20 sec denaturing at 95°, 30 sec annealing at 38° and 1 min extension at 72°. The program was completed by 30 cycles of 20 sec denaturing at 95°, 30 sec annealing at 42° and an extension time of 1 min at 72°. Amplified products were analyzed on a 1% agarose gel, revealing one or more distinct fragments of ≥0.6 kb. The reactions were directly cloned using the TA and TOPO-TA Cloning kits (Invitrogen, Carlsbad, CA). Clones were analyzed based on size and all fragments over 0.6 kb were sequenced using fluorescent sequencing methods by the Keck Foundation Biotechnology Resource Laboratory at Yale University.

In *Papaver nudicaule*, *Dicentra eximia*, *Ranunculus bulbosus* and *Pachysandra terminalis*, 3′ primers targeted to *AP3*- or *PI*-specific C-terminal motifs were used in conjunction with the degenerate 5′ primer to specifically amplify *AP3* and *PI* homologs from the primary PCR reaction (same conditions as above). The *PI*-specific primer has the sequence 5′-TGIA(A/G)(A/G)TTIGGITGIA(A/T)(T/G)GGITG and the *AP3*-specific primer has the sequence 5′-CIAGICGIAG(A/G)TC(A/G)T. For these species, clones from both the primary PCR reaction as well as the *AP3*- and *PI*-specific secondary amplifications were analyzed. The complete 3′ sequence of the clones generated with the *AP3*- or *PI*-specific primers was obtained using 3′ RACE (3′ RACE primer sequences available upon request). The complete cDNA sequences of *PnAP3-2* and *PnPI-1* were obtained using the 5′ RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (Gibco BRL, Gaithersburg, MD), in conjunction with several different oligonucleotides (sequences available upon request).

For the cloning of *LeAP3* from *Lycopersicon esculentum*, a Solanaceae *AP3*- specific primer with the sequence 5′-A(A/G)IGC(A/G)AAIGTIGTIAT(A/G)TC was designed from the C-terminal consensus D(I/L)TTFAL. This 3′ primer was used in conjunction with the degenerate 5′ MADS-box primer to specifically amplify *LeAP3*. The complete 3′ sequence was obtained using 3′ RACE (primer sequence available upon request).

Phylogenetic analysis: The sequences of all of the published B group representatives were obtained from GenBank (see Tables 2 and 3 for accession numbers). Protein sequences were aligned using CLUSTALW and refined by hand, taking both nucleotide and amino acid sequences into consideration (see appendices 1 and 2). The alignments were further modified (for phylogenetic analysis) by encoding gaps as single characters in a supplemental data matrix. In this approach, gaps are treated as single events, thus preventing their overweighting (based on gap length) in the subsequent phylogenetic analyses.

We generated parsimony trees based on the aligned protein sequences using the PAUP 4.0* package (Phylogenetic Analysis Using Parsimony, Version 61, used by permission of the author, Swofford 1993). Parsimony trees were found using the heuristic search algorithm, generating 1000 replicate runs using random stepwise addition of sequences. Multiple equal-length parsimony trees were collapsed into 50% majority rule consensus trees. Bootstrap values for all resolved nodes in

TABLE 1

Collection information

Sample	Voucher or source	Clones/species[a]
Lycopersicon esculentum cv. Celebrity	Cultivated, Kramer 108	5 (5)
Syringa vulgaris	Kramer 106	11 (11)
Pachysandra terminalis	Kramer 105	22 (9)
Papaver californicum	Kramer 100	4 (4)
Papaver nudicaule	Cultivated, Kramer 101	56 (36)
Dicentra eximia	Kramer 102	30 (20)
Delphinium ajacis	Cultivated, Kramer 104	5 (5)
Caltha palustris	Kramer 103	16 (16)
Ranunculus bulbosus	Kramer 107	31 (22)
Michelia figo	Yale Marsh Gardens	10 (10)
Liriodendron tulipifera	Yale Marsh Gardens	6 (6)
Peperomia hirta	Yale Marsh Gardens	12 (12)
Piper magnificum	Yale Marsh Gardens	8 (8)

[a] First value listed under the Clones/species heading indicates the total number of isolates characterized by restriction analysis and sequencing; the value in parentheses is the number of sequenced clones for each species (does not include 5′ or 3′ RACE fragments; see materials and methods).

the consensus tree were derived from the partition functions obtained from 1000 replicate bootstrapping runs. These runs were similarly produced using a heuristic search via random stepwise addition and under the TBR (tree-bisection-reconnection) algorithm for branch swapping.

Distance matrices were derived from the *PI* and *AP3* datasets under a mean character difference criterion. Amino acid substitutions were weighted in accordance with the BLOSUM substitution matrix. Trees were subsequently generated via the Neighbor-Joining algorithm (NJ) as implemented in PAUP. The resulting trees represent the "minimum evolution" networks connecting the sequences. Bootstrap values for resolved nodes are derived from 1000 replicate runs, again using the NJ algorithm.

For both the parsimony and distance analyses, the *AP3* trees were rooted using six *PI* sequences (*GLO*, *PI*, *DaPI*, *LtPI*, *PhPI* and *OsMADS2*); conversely, six *AP3* sequences (*AP3*, *DEFA*, *SLM3*, *MfAP3*, *PhAP3* and *CpAP3*) were used to root the *PI* trees. These choices for outgroups were based, first of all, on the fact that *AP3* and *PI* are known to be paralogous lineages and are, therefore, each other's natural outgroup. Secondly, while the *AP3* and *PI* orthologs can be reasonably aligned to each other throughout their entire length, unambiguous alignments cannot be generated between the B group sequences and the remaining members of the MADS-box gene family.

Similar analyses were conducted using nucleotide data sets aligned in accordance with the protein alignments (data not shown, see results).

RESULTS

Cloning of B-group genes from lower eudicots and magnolid dicots: Our strategy for cloning B group gene members from various lower eudicot and magnolid dicot species depends on the presence of a highly conserved sequence in the putative DNA-binding α-helix of the MADS-box domain (Shore and Sharrocks 1995). We designed a degenerate primer based on the decapeptide sequence NRQVTYSKR using the sequence of the published higher eudicot *AP3* and *PI* orthologs. This region of the MADS domain is largely invariant across the predicted products of the plant MADS-box genes examined to date and includes three amino acids that are invariant across all members of the MADS-box family (Doyle 1994; Shore and Sharrocks 1995). The tyrosine residue within this domain is diagnostic for the B group genes. The use of this B-group specific 5′ primer in conjunction with the completely nonspecific poly-T 3′ primer allowed for the amplification of *AP3* and *PI* orthologs with little bias against divergent paralogs. Since the predicted length of the eudicot B group members ranges from 180 to 230 amino acids, we examined all clones longer than approximately 600 bp.

We used *Syringa vulgaris* (lilac), a higher eudicot, as a positive control to test the primary PCR reaction and cloning strategy. Both *AP3* and *PI* orthologs were isolated from Syringa and their sequences showed very high similarity to that of the other Asterid B group representatives. For each lower eudicot and magnolid dicot species, we analyzed 10 to 36 clones (Table 1). It should be noted that in several species, an exhaustive survey has not been undertaken (<10 clones analyzed, Table 1). In the case of Lycopersicon, this is due to the fact that we were only interested in establishing the sequence of one gene (see below). Analyses of the Papaveraceae and Ranunculaceae were initiated with *Papaver californicum* and *Delphinium ajacis* but subsequent analyses were carried out with more convenient representative species (*P. nudicaule*, *R. bulbosus* and *Caltha palustris*).

A small number of the sequenced clones obtained from the primary PCR reactions proved not to be MADS-box gene representatives (5/171 total clones sequenced) and two were found to be members of the AGAMOUS-like family (one from *P. nudicaule* and one from *Delphinium ajacis*). Conceptual translation of all of the

TABLE 2

***PISTILLATA*-like genes**

Subclass	Family	Species	Gene	Ref.	Acc. no.
Asteridae	Scrophulariaceae	*Antirrhinum majus*	*GLO*	(Trobner *et al.* 1992)	S28062
	Solanaceae	*Petunia hybrida*	*FBP1*	(Angenent *et al.* 1992)	M91190
			PMADS2	(Kush *et al.* 1993)	X69947
		Nicotiana tabacum	*NTGLO*	(Hansen *et al.* 1993)	X67959
	Oleaceae	*Syringa vulgaris*	*SvPI*	This study	AF052861
Dilleniidae	Brassicaceae	*Arabidopsis thaliana*	*PI*	(Goto and Meyerowitz 1994)	D30807
Caryophyllidae	Caryophyllaceae	*Silene latifolia*	*SLM2*	(Hardenack *et al.* 1994)	X80489
Ranunculidae	Papaveraceae	*Papaver nudicaule*	*PnPI-1*	This study	AF052855
			PnPI-2	This study	AF052856
	Fumariaceae	*Dicentra eximia*	*DePI*	This study	AF052857
	Ranunculaceae	*Caltha palustris*	*CpPI*	This study	AF052858
		Ranunculus bulbosus	*RbPI-1*	This study	AF052859
			RbPI-2	This study	AF052860
		Delphinium ajacis	*DaPI*	This study	AF052862
Magnoliidae	Magnoliaceae	*Michelia figo*	*MfPI*[a]	This study	AF052863
		Liriodendron tulipifera	*LtPI*	This study	AF052864
	Piperaceae	*Peperomia hirta*	*PhPI*	This study	AF052865
		Piper magnificum	*PmPI-1*	This study	AF052866
			PmPI-2	This study	AF052867
Commelinidae	Poaceae	*Oryza sativa*	*OsMADS2*	(Chung *et al.* 1995)	L37526
			OsMADS4	(Chung *et al.* 1995)	L37527

[a] *MfPI* was only represented by a partial sequence at the time of submission and was not included in the phylogenetic analysis.

other cDNAs showed that they encode B group-type MADS-box gene products (164/171 total clones sequenced). We assigned the novel genes to the *AP3* or *PI* classes based on the overall sequence similarity to the known *AP3* and *PI* representatives and on the presence of specific diagnostic sites in the MADS, K and C domains. AP3 and PI-like proteins can be distinguished from one another at MADS-box residues 29, 35 and 47. Within the K box, PI homologs possess a highly conserved sequence KHExL (appendix 1, residues 88 to 92). The comparable sequence in the K box of the AP3 homologs is (H/Q)YExM (appendix 2, residues 85 to 89). We also found that the C-terminal portions of the predicted proteins contained diagnostic motifs for each lineage (see below). Each unique cDNA was named using the first letter of the genus and species from which it was isolated followed by either *AP3* or *PI*, depending on its sequence similarity.

Sequence and phylogenetic analysis of *PI* homologs: We have identified a total of twelve new *PI*-like genes which have been cloned from nine species (Table 2). In most of the species surveyed, only one *PI*-like gene was identified, while in *P. nudicaule*, *R. bulbosus* and *Piper magnificum*, two distinct *PI*-like genes were found. The sequences of the predicted products of the new *PI* clones align well with the previously studied higher eudicot and monocot representatives. Particularly striking is an approximately twenty amino acid region at the C-terminal end of the predicted proteins that displays extremely high conservation (Figure 3A). This domain, which we will refer to as the PI motif, has a core consensus sequence of MPFxFRVQPxQPNLQE. Four of the positions are completely invariant and the remainder show very strong conservation of chemical characteristics. Overall, the PI motif appears to be a strongly hydrophobic domain, although several charged and polar amino acids are present within the region. The motif bears no strong similarity to any known structural elements and a BLAST search for similar sequences in GenBank yields only *PI* homologs.

There are only two instances of marked divergence in sequence or structure among the newly isolated *PI* homologs. One is found in the predicted products of the *D. eximia* (Fumariaceae) and *P. nudicaule* (Papaveraceae) genes, *DePI* and *PnPI-1*, both of which contain an approximately 20 amino acid insertion upstream of the PI motif as well as a 10 to 12 amino acid addition at the C-terminal end of the protein (Figure 3A). These novel regions are characterized by stretches of A/C-rich repetitive DNA sequence. At the nucleotide level, the two cDNAs clearly align with one another through these novel regions, sharing 58% identity in the 100 bp upstream of the PI motif and 70% identity in the downstream region. It is likely that the event(s) that produced these insertions occurred before the last common ancestor of the Papaveraceae and Fumariaceae, which are sister families within the Papaverales. The second example of major sequence deviation within the *PI* homologs is seen in the *P. nudicaule* gene *PnPI-2*. The *PnPI-2* cDNA encodes a truncated protein of only 164 amino acids,

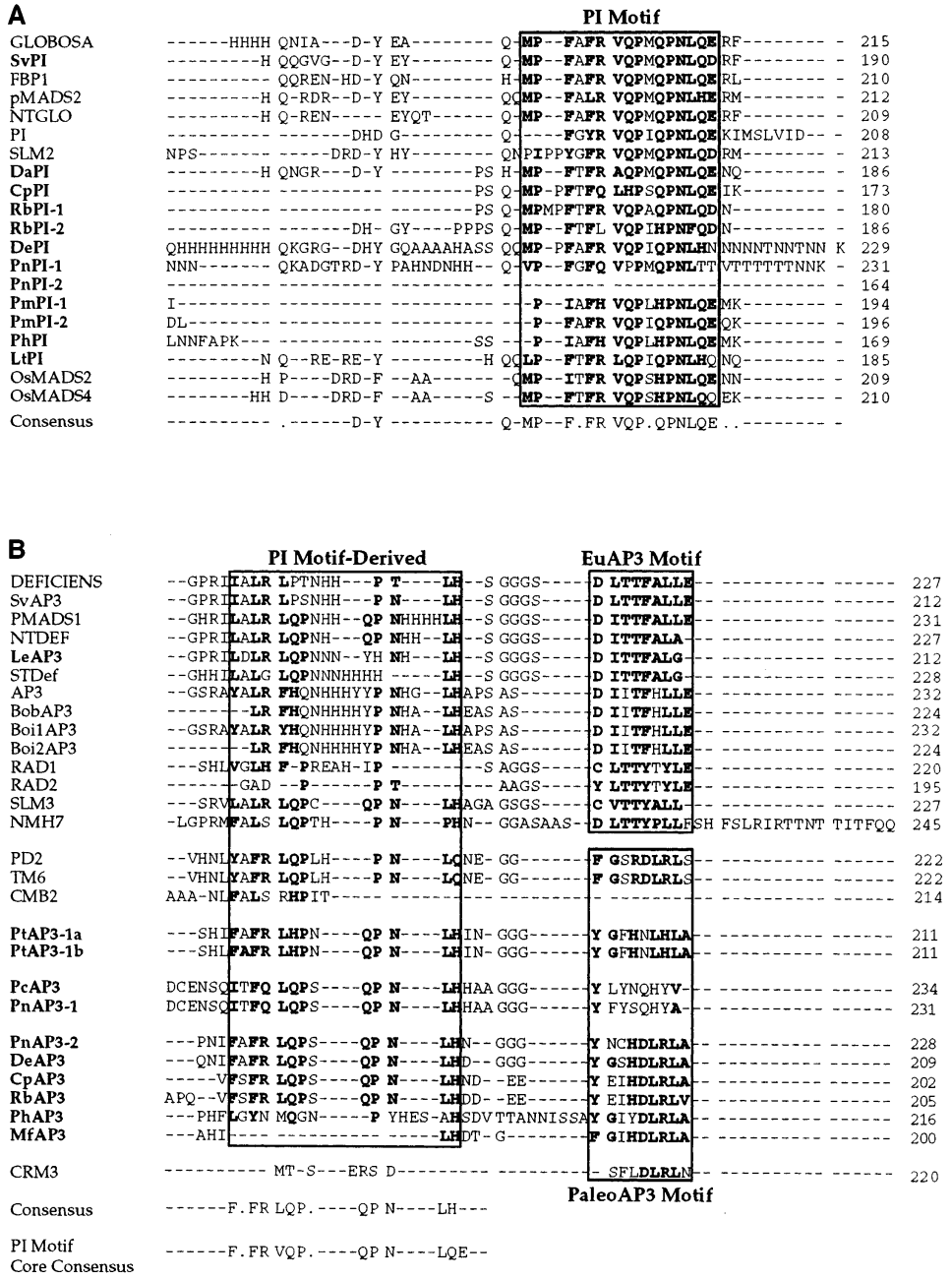

Figure 3.—Alignment of C-terminal regions of the predicted protein sequences analyzed in this study. The names of genes cloned in this study are highlighted in bold. The sequences of the unpublished genes *AsAP3* and *SILKY-1* are not shown. See Tables 1 and 2 for information on these genes. (A) Predicted PI protein sequences. The region which we have designated as the PI motif is boxed and the consensus is shown below. Residues which show chemical conservation with the consensus are highlighted in bold. (B) Predicted AP3 protein sequences. Genes are grouped according to lineage as indicated by phylogenetic analysis. The region which bears similarity to the PI motif is boxed and defined as PI Motif-Derived. Residues which show chemical conservation with the PI motif core consensus are highlighted in bold. The C-terminal euAP3 and paleoAP3 motifs are also indicated with boxes. Residues in each region which show chemical conservation with the motif consensus sequences defined in the text are highlighted in bold.

as opposed to the usual length of 180 to 210 amino acids. Alignment of the *PnPI-1* and *PnPI-2* nucleotide sequences reveals that the similarity between the *PnPI-1* and *PnPI-2* transcripts is quite high (80% identity) up to the point of the stop codon in *PnPI-2*, after which the similarity declines considerably.

Parsimony and distance-based phylogenies of the *PI* sequences were produced by analysis of the complete *PI* protein data set using several *AP3*-like sequences as the outgroup. The parsimony analysis resulted in 8 trees of equivalent length, from which a 50% majority rule consensus tree with a consistency index of 0.972 was derived, shown in Figure 4A. The topology of the parsimony tree is similar in many ways to that of the distance-based tree shown in Figure 4B. In both trees, all of the higher eudicot *PI* sequences are grouped into a single clade. The distance and parsimony trees do differ in the placement of the *PnPI-1* / *PnPI-2* paralog pair relative to the other Ranunculid *PI* orthologs. One unexpected result is the position of the *P. magnificum* gene, *PmPI-2*, as sister to the higher eudicot clade in the parsimony analysis. This is not the expected position of a representative of a basal magnolid dicot. The other Piperaceae representatives, *PmPI-1* and *PhPI*, are located at the base of the tree. In diverging from its paralog, *PmPI-2* may have independently acquired sequence motifs characteristic of the higher eudicot *PI* representatives. In the distance analysis, *PmPI-2* is placed at the base of the

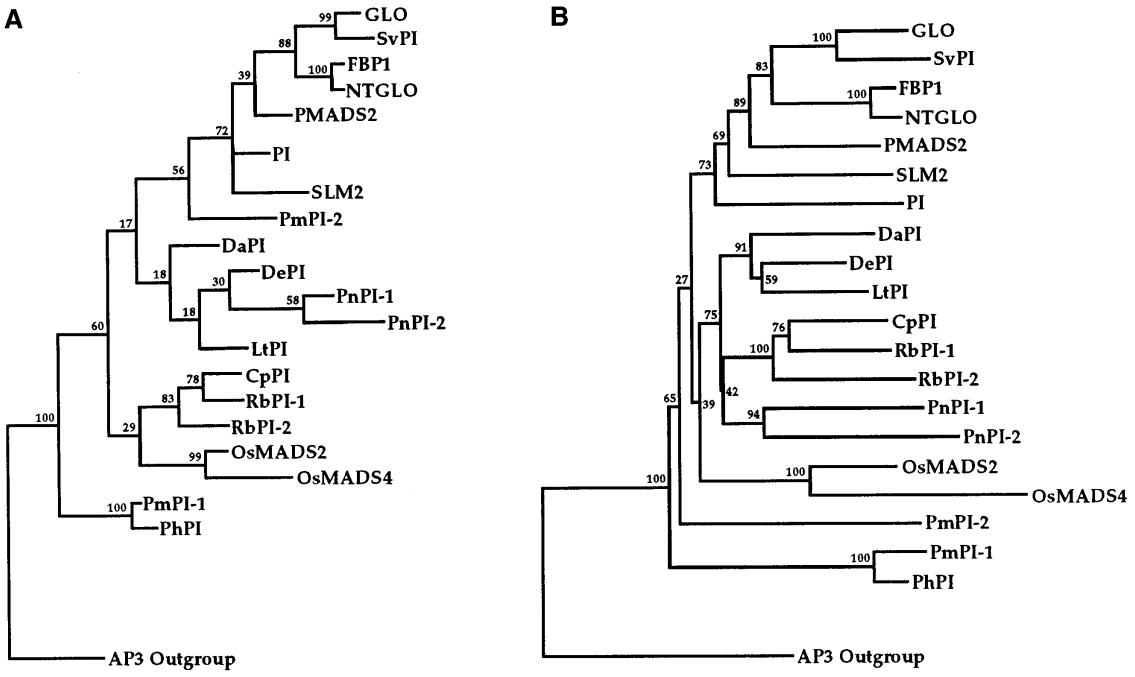

Figure 4.—Phylogenetic trees of the complete *PI* amino acid data set. The numbers next to the nodes give bootstrap values from 1000 replicates. (A) Parsimony analysis. The tree length is 1526 steps. (B) Distance analysis.

tree, closer to the other Piperaceae representatives. We note that the low bootstrap support seen for certain nodes reflects the restricted number of characters on which the node is based, rather than on the existence of alternative, better supported topologies. Analysis of a nucleotide *PI* data set yielded trees which displayed a significant number of unresolved polytomies due largely to the effects of saturation (data not shown). Where resolution was obtained using the nucleotide sequences, the structure of the consensus tree did not differ from that found using the protein sequences.

Sequence and phylogenetic analysis of the *AP3* homologs: The higher eudicot *L. esculentum* (tomato) has been previously found to contain an *AP3* paralog, *TM6*, which is considered to be orthologous to the *Solanum tuberosum* gene *PD2* (Pnueli *et al.* 1991; Garcia-Maroto *et al.* 1993). Solanum (potato) contains another gene, *STDEF*, which appears to be orthologous to the *AP3* lineage members of the other Asterids and higher eudicots. Lycopersicon was examined in an attempt to find an *AP3*-like gene which would be more similar to the other higher eudicot *AP3* orthologs than the previously described *TM6*. We recovered several clones of a cDNA, *LeAP3* whose predicted product displays 93% amino acid identity to that of *STDEF* from Solanum (Garcia-Maroto *et al.* 1993). This level of similarity is comparable to that seen between the TM6 and PD2 gene products from Lycopersicon and Solanum, respectively. This result confirms the presence of two separate *AP3*-related genes in both tomato and potato.

We have also cloned eleven *AP3*-like genes from a total of nine lower eudicot and magnolid dicot species (Table 3). *P. terminalis* and *P. nudicaule* were the only species that we found to have two *AP3*-like genes. The Ranunculid and magnolid dicot *AP3* representatives show many differences when compared to the higher eudicot *AP3* orthologs. Foremost among these are the considerable differences in the sequences of the predicted C-termini. The higher eudicot *AP3* gene products display a highly conserved C-terminal motif with the consensus sequence D(L/I)TTFALLE (euAP3 motif, Figure 3B). The majority of the *AP3* clones from the Ranunculidae and the magnolid dicots have a completely different predicted C-terminal sequence with the consensus YGxHDLRLA (paleoAP3 motif, Figure 3B). The predicted product of an *AP3*-like gene from *Zea mays, SILKY-1*, also displays the paleoAP3 motif with only one amino acid difference (C. Padilla and R. Schmidt, personal communication). The paleoAP3 motif is slightly divergent, but recognizable, in *PtAP3-1* and *PtAP3-2* from *P. terminalis*. Interestingly, the paleoAP3 motif aligns very well with the C-terminal end of the predicted *TM6* and *PD2* proteins (Figure 3B). Similarity to the paleoAP3 motif is also seen in the product of a putative B group-related gene which has been isolated from the Pteridophyte (fern) Ceratopteris (Munster *et al.* 1997; Figure 3B). The product of this fern gene, CRM3, has relatively low similarity to the *AP3* gene products through most of its length, but the predicted C-terminal end aligns well with the paleoAP3 motif. This observation is remarkable considering the fact that the Pteridophytes diverged from the land plants

TABLE 3

***APETALA3*-like genes**

Subclass	Family	Species	Gene	Ref.	Acc. no.
Asteridae	Scrophulariaceae	*Antirrhinum majus*	*DEF*	(Schwartz-Sommer *et al.* 1992)	S12378
	Solanaceae	*Lycopersicon esculentum* cv. Celebrity	*LeAP3*	This study	AF052868
		Lycopersicon esculentum cv. Tiny Tim	*TM6*	(Pnueli *et al.* 1991)	X60759
		Petunia hybrida	*PMADS1*	(van der Krol *et al.* 1993)	X69946
		Solanum tuberosum	*STDEF*	(Garcia-Maroto *et al.* 1993)	X67511
			PD2	(Garcia-Maroto *et al.* 1993)	None
		Nicotiana tabacum	*NTDEF*	(Davies *et al.* 1996)	X96428
	Asteraceae	*Argyroxiphium sandwicense*	*AsAP3*	Unpublished results[b]	
	Oleaceae	*Syringa vulgaris*	*ScAP3*	This study	AF052869
Dilleniidae	Brassicaceae	*Arabidopsis thaliana*	*AP3*	(Jack *et al.* 1992)	A42095
		Brassica oleraceae var. boytris	*BobAP3*	(Carr and Irish 1997)	BOU67456
		Brassica oleraceae var. italica	*Boi1AP3*	(Carr and Irish 1997)	BOU67453
			Boi2AP3	(Carr and Irish 1997)	BOU67455
Caryophyllidae	Caryophyllaceae	*Silene latifolia*	*SLM3*	(Hardenack *et al.* 1994)	X80490
		Dianthus caryophyllus	*CMB2*	Unpublished results[c]	L40405
	Polygonaceae	*Rumex acetosa*	*RAD1*	(Ainsworth *et al.* 1995)	X8913
			RAD2	(Ainsworth *et al.* 1995)	X89108
Rosidae	Fabaceae	*Medicago sativa*	*NMH7*	(Heard and Dunn 1995)	L41727
	Buxaceae	*Pachysandra terminalis*	*PtAP3-1*	This study	AF052870
			PtAP3-2	This study	AF052871
Ranunculidae	Papaveraceae	*Papaver californicum*	*PcAP3*	This study	AF052872
		Papaver nudicaule	*PnAP3-1*	This study	AF052873
			PnAP3-2	This study	AF052874
	Fumariaceae	*Dicentra eximia*	*DeAP3*	This study	AF052875
	Ranunculaceae	*Caltha palustris*	*CpAP3*	This study	AF052854
		Ranunculus bulbosus	*RbAP3*	This study	AF052876
Magnoliidae	Magnolicaceae	*Michelia figo*	*MfAP3*	This study	AF052877
		Liriodendron tulipifera	*LtAP3*[a]	This study	AF052878
	Piperaceae	*Peperomia hirta*	*PhAP3*	This study	AF052879
Commelinidae	Poaceae	*Zea mays*	*SILKY-1*	Unpublished results[d]	

[a] *LtAP3* was only represented by a partial sequence at the time of submission and was included in the phylogenetic analysis.
[b] M. Barrier and M. Purugganan, personal commmunication.
[c] S. C. Baudinette and K. W. Savin, unpublilshed results.
[d] C. Padilla and R. Schmidt, personal communication.

roughly 400 million years ago (mya) (Stewart and Rothwell 1993). The Papaver clones *PnAP3-1* and *PcAP3* are unique among the Ranunculid *AP3* genes in that they do not show high conservation of the paleoAP3 motif.

One other motif that we recognized in the new *AP3* gene products is a region with similarity to the PI motif. The Ranunculid *AP3* representatives have a region upstream of the paleoAP3 motif with the consensus sequence FxFRLQPSQPNLH. This is very similar to the core of the PI motif consensus sequence (Figure 3B). When all of the *AP3*-like gene products are aligned, the PI motif-derived sequence can be observed in almost all of the *AP3*-like proteins. While this sequence has been very highly conserved in the *PI* lineage, in the *AP3* lineage it has diverged to differing degrees. *TM6* and *PD2* show a much greater conservation of the PI motif than do the other higher eudicot *AP3* gene products. The PI motif is particularly divergent in *RAD1* and *RAD2* from Rumex, while in *CMB2*, an *AP3* paralog from Dianthus, a truncation has eliminated most of the PI motif-derived sequence. In the Michelia and Peperomia *AP3* representatives, the PI motif sequence is unrecognizable.

Figure 5, A and B, show the results of parsimony and distance analysis of the complete *AP3* protein data set. The parsimony analysis produced 27 trees of equivalent length, from which we generated a 50% majority rule consensus tree with a consistency index of 0.89. The topmost clade in both the parsimony and distance trees contains a grouping of the higher eudicot Brassicaceae, Asterid and Caryophyllid *AP3* orthologs (Figure 5A). The members of this clade will be referred to as the eu*AP3* lineage. The second main clade in both trees is made up of several other higher eudicot *AP3*-like genes: *TM6*, *PD2*, *AsAP3* and *CMB2*. The orthology of *PD2* and

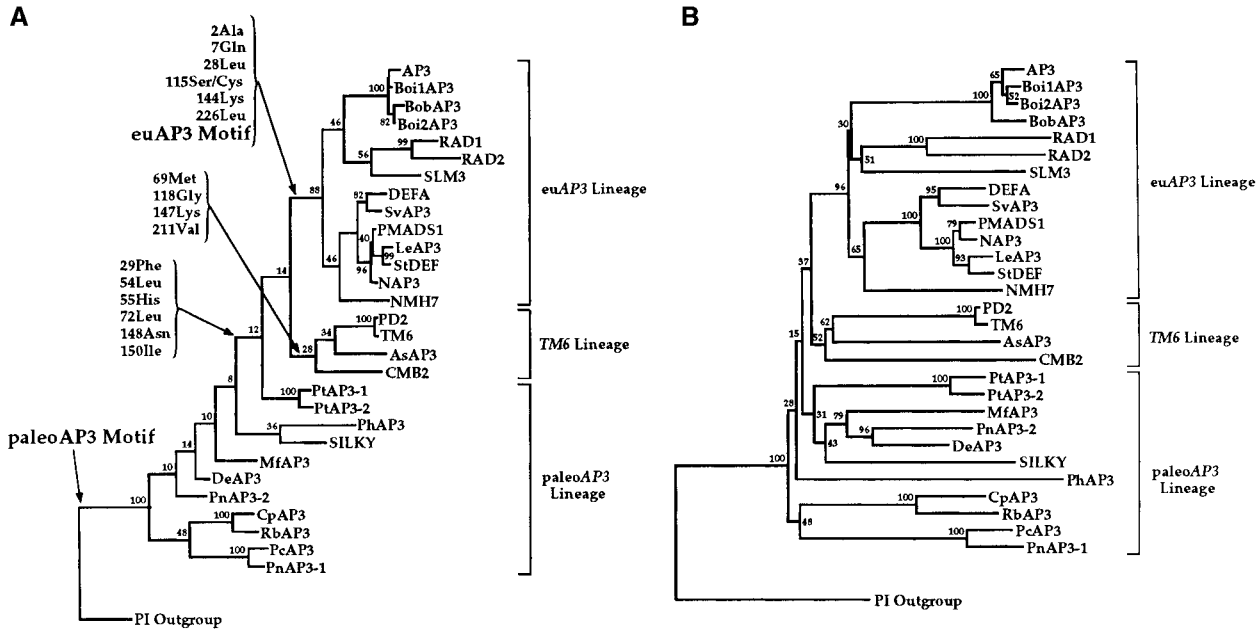

Figure 5.—Phylogenetic trees of the complete *AP3* data set. The numbers next to the nodes give bootstrap values from 1000 replicates. (A) Parsimony analysis. The tree length is 1781 steps. The distribution of diagnostic sequence characters is mapped onto the tree. (B) Distance analysis.

TM6 has been suggested previously (Garcia-Maroto *et al.* 1993). Likewise, a recent analysis of the relationships of the MADS-box genes also found an association between *AsAP3* and *TM6* (Purugganan 1997). *CMB2* is an *AP3*-like gene that was isolated from *Dianthus caryophyllus.* The inclusion of *CMB2* in the *TM6* lineage is supported by several positions which are synapomorphic for this clade, including 69Met, 118Gly, 147Lys and 211Val. Our analyses define all of these genes as descendants of a lineage that we will refer to as the *TM6* lineage.

In the parsimony analysis, the remaining *AP3*-like genes of the lower eudicots and magnolid dicots constitute a paraphyletic group at the base of the tree. The two very similar Pachysandra genes are paired together and are the most closely related to the higher eudicot clades. In contrast, the distance analysis defines a clade composed of *PtAP3-1*, *PtAP3-2* and the Ranunculid and Magnoliaceae sequences. This result highlights the similarities between the Pachysandra and other lower eudicot sequences. In both trees, *PnAP3-1* and *PcAP3* from *P. nudicaule* and *P. californicum*, respectively, are paired together with very high certainty, suggesting that they define a paralogous lineage. Similar to the *PI* analysis, the low bootstrap support found for several nodes reflects the limited number of characters on which the node is based. All resolved nodes shown in Figure 5A were present in all equally parsimonious trees. As with *PI*, a nucleotide *AP3* data set was analyzed, but the resulting trees were highly unresolved (data not shown). Where resolution was obtained using the nucleotide sequences, the structure did not differ from that found using the protein sequences.

If the character of each representative's predicted C-terminal motif is mapped on Figure 5A, we see that the paleoAP3 motif is present throughout the basal paraphyly and along the *TM6* lineage. We will refer to the lower dicot and magnolid dicot genes which exhibit the paleoAP3 motif as the paleo*AP3* lineage. The euAP3 motif is synapomorphic for the clade defining what we have called the eu*AP3* lineage. Diagnostic amino acid character states from throughout the protein sequence can be mapped onto the tree as well. Although the C-termini of the predicted *PtAP3-1* and *PtAP3-2* proteins resemble the paleoAP3 motif, in other positions the sequence is more similar to the eu*AP3* genes. These eu*AP3*-like positions include 54Leu, 55His, 72Leu, 148-Asn and 150Ile. The eu*AP3* clade is also defined by several synapomorphic residues, aside from the C-terminal euAP3 motif. These changes include 2Gly→Ala, 7Glu→Gln, 28Ile→Leu, 115Asp→Ser/Cys, 144His→Lys and 226Phe→Leu. Interestingly, the primitive states of positions 2 and 7 in the MADS box are identical to the character state in the *PI* lineage members.

DISCUSSION

Morphological evolution proceeds by changes in developmental pathways due to alterations in the structure and regulation of particular developmental genes. The organ identity genes of the A, B, and C classes are logical starting points for a study of changing floral morphologies. The B group genes, in particular, appear to be a likely site of developmental flexibility in floral evolution. While A and C group members might be expected to

show very high levels of constraint due to their pleiotropic roles in floral meristem identity and determinacy, as well as in organ identity, the functions of the B group genes appear to be limited to establishing organ identity. Moreover, the fact that the petals and stamens display enormous morphological plasticity within the angiosperms may be directly reflected as changes in the B group genes. These predictions are supported by substitution rate analysis, which shows that the *AP3*/*PI* group is evolving 20–40% faster than are all of the other plant MADS-box genes (Purugganan 1997). Furthermore, although the B group genes appear to have a conserved function in determining petal identity in the higher eudicots, these petals are not thought to be homologous with those of the lower eudicots, magnolid dicots and monocots. The independent derivation events that gave rise to the different petals of the angiosperms may be reflected in the diversified structure and function of the B group genes.

In this study we present the isolation of twenty-six new B group genes from a total of thirteen species distributed throughout the angiosperms. One of the most striking findings of the analysis of these genes is the high frequency of gene duplications within both the *AP3* and *PI* lineages. We have identified several key ancestral duplications, in addition to what appear to be more recent duplication events. Pinpointing the exact time at which these duplications have occurred is difficult, however. The functional redundancy that results from having multiple gene copies can serve to release constraint on the sequence of the paralogs. Thus, although a high sequence similarity between paralogs would seem to suggest a relatively recent duplication event, low sequence similarity could reflect either an ancient duplication or the rapid divergence of a paralogous lineage. If we can identify orthologs of both duplication products in more than one species, then it is clear that the duplication event occurred before the last common ancestor of the species in question.

The B group gene ancestor and the *AP3*/*PI* duplication event: The *AP3* and *PI* genes have previously been shown to be members of closely paralogous gene lineages (Doyle 1994; Purugganan *et al.* 1995; Purugganan 1997; Theissen *et al.* 1996). We report further evidence for the paralogy of the *AP3* and *PI* lineages, most notably the presence of the PI motif-derived sequence in the *AP3* homologs. Based on the common sequence characteristics of the *AP3* and *PI* lineages, we can reconstruct some of the characters of the B group ancestral gene lineage. The basal representatives of both lineages contain the PI motif as well as several diagnostic amino acids in the MADS domain (2Gly, 7Glu). This would suggest, based on parsimony, that the most recent B group ancestor also displayed these traits. The predicted product of the *CRM3* gene from Pteridophyte Ceratopteris may reveal another aspect of the B group ancestor. If the CRM3 C-terminal domain is homologous to the paleoAP3 motif, then we must add possession of the paleoAP3 motif to the list of characteristics of the B group ancestor. The fact that the predicted product of *CRM3* does not possess the PI motif could reflect an origin of this domain that postdates the split of the ferns from the land plants (Figure 6).

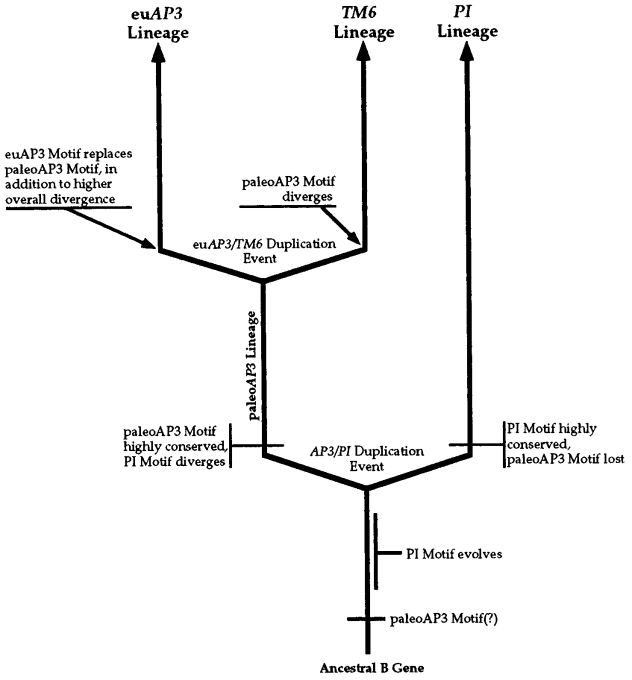

Figure 6.—Model of B Group MADS-box gene evolution. The hypothesized appearance and divergence of various diagnostic sequences are indicated on a simplified lineage tree.

All of the species which we have thoroughly surveyed contain both an *AP3* and a *PI* representative, suggesting that the duplication event which produced these two lineages predates the diversification of the angiosperms. Initially after the duplication event, both of the products would have exhibited the PI motif, the characteristic MADS-box residues and, possibly, the paleoAP3 motif. In the *PI* lineage, the PI motif and the MADS-box characters appear to have been highly conserved throughout while the paleoAP3 motif was lost, possibly by a single truncation event. In contrast, along the *AP3* lineage, the PI motif diversifies dramatically, along with additional changes throughout the protein (Figure 6).

The *PI* lineage: The high degree of conservation of the PI motif throughout the members of the *PI* lineage suggests that it has a critical function. The requirement of the C terminus for *in vivo* function has been demonstrated for the products of several plant MADS-box genes, including *AP3* (Krizek and Meyerowitz 1996; Riechmann and Meyerowitz 1997). *In vitro* studies have shown, however, that the C regions of the AP3 and PI proteins are not required for heterodimerization or the binding of the complex to DNA (Krizek and Meyerowitz 1996; Riechmann *et al.* 1996b). Furthermore,

none of the *PI* gene mutants isolated in Arabidopsis or Antirrhinum has lesions in this region (Goto and Meyerowitz 1994; Jack *et al.* 1992; Sommer *et al.* 1990; Trobner *et al.* 1992). For these reasons, it is difficult to speculate on the functions of the PI motif, aside from suggesting that it is involved in protein interactions which are important for the overall function of the *PI* orthologs.

Many other duplication events have followed the creation of the separate *AP3* and *PI* lineages. In the *PI* lineage, five different paralog pairs have been identified, in *Petunia hybrida, R. bulbosus, P. nudicaule, P. magnificum* and *Oryza sativa.* The phylogenetic analysis suggests that these are the products of independent duplication events which occurred after the last common ancestor of any of the species included in this analysis. In two of these species, *P. hybrida* and *O. sativa*, there is some evidence for functional divergence of the paralogs (Angenent *et al.* 1993; Chung *et al.* 1995). Although polyploidy is a common cause for the presence of paralogs within plant genomes, none of these species are known polyploids (Bennett and Smith 1976, 1991; Bennett and Leitch 1995). Duplications such as these, as well as insertion events like those seen in the *PI* orthologs of *P. nudicaule* and *D. eximia,* may be useful as synapomorphies to define taxonomic groups.

The *AP3* lineage: We have defined two distinct *AP3* lineages which appear to be present throughout the higher eudicots. These are clearly the products of an ancient duplication event that occurred before the diversification of the higher eudicots. We cannot rule out the possibility that this duplication event occurred very early in the angiosperms and we have simply not detected the members of the eu*AP3* lineage in the lower eudicots or magnolid dicots. The sequence and phylogenetic analyses do not support an extremely early duplication, however. Most notably, the Pachysandra *PtAP3-1* and *PtAP3-2* genes appear to represent a mosaic of paleo*AP3* characters and characters of the eu*AP3* and *TM6* lineages of the higher eudicots. Although the C-terminal regions of the predicted products of the Pachysandra genes bear more similarity to the paleoAP3 motif, other positions in the protein sequences ally them with the higher eudicot genes. Based on these observations, we propose that the duplication which produced the eu*AP3* and *TM6* lineages from a paleo*AP3* ancestor occurred after the last common ancestor of the Buxaceae and the higher eudicots but before the diversification of the major higher eudicot subclasses (Figure 7).

We see the results of additional taxon-specific duplication events in the *AP3* lineages of *Brassica oleracea, Rumex acetosa,* and *P. terminalis.* The *AP3* paralog pairs of these species all show relatively high levels of sequence similarity. We have also identified several more divergent paralogous *AP3* lineages. One such lineage is defined by *PcAP3* and *PnAP3-1* from *P. californicum* and *P. nudicaule,* respectively. Other than being florally expressed, the

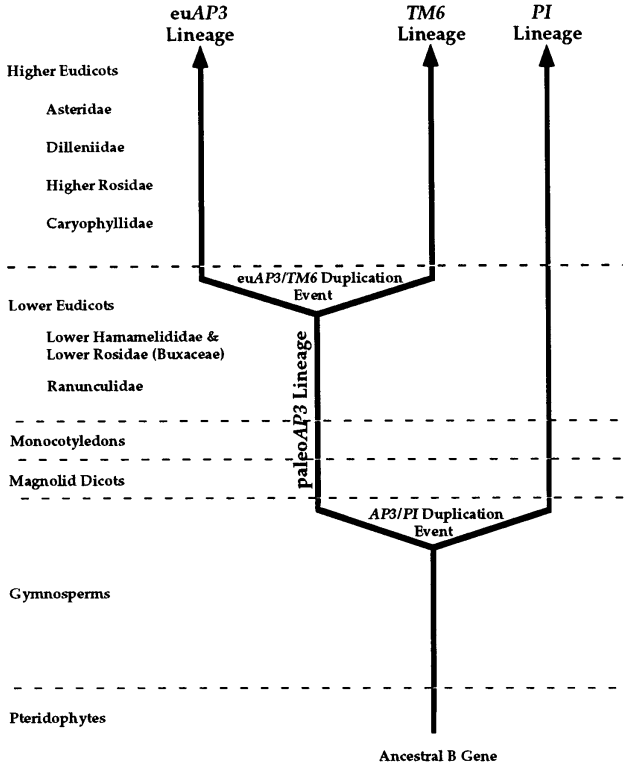

Figure 7.—Model showing correlation of B Group gene duplication events and land plant evolution. The divisions of the simplified gene phylogeny indicate which B group lineage representatives have been identified in the extant plant groups listed on the left.

role of these Papaver genes is unknown. Evidence for a functionally divergent *AP3* paralog is seen in the *Medicago sativa* gene, *NMH7* (Heard and Dunn 1995). The sequence of *NMH7* is highly diverged from the sequence of the other eu*AP3* lineage members, and may reflect the fact that the function of this paralog has also diverged considerably, being involved in mediating root nodulation (Heard and Dunn 1995). Our analysis, which is based on amplifying floral cDNAs, would not have detected any paralogs such as *NMH7.*

Implications of the *AP3* duplication event and subsequent divergence: While the *PI* lineage appears to be highly conserved, the *AP3* lineage has experienced significant diversification, some of which clearly correlates with duplication events. If the fern *CRM3* gene is, in fact, a representative of the B group ancestral lineage which possesses a paleoAP3 motif, then this small region has been conserved, to some degree, from the Pteridophytes, dating back some 400 mya, up through the lower Rosids, which first appeared approximately 80 mya (Drinnan *et al.* 1994; Stewart and Rothwell 1993). This highly conserved motif was lost, however, in the eu*AP3* lineage where we see fixation of a new conserved C-terminal motif. In addition, we see changes throughout the predicted products of the eu*AP3* lineage genes, including the MADS-box domain. The eu*AP3* lineage

would, therefore, appear to have undergone considerable sequence divergence in regions that were highly conserved, some for at least 300 million years. We assume that this reflects a shift in the functional repertoire of the eu*AP3* lineage members relative to the ancestral paleo*AP3* representatives.

There are many possible ways in which the functions of the eu*AP3*, paleo*AP3* and *TM6* lineage members may differ. The eu*AP3* lineage members function primarily in the establishment of petal and stamen identity. The role played by eu*AP3* representatives in stamen identity is likely to be the ancestral function of these genes since stamens are thought to have evolved only once, before the diversification of the angiosperms. In contrast, the function of eu*AP3* lineage members in petal identity may be more recently acquired, reflecting a *de novo* evolution of petals at the base of the higher eudicot radiation. Similarly, it seems likely that the paleo*AP3* lineage members also function in stamen identity, but each time that petals were derived, whether from stamens or from bracts, the paleo*AP3* ortholog(s) may have been recruited to new roles. In the Ranunculidae, for example, andropetals are thought to have been derived many times, even within single families (Drinnan *et al.* 1994). Each of these events may have resulted in changes of the functions of the B group genes. The functions of the *TM6* lineage members in the higher eudicots are not well understood. Expression patterns have only been characterized for *TM6* itself, which is highly expressed in the developing petal, stamen and carpel, but a role in the formation of these organs has not been established (Pnueli *et al.* 1991). Our analysis reveals that the *TM6* lineage members still retain sequence similarity to the ancestral paleo*AP3* lineage, but the individual representatives appear to have undergone significant diversification since the eu*AP3*/*TM6* duplication event.

These results reveal a complex pattern of gene duplication and divergence in the *AP3* and *PI* lineages. While it appears that the euAP3 motif is found in all of the higher eudicot *AP3* homologs that are known to play a role in determining petal and stamen identity, it is the paleoAP3 motif that defines the *AP3* homologs of the lower eudicots and magnolid dicots. To understand what changes in function may underlie the sequence divergence observed in the eu*AP3* lineage relative to the ancestral paleo*AP3* lineage, investigation of the function of the paleo*AP3* lineage members in the lower eudicots and magnolid dicots will be necessary. These and other comparative developmental approaches are critical to defining the relationships between gene duplication, gene expression and the morphological diversification that is the hallmark of the angiosperm radiation.

The authors thank the members of the Irish lab for helpful discussions. We are very grateful to C. Padilla, R. Schmidt, M. Barrier and M. Purugganan for providing unpublished sequence information. We also thank David Swofford for kindly allowing us to use Version 61 of PAUP 4.0*. This work was supported by grants from the National Science Foundation (IBN 9630916) and United States Department of Agriculture (97 01286) to V.F.I., a National Science Foundation graduate fellowship to E.M.K. and the Yale Science Development Fund to R.L.D.

LITERATURE CITED

Ainsworth, C., S. Crossley, V. Buchanan-Wollaston, M. Thangavelu and J. Parker, 1995 Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression. Plant Cell **7:** 1583–1598.

Angenent, G. C., M. Busscher, J. Franken, J. N. M. Mol and A. J. van Tunen, 1992 Differential expression of two MADS box genes in wild-type and mutant petunia flowers. Plant Cell **4:** 983–993.

Angenent, G. C., J. Franken, M. Bussher, L. Colombo and A. J. van Tunen, 1993 Petal and stamen formation in petunia is regulated by the homeotic gene fbp1. Plant J. **4:** 101–112.

Bennett, M. D., and I. J. Leitch, 1995 Nuclear DNA amounts in angiosperms. Ann. Bot. **76:** 113–176.

Bennett, M. D., and J. B. Smith, 1976 Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. Lond. B Biol. Sci. **274:** 227–274.

Bennett, M. D., and J. B. Smith, 1991 Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. Lond. B Biol. Sci. **334:** 309–345.

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.

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

Carr, S. M., and V. F. Irish, 1997 Floral homeotic gene expression defines developmental arrest stages in *Brassica oleracea* L. vars. *botrytis* and *italica.* Planta **201:** 179–188.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les *et al.*, 1993 Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Ann. Miss. Bot. Gard. **80:** 528–580.

Chung, Y.-Y., S.-R. Kim, H.-G. Kang, Y.-S. Noh, M. Park *et al.*, 1995 Characterization of two rice MADS box genes homologous to *GLOBOSA.* Plant Sci. **109:** 45–56.

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

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

Cronquist, A., 1981 *An Integrated System of Classification of Flowering Plants.* Columbia University Press, New York.

Cronquist, A., 1988 *The Evolution and Classification of Flowering Plants.* The New York Botanical Garden, New York.

Dahlgren, R. M. T., H. T. Clifford and P. F. Yeo, 1984 *The Families of the Monocotyledons.* Springer Verlag, New York.

Davies, B., and Z. Schwarz-Sommer, 1994 Control of floral organ identity by homeotic MADS-box transcription factors. Results Probl. Cell Differ. **20:** 235–258.

Davies, B., A. Di Rosa, T. Eneva, H. Saedler and H. Sommer, 1996 Alteration of tobacco floral organ identity by expression of combinations of *Antirrhinum* MADS-box genes. Plant J. **10:** 663–677.

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

Drinnan, A. N., P. R. Crane and S. B. Hoot, 1994 Patterns of floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots), pp. 93–122 in *Early Evolution of Flowers,* edited by P. K. Endress and E. M. Friis. Springer-Verlag, New York.

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

Friis, E. M., and P. K. Endress, 1990 Origin and evolution of angiosperm flowers. Adv. Bot. Res. **17:** 99–162.

Garcia-Maroto, F., F. Salamini and W. Rohde, 1993 Molecular cloning and expression patterns of the *Deficiens* homologous gene *St-Deficiens* from *Solanum tuberosum.* Plant J. **4:** 771–780.

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

Hansen, G., J. Estruch, H. Sommer and A. Spena, 1993 *NTGLO*: a tobacco homologue of the *GLOBOSA* floral homeotic gene of *Antirrhinum majus*: cDNA sequence and expression pattern. Mol. Gen. Genet. **239:** 310–312.

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

Heard, J., and K. Dunn, 1995 Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc. Natl. Acad. Sci. USA **92:** 5273–5277.

Hill, T., C. D. Day, S. C. Zondlo, A. G. Thackeray and V. F. Irish, 1998 Discrete spatial and temporal cis-acting elements regulate transcription of the *Arabidopsis* floral homeotic gene *APETALA3.* Development **125:** 1711–1721.

Irish, V. F., and E. M. Kramer, 1998 Genetic and molecular analysis of angiosperm flower development. Adv. Bot. Res. **28:** 197–230.

Irish, V. F., and Y. T. Yamamoto, 1995 Conservation of floral homeotic gene function between *Arabidopsis thaliana* and *Antirhinum majus.* Plant Cell **7:** 1635–1644.

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

Komaki, M., K. Okada, E. Nishino and Y. Shimura, 1988 Isolation and characterization of novel mutants of *Arabidopsis thaliana* defective in flower development. Development **104:** 195–203.

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.

Kush, A., A. Brunelle, D. Shevell and N.-H. Chua, 1993 The cDNA sequence of two MADS box proteins in petunia. Plant Physiol. **102:** 1051–1052.

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.

McGonigle, B., K. Bouhidel and V. F. Irish, 1996 Nuclear localization of the Arabidopsis APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. Genes Dev. **10:** 1812–1821.

Meyerowitz, E. M., J. L. Bowman, L. L. Brockman, G. N. Drews, T. Jack *et al.*, 1991 A genetic and molecular model for floral development in *Arabidopsis thaliana.* Development **1:** 157–167.

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.

Norman, A., M. Runswick, R. Pollock and R. Treisman, 1988 Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. Cell **55:** 989–1003.

Pnueli, L., M. Abu-Abeid, D. Zamir, W. Nacken, Z. Schwarz-Sommer *et al.*, 1991 The MADS box gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from *Antirrhinum* and *Arabidopsis.* Plant J. **1:** 255–266.

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

Purugganan, M. D., 1997 The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. J. Mol. Evol. **45:** 392–396.

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., M. W. Chase, D. H. Les and C. R. Parks, 1993 Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene rbcL. Ann. Miss. Bot. Gard. **80:** 587–606.

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

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.

Samach, A., S. E. Kohalmi, P. Motte, R. Datla and G. W. Haughn, 1997 Divergence of function and regulation of class B floral organ identity genes. Plant Cell **9:** 559–570.

Schwarz-Sommer, Z., I. Hue, P. Huijser, P. J. Flor, R. Hansen *et al.*, 1992 Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J. **11:** 251–263.

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

Sommer, H., J.-P. Beltrán, P. Huijser, H. Pape, W.-E. Lönnig *et al.*, 1990 Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. EMBO J. **9:** 605–613.

Stewart, W. N., and G. W. Rothwell, 1993 *Paleobotany and the Evolution of Plants*, Ed. 2, Cambridge University Press, Cambridge.

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

Takhtajan, A., 1991 *Evolutionary Trends in Flowering Plants.* Columbia University Press, New York.

Theissen, G., and H. Saedler, 1995 MADS-box genes in plant ontogeny and phylogeny: Haekel's 'biogenetic law' revisited. Curr. Opin. Gen. Dev. **5:** 628–639.

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.

Trobner, W., L. Ramirez, P. Motte, I. Hue, P. Huijser *et al.*, 1992 Globosa—a homeotic gene which interacts with deficiens in the control of antirrhinum floral organogenesis. EMBO J. **11:** 4693–4704.

van der Krol, A. R., and N.-H. Chua, 1993 Flower development in petunia. Plant Cell **5:** 1195–1203.

van der Krol, A. R., A. Brunelle, S. Tsuchimoto and N. H. Chua, 1993 Functional analysis of petunia floral homeotic mads box gene PMADS1. Genes Dev. **7:** 1214–1228.

Communicating editor: E. Meyerowitz

APPENDIX 1

Alignment of all *PI* lineage members with selected *AP3* outgroup representitives

```
                1                                                          60
GLO             MGRGKIEIKR IENTSNRQVT YSKRKNGIMK KAKEISVLCD AHVSVIIFAS SG--KMHEFC
SvPI            ---------- ---------- -----NGLMK KAKEISVLCD AQVSVVIFAS SG--KMHEFC
FBP1            MGRGKIEIKR IENSSNRQVT YSKRRNGILK KAKEISVLCD ARVSVIIFAS SG--KMHEFS
pMADS2          MGRGKIEIKR IENSSNRQVT YSKRRNGIIK KAKEITVLCD AKVSLIIFGN SG--KMHEYC
NTGLO           MGRGKIEIKR IENSSNRQVT YSKRRNGILK KAKEISVLCD ARVSVIIFAS SG--KMHEFS
PI              MGRGKIEIKR IENANNRVVT FSKRRNGLVK KAKEITVLCD AKVALIIFAS NG--KMIDYC
SLM2            MGRGKIEIKR IENSTNRQVT YSKRRNGIIK KAGEITVLCE AKVSLIIFSN NG--KMHAYH
DaPI            ---------- ---------- -----NGILK KAKEITVLCD AEVSLVVFSS TG--KMSEFI
CpPI            ---------- ---------- -----NGIIK KAKEIAILCA AEVSLVIFSS SD--KMSEF-
RbPI-1          ---------- ---------- -----NGILK KAKEIAILCE AKVSLVIFTS NS--KMFEF-
RbPI-2          ---------- ---------- -----NGILK KAKEISILCS AEVSLVIFSC NN--KMSEF-
DePI            ---------- ---------- -----NGIMK KAKEITVLCD AEVSLVIFSS TG--KMSEYC
PnPI-1          MGRGKIEIKR IENSTNRQVT YSKRKNGILK KAKEITILCD AHVSLVIFSS TG--KMNEYC
PnPI-2          ---------- ---------- -----NGILK KAKEISILCD ANLSLVMISE AG--SIDEYS
PmPI-1          ---------- ---------- -----KGIIK KAQEISVLCD TQVSLVIFSS AG--NMGEFC
PmPI-2          ---------- ---------- -----AGLLK KAREISVLCD AEVALVIVSS SANVKVEEFC
PhPI            ---------- ---------- -----KGIIK KAQEISVLCD THVSVLIFSS AG--NMGEFC
LtPI            ---------- ---------- -----GGILK KAKEITVLCD AQVSLVIFSS TG--KISEYC
OsMADS2         MGRGKIEIKR IENSTNRQVT FSKRRSGILK KAREISVLCD AEVGVVIFSS AG--KLYDYC
OsMADS4         MGRGKIEIKR IENSTNRQVT FSKRRSGILK KAREIGVLCD AEVGVVIFSS AG--KLSDYC

AP3             MARGKIQIKR IENQTNRQVT YSKRRNGLFK KAHELTVLCD ARVSIIMFSS SN--KLHEYI
DEFA            MARGKIQIKR IENQTNRQVT YSKRRNGLFK KAHELSVLCD AKVSIIMISS TQ--KLHEYI
SLM3            MARRKIQIKK IENLTNRQVT YSKRRNGLFK KANELTVLCD ATVSIIMLSS NL--KLHEFL
CpAP3           ---------- ---------- -----AGIMK KAKELTVLCD AKVSLIMFSS TG--KCVDFI
PhAP3           ---------- ---------- -----NGLFK KAQELTVLCD AQISIILISS TN--RLYDYC
MfAP3           ---------- ---------- -----GGIMK KAKELTVLCD AQVSLVMFSS TG--KFSEYC

                61                                                         12
GLO             SP--STTLVD MLDHYHKLS- GKRLWDPKHE HLDNEINRVK KENDSMQIEL RHLKGE---D
SvPI            SP--STTLID MLDQYHKLS- GKKLWDAKHE HLDNEINRIK KENDRMQTEL RHLVGE---D
FBP1            S----TTLVD ILDQYHKLT- GRRLWDAKHE NLDNEINKVK KDNDNMQIEL RHLKGE---D
pMADS2          SP--STTLPD MLDGYQKTS- GRRLWDAKHE NLSNEIDRIK KENDNMQVKL RHLKGE---D
NTGLO           S----TTLVD ILDQYHKLT- GRRLWDAKHE NLDNEINKVK KDNDNMQIEL RHLKGE---D
PI              CP--SMDLGA MLDQYQKLS- GKKLWDAKHE NLSNEIDRIK KENDSLQLEL RHLKGE---D
SLM2            SP--ETAVED ILDQYHKIS- GKRLWDAKHE NLSNEIDRVK KENDNMQIEL RHLKGE---D
DaPI            SP--NASLIK ILDKYQRTS- GRRLWDAKHE YLNSEVERIK KENDNMQIEL RHLKGE---D
CpPI            N---SRPLPQ ILEKYQKSS- GNKLWDAKHE YLSQEIARVE KENQSMRQEL KHLKGE---E
RbPI-1          N---SHPLPE ILHKYQKDT- GNKLWDAKHE YLHQECARIK KENESMQREL GHLKGE---D
RbPI-2          H---SHSLIN SLKRYHKLCP GKRLWDAKHE FLHQEIERIE NENKSMQIEL RHLKGE---D
DePI            SP--STTLIK ILDRYQKAS- GKRLWDAKHE YLSSEVDRVK KENDNMQIEL RHLKGE---D
PnPI-1          S---S-PLIK QLDRYQKAS- GNKLWDAKHE YLSAEVDRVK KENDNMQIEL RHLKGE---D
PnPI-2          S---S-PLVE QLARYQKET- GNKLWDAKHE RLSAEIDRVK KENSNLQIEI RHLKGE---D
PmPI-1          SP--KTTMDA ILTRYQNSK- GSQLWNAKHE YLKQEKERIE KENGRLQLRL RQLKGE---D
PmPI-2          SP--QNKLED VLKKYQDLS- GNKLWDAKHE CLSQEIDRIR KENDAMSIRL RHLEGI---D
PhPI            SP--KTTMDA ILTRYQNST- GTQLWNAKHE SLKQEKERIE KENGRLQLRL RQLKGE---D
LtPI            SP--STTLVK ILDRYQKSS- GKKLWDAKHE HLSNEVERIK KENDSMQIKL RHLKGE---D
OsMADS2         SP--KTSLSR ILEKYQTNS- GKILWDEKHK SLSAEIDRIK KENDNMQIEL RHLKGE---D
OsMADS4         TP--KTTLSR ILEKYQTNS- GKILWDEKHK SLSAEIDRVK KENDNMQIEL QLPYAAAQLD

AP3             SP--NTTTKE IVDLYQTIS- DVDVWATQYE RMQETKRKLL ETNRNLRTQI KQRLGE---C
DEFA            SP--TTATKQ LFDQYQKAV- GVDLWSSHYE KMQEHLKKLN EVNRNLRREI RQRMGE---S
SLM3            SPGSNLTTKD VYDRYQKAL- GVDIWVTHEK RMQDDLQKLN ELNRKLQTDI RQRMGD---C
CpAP3           SP--SISPKA FYDKYRDVT- GDDLWKSQYD KMQQELKTLV ETNRKLRREI GQRVGE---D
PhAP3           SP--STSHKK VYDQYQDGR- KVDLWKKRYE NMKHQLNEQS ERSNKLKKEI RQFMGE---E
MfAP3           SP--STTTKN IFDRYQQAS- GTSLWNSHYE RMQGHLIKLK EENNNLRREI RQRIGE---D
```

(*Continued*)

APPENDIX 1 (Continued)

```
          121                                                                        180
GLO       ITTLNYKE-- LMVLEDALEN GTSA--LKNK QMEFV----R MMRKHNE--- MVE-EENQSL
SvPI      ITTLNYKE-- LMVLEEVLEN GISS--LKAK QMEFV----P AMRKHNE--- MLGGRRTGGL
FBP1      ITSLNHRE-- LMILEDALEN GLTS--IRNK QNEVL----R MMRKKTQ--- SME-EEQDQL
pMADS2    INSLNHKE-- LMVLEEGLTN GLSS--ISAK QSEIL----R MVRKNDQ--- ILE-EEHKQL
NTGLO     ITSLNHRE-- LMMLEDALDN GLTS--IRNK QNDLL----R MMRKKTQ--- SME-EEQDQL
PI        IQSLNLKN-- LMAVEHAIEH GLDK--VRDH QMEIL----I SKRRNEK--- MMA-EEQRQL
SLM2      ITSLPYPD-- LMRLEDALEN GLVG--VREK QMEMY----K LHKKNHK--- MLE-DENNQL
DaPI      LTSLNPKE-- LIPIEAALQN GLTE--VRAK QAEVW----K MMKKNDR--- LLE-EENKHL
CpPI      INSLQPKE-- LIPIEHALEN GITK--VKEK QMEIY----R MMKRNDR--- KLE-EDNKRL
RbPI-1    INSLQPIE-- LIPIEQALDD GIAR--VKER KNEIY----R MMKRNDK--- MLE-EEN-KL
RbPI-2    INSLQPRE-- LIPIERALDN GITK--VRAK IDEIP----R ILEKNGR--- MIE-EEN-KL
DePI      LTSLHPKE-- LISIEDALQN GLVG--VRAK QMEFM----K MMKKNER--- MLE-EENKRL
PnPI-1    LTPLNPRE-- LIPIESALDD GIVG--VKAK IKEHY----R ALKKRTR--- MLE-EDNMRL
PnPI-2    LKPLGPRE-- LYAIENDLED GYAC--VRDK IMEQW----K KLKRNGR--- RLE-AENKHL
PmPI-1    ITSLKPEE-- LIEIENILED GLTN--IRNK SDGLL----E GSNQEYK--- GFGGEEP-AH
PmPI-2    VSSLNHQE-- LEVLEKTLEN GLAS--IRRQ EMEIR----D ARENNIK--- TEE-EEKRRL
PhPI      ITSLKPEE-- LIEIESILDD GLTN--IRNK QD-------- ------K--- GFGRKEP-AL
LtPI      ITSLHPRE-- LLPIEEAFQN GLAC--VRSK QMEYL----K MLKKNER--- TLE-EENKRL
OsMADS2   LNSLQPKE-- LIMIEEALDN GIVN--VNDK LMDHW----E RHVRTDK--- MLE-DEN-KL
OsMADS4   LLLVHPKSQH LVLVLQHLLI PFMHPPVVHH LVPQI----G LAV-VER--- LL--DRD-QL

AP3       LDELDIQE-- LRRLEDEMEN TFKL--VRER KFKSLGNQIE TTKKKNK--S QQDIQKNLIH
DEFA      LNDLGYEQ-- IVNLIEDMDN SLKL--IRER KYKVISNQID TSKKKVR--N VEEIHRNLVL
SLM3      LEDLSFEE-- LCRLGQEMQE AVTL--IRER KYKKIDNQID TTKKKVR--N GQEVHKGLLQ
CpAP3     LSNLSIKE-- LRGLEQDLRD TEKV--VRQR KFGLLSSQGE TQRKKIR--N LAEINGNLWQ
PhAP3     LDGLSFEQ-- LHGLEQKVER ASNI--VRER KEKAISTKVD TLNKKVK--- GYEKQHEGYR
MfAP3     LDDLEIEE-- LRGLEQNLES SIKV--VRER KYHVIQTQTE TYKKKLRSLN GRTGKINFVY

          181                                                                        240
GLO       QFKLRQMH-- -LDPMNDNVM ESQA---VYD ------HHHH QNIA---D-Y EA--------
SvPI      QFKLRQMH-- -LDPTMDEHN VLENG--VY- ---------H QQGVG--D-Y EY--------
FBP1      NCQLRQLE-- -IATMNRNMG EIGE---VF- ---------- QQREN-HD-Y QN--------
pMADS2    QYALHQKE-- -MAAMGGNMR MIEE---VY- ---------H Q-RDR--D-Y EY--------
NTGLO     NWQLRQLE-- -IASMNRNMG EIGE---VF- ---------H Q-REN----- EYQT------
PI        TFQLQQQE-- -MAIASNARG MMMR------ ---------- -------DHD G---------
SLM2      AYMLHKQE-- -MD---GNMR EMEAG--VCS NPS------- -----DRD-Y HY--------
DaPI      NYVLQKQQQ- -MH-MNGNAR DLENE---Y- ---------H QNGR---D-Y --------PS
CpPI      VLKLQHQQQL -MN---GNGR EY-------- ---------- ---------- --------PS
RbPI1     -LVLKLQQ-- -MMN--GNSR EH-------- ---------- ---------- --------PS
RbPI2     -LAFKLEQQQ QQIN--GNVG ENGR--NV-- ---------- -------DH- GY----PPPS
DePI      SYILHNQQIE -MAD--GSNV GE-MENNGYH QHHHHHHHHH QKGRG--DHY GQAAAAHASS
PnPI1     LT-RLSQQQ- -MEIH-GRVE MDTMENTGYN NNN------- QKADGTRD-Y PAHNDNHH--
PnPI2     SVINQHQQQN KWSWKPTHNA F--------- ---------- ---------- ----------
PmPI1     GLYL-HQQ-- -MGLYSGETK GLLNKCAPKS I--------- ---------- ----------
PmPI2     TYIMHEQH-- -MRM--GESL RDLESGSMKR DL-------- ---------- ----------
PhPI      GLHL-EQQ-- -MGVYSGET- -------KYL LNNFAPK--- ---------- --------SS
LtPI-1    SYIL-HQQQL AMD---GNAR EMDH---GY- ---------N Q--RE-RE-Y H---------
OsMADS2   LAFKLHQQDI ALS---GSMR DLEL---GY- ---------H P----DRD-F --AA------
OsMADS4   LGLQGVQILP FH------MP --EL---GY- --------HH DD----RD-F --AA-----S

AP3       -ELELRAEDP HY----GLVD NGG------- ---------- -------D-Y DSVLGYQIEG
DEFA      -EFDARREDP HF----GLVD NEG------- ---------- -------D-Y NSVLGFPNGG
SLM3      -EFEIPKDEP PY----GLVD N-G------- ---------- -------D-Y SNVMGYNDAS
CpAP3     -EYQERMEDE -Y----ALAN --G------- ---------- ---------M STLELGNG--
PhAP3     QKLEMMEVD- -Y----G--- --GT------ ---------- -------Q-Y KGIPMCGPHF
MfAP3     WKVKLRMG-P -T----GLVD NGGP------ ---------- -------D-Y ESALVLANGG
```

(*Continued*)

APPENDIX 1 (Continued)

```
                241                                                                 291
GLO             Q-MP--FAFR VQPMQ----P --NLQ----- --------ER F----------
SvPI            Q-MP--FAFR VQPMQ----P --NLQ----- --------DR F----------
FBP1            H-MP--FAFR VQPMQ----P --NLQ----- --------ER L----------
pMADS2          QQMP--FALR VQPMQ----P --NLH----- --------ER M----------
NTGLO           Q-MP--FAFR VQPMQ----P --NLQ----- --------ER F----------
PI              Q-----FGYR VQPIQ----P --NLQ----- --------EK IMSL--VID--
SLM2            QNPIPPYGFR VQPMQ----P --NLQ----- --------DR M----------
DaPI            H-MP--FTFR AQPMQ----P --NLQ----- --------EN Q----------
CpPI            Q-MPP-FTFQ LHPSQ----P --NLQ----- --------EI K----------
RbPI1           Q-MPMPFTFR VQPAQ----P --NLQ----- --------DN -----------
RbPI2           Q-MP--FTFL VQPAQ----P --NLQ----- --------DN -----------
DePI            QQMPP-FAFR VQPIQ----P --NLH----- --------NN NNNT-NNTNNK
PnPI1           Q-VP--FGFQ VPPMQ----P --NLT----- --------TV TTTT--TTNNK
PnPI2           ---------- ---------- ---------- ---------- -----------
PmPI1           ---P--IAFH VQPLH----P --NLQ----- --------EM K----------
PmPI2           ---P--FAFR VQPIQ----P --NLQ----- --------EQ K----------
PhPI            ---P--IAFH VQPLH----P --NLQ----- --------EM K----------
LtPI-1          QQLP--FTFR LQPIQ----P --NLH----- --------QN Q----------
OsMADS2         Q-MP--ITFR VQPSH----P --NLQ----- --------EN N----------
OsMADS4         --MP--FTFR VQPSH----P --NLQ----- --------QE K----------

AP3             SRA---YALR FHQNHHHYYP NHGLHAPSAS --------DI ITFHLLE----
DEFA            PRI---IALR LPTNHH---P --TLH--SGG GS------DL TTFALLE----
SLM3            RV----LALR LQPCQ----P --NLHAGAGS GS------CV TTYALL-----
CpAP3           -V----FSFR LRPSQ----T --NLHND--E E------YEI HDLRLA-----
PhAP3           ------LGYN MQGN--PY-H ES-AHSDVTT ANNISSAYGI YDLRLA-----
MfAP3           ---------- ---AHI---- ---LHDTG-- -------FGI HDLRLA-----
```

Incomplete sequence of *MfPI* is not shown (see Table 2). Amino acids 25–291 were used in the phylogenetic analysis.

APPENDIX 2

Alignment of all *AP3* lineage members with selected *PI* outgroup representatives

```
          1                                                                 60
DEFA      MARGKIQIKR IENQTNRQVT YSKRRNGLFK KAHELSVLCD AKVSIIMISS TQKLHEYISP
SvAP3     ---------- ---------- -----NGLFK KAHELTVLCD AKVSIIMISS TQKIHEYISP
pMADS1    MARGKIQIKR IENQTNRQVT YSKRRNGLFK KANELTVLCD AKVSIIMISS TGKLHEFISP
NTDEF     MARGKIQIKR IENQTNRQVT YSKRRNGLFK KANELTVLCD AKVSIIMISS TGKLHEFISP
LeAP3     ---------- ---------- -----NGLFK KANELTVLCD AKVSIVMISS TGKLHEFISP
StDEF     MARGKIQIKK IENQTNRQVT YSKRRNGLFK KANELTVLCD AKVSIVMISS TGKLHEFISP
AP3       MARGKIQIKR IENQTNRQVT YSKRRNGLFK KAHELTVLCD ARVSIIMFSS SNKLHEYISP
BobAP3    MARGKIQIKR IENQTTGQVT YSKRRNGLFK KAHELTVLCD ARVSIIMFSS SNKLHEFISP
Boi1AP3   MARGKIQIKR IENQTNRQVT YSKRRNGLFK KAHELTVLCD ARVSIIMFSS SNKLHEFISP
Boi2AP3   MARGKIQIKR IENQTNRQVT YSKRRNGLFK KAHELTVLCD ARVSIIMFSS SNKLHEFISP
RAD1      MARGKIQIKR IENDTNRQVT YSKRRSGLFK KAKELTILCD AKVSIIMISN TNKLHEFISP
RAD2      MTRGQIQIRR IENITNRQVT YSKRRNGLFK KAQELVVLCD AKVSIIMISS RNKLHEFTTP
SLM3      MARRKIQIKK IENLTNRQVT YSKRRNGLFK KANELTVLCD ATVSIIMLSS NLKLHEFLSP
NMH7      --RGKIQIKR IENTTNRQVI YSKRRNGLFK KANELTVLCD AKVSIIMFSS TGKLHEYISP
PD2       ---GKIEIKK IENSTNRQVT YSKRRNGIFK KAKELTVLCD AKISLIMLSS TRKYHEYTSP
TM6       ---GKIEIKK IENSTNRQVT YSKRRNGIFK KRKELTVLCD AKISLIMLSS TRKYHEYTSP
CMB2      MGRGKLEIRK IENKTNRQVT FSKRRNGIMK KAQELTVLCD AKVSLLMISS THKLHHYLSP
PtAP31a   ---------- ---------- -----NGIFK KALELSVLCD AKVSIIMVAT NRKLHEYTSP
PtAP31b   ---------- ---------- -----NGIFK KAKELSVLCD AKVSIIMVAT NRKLHEYTSP
PcAP3     ---------- ---------- -----SGIFK KAKELTILCD AQVCLIMFSN TGKVCEYVSP
PnAP31    ---------- ---------- -----SGIFK KAKELTILCD AQVCLIMFSN TGKVCEYVSP
PnAP32    MGRGKIEIKR IENATNRQVT YSKRRSGILK KAKELTVLCD AEVSLIMFSS TGKMTEYLSP
DeAP3     ---------- ---------- -----AGIMK KARELTVLCD AEVSLIMFSS TGKFSEYISP
CpAP3     ---------- ---------- -----AGIMK KAKELTVLCD AKVSLIMFSS TGKCVDFISP
RbAP3     ---------- ---------- -----AGIIK KAQELTVLCD AEVSLIMVSS SGKCVDFISP
PhAP3     ---------- ---------- -----NGLFK KAQELTVLCD AQISIILISS TNRLYDYCSP
MfAP3     ---------- ---------- -----GGIMK KAKELTVLCD AQVSLVMFSS TGKFSEYCSP

GLO       MGRGKIEIKR IENSSNRQVT YSKRRNGIMK KAKEISVLCD AHVSVIIFAS SGKMHEFCSP
PI        MGRGKIEIKR IENANNRVVT FSKRRNGLVK KAKEITVLCD AKVALIIFAS NGKMIDYCCP
DaPI      ---------- ---------- -----NGILK KAKEITVLCD AEVSLVVFSS TGKMSEFISP
PhPI      ---------- ---------- -----KGIIK KAQEISVLCD THVSVLIFSS AGNMGEFCSP
LtPI-1    ---------- ---------- -----GGILK KAKEITVLCD AQVSLVIFSS TGKISEYCSP
OsMADS2   MGRGKIEIKR IENSTNRQVT YSKRKSGILK KAREISVLCD AEVGVVIFSS AGKLYDYCSP

          61                                                                120
DEFA      T--TATKQLF DQYQKAVGVD LWSSHYEKMQ EHLKKLNEVN RNLRREIRQR M-GESLNDLG
SvAP3     T--SSTKQLF DLYQTTVGVD LWITHYERMQ EHLRKLKDIN KNLRREIRQR M-GESLNDLN
pMADS1    S--ITTKQLF DLYQKTVGVD LWNSHYEKMQ EQLRKLKEVN RNLRKEIRQR M-GESLNDLN
NTDEF     S--VTTKQLF DLYQKTVGVD LWNSHYEKMQ EQLRKLKDVN RNLRREIRQR M-GESLNDLN
LeAP3     S--ITTKQLF DLYQKTIGVD IWTTHYEKMQ EQLRKLKDVN RNLRKEIRQR M-GESLNDLN
StDEF     S--ITTNNLF DLYQKTIGVD IWTSHYEKMQ EQLRKLKDVN RNLRKEIRQR M-GESLNDLN
AP3       N--TTTKEIV DLYQTISDVD VWATQYERMQ ETKRKLLETN RNLRTQIKQR L-GECLDELD
BobAP3    N--TTTKEIL DLYQTVSDVD VWNAHYERMQ ETKRKLLETN RNLRTQIKQR L-GECLDEFD
Boi1AP3   N--TTTKEII DLYQTVSDVD VWSAHYERMQ ETKRKLLETN RKLRTQIKQR L-GECLDELD
Boi2AP3   N--TTTKEIL DLYQTVSDVD VWSAHYERMQ ETKRKLLETN RNLRTQIKQR L-GECLDEFD
RAD1      N--ITTKQVY DAYQTTFSPD LWTSHYAKME QELRNLNEVN RQIRKEIRRR M-GCCLEDMS
RAD2      G--TTTKQIY DMYQQLSGND VWSSQYAMML EELRKIKEAN GNIRKEIRRR M-GFSMEDMS
SLM3      GSNLTTKDVY DRYQKALGVD IWVTHEKRMQ DDLQKLNELN RKLQTDIRQR M-GDCLEDLS
NMH7      S--ASTKQFF DQYQTTVGID LWNSHYENMQ ENLKKLKDVN RNLRKEIRQG M-GECLNDLS
PD2       N--TTTKKMI DQYQSALGVD IWSIHYEKMQ ENLKRLKEIN NKLRREIRQR T-GEDMSGLN
TM6       N--TTTKKMI DQYQSALGVD IWSIHYEKMQ ENLKRLKEIN NKLRREIRQR T-GEDMSGLN
CMB2      G--VSLKKMY DEYQKIEGVD LWRKQWERMQ EQHRKVLELN SLLRREISRR M-GGDLEGLT
PtAP31a   H--TTTKELY DLYQKASGKS LWNSHYERMN DNLNKLKEIN NKLRTEIRQR M-GEDLNELR
PtAP31b   H--TTTKDLY DLYQKASGNS LWNSHYERMK DNLNKLKDIN NKLRTEIRQR M-GEDLNDLR
PcAP3     S--TTMKEFF DRFRRVTNID LWASQYETLQ EELKKQKEIN SRLKKEIRQR TGQDDLNELT
PnAP31    S--TTMKEFF DRFRRITNID LWASQYETLQ EELKTQKEIN NKLKKEIRQR TGQDDLSELS
PnAP32    SLNGNTKRVY DKYQQLSGIS LWNSHYESLQ NALNKQKEIN RRLRREIRQR M-GEDLDELT
DeAP3     S--ATTKRMF DRYQQVSGVN LWNSHYERMQ DNLNKQKEIN NKLRREIRQR M-GEDLNDLS
CpAP3     S--ISPKAFY DKYRDVTGDD LWKSQYDKMQ QELKTLVETN RKLRREIGQR V-GEDLSNLS
RbAP3     T--ISQKEFY DKYQKITKQD LWKSQYDEMQ ERFKHLMETN RKLRREIGQR V-GEDLEGLG
PhAP3     S--TSHKKVY DQYQDGRKVD LWKKRYENMK HQLNEQSERS NKLKKEIRQF M-GEELDGLS
MfAP3     S--TTTKNIF DRYQQASGTS LWNSHYERMQ GHLIKLKEEN NNLRREIRQR I-GEDLDDLE

GLO       S--TTLVDML DHYHKLSGKR LWDPKHEHLD NEINRVKKEN DSMQIELRHL K-GEDITTLN
PI        S--MDLGAML DQYQKLSGKK LWDAKHENLS NEIDRIKKEN DSLQLELRHL K-GEDIQSLN
DaPI      N--ASLIKIL DKYQRTSGRR LWDAKHEYLN SEVERIKKEN DNMQIELRHL K-GEDLTSLN
PhPI      K--TTMDAIL TRYQNSTGTQ LWNAKHESLK QEKERIEKEN GRLQLRLRQL K-GEDITSLK
LtPI-1    S--TTLVKIL DRYQKSSGKK LWDAKHEHLS NEVERIKKEN DSMQIKLRHL K-GEDITSLH
OsMADS2   K--TSLSRIL EKYQTNSGKI LWDEKHKSLS AEIDRIKKEN DNMQIELRHL K-GEDLNSLQ
```

(*Continued*)

APPENDIX 2 (Continued)

```
          121                                                                 180
DEFA      YEQIVNLIED MDNSLKLIRE RKYKVISNQI DTSKKKVRNV EEIHRNLVLE FDARRE-DP-
SvAP3     YDQIVSLIED VDDSLRKIRE RKYKVIGNQI ETSKKKLRNV EEIHRNILLE FDARQE-DP-
pMADS1    YEQLEELMEN VDNSLKLIRE RKYKVIGNQI ETFKKKVRNV EEIHRNLLLE FDARQE-DP-
NTDEF     YEQLEELNEN VDNSLKLIRE RKYKVIGNQI DTYKKKVRNV EEIHRNLLLE FDARQE-DP-
LeAP3     YEQLEELMEN VDNSLKLIRE RKFKVIGNQI ETYRKKVRNV EEINRNLLLE FDARQE-DP-
StDEF     FEQLEELMEN VDNSLKLIRE RKYKVIGNQI ETYRKKVRNV EEIHRNLLLE FDARQE-DP-
AP3       IQELRRLEDE MENTFKLVRE RKFKSLGNQI ETTKKKNKSQ QDIQKNLIHE LELRAE-DP-
BobAP3    IQELCSLEEE MENTFKLVRE RKFKSLGNQI ETTKKKTRAS KTYKKNLIHE LELRAE-DP-
Boi1AP3   IQELRSLEEE MENTFKLVRE RKFKSLGNQI ETTKKKNKSQ QDIQKNLIHE LELRAE-DP-
Boi2AP3   IQELLSLEEE MENTFKLVRE RKFKSLGNQI ETTKKKNKSQ QDIQKNLIHE LELRAE-DP-
RAD1      YQELVFLQQD MENAVTNLSE RKYKVLSNQI ETGKKKLRNV QGIRQNLMQA YDALRE-DP-
RAD2      FRELVILQQD MQDSVAKISE RKYKAIANQI ETTRKKLRNS HGIHRSLVHA FGALN-----
SLM3      FEELCRLGQE MQEAVTLIRE RKYKKIDNQI DTTKKKVRNG QEVHKGLLQE FEIPKD-EP-
NMH7      MEELRLLEDE MDKALKAIRE RKYKVITNQI DTQRKKFNNE REVDNRLLRD LDARAE-DP-
PD2       LQELCHLQEN ITESVAEIRE RKYHVIKNQT DTCRKKVRNL EEQHGTLVLD LEAKCE-DP-
TM6       LQELCHLQEN ITESVAEIRE RKYHVIKNQT DTCKKKARNL EEQNGTLVLD LEAKCE-DP-
CMB2      LVELSALQQE MEEAIIQIRN KKYHTIKNQT GTTRKKIKNL EERHTDLVME LEAKFR-GP-
PtAP31a   LDELRGLEQN MEECLKNIRD RKEHQLRNQI GTSKKKTRNA EEINRKLIRR LD-GMDDDS-
PtAP31b   LEELRGLEQN IQESLMIVGD RKEHQLRNQI GTYKKKSRNA EEINRKLLRR LD-GIDDNS-
PcAP3     FEELRSLEAN LLSSVEIVRL RKFHVLGSHT ETSKKRNKAM EETHKNLLRA TYSQADRDE-
PnAP31    LDEMRILEKN LIDSADIVRN RKNHVLNSHT ETSKKRNKAQ EETYKNLVRA LHSQADREEQ
PnAP32    IEELRSLEQN LEASVKVVRD RKYHVIITQT ETTRKKLRNH TEQNHGLLRE FEPILDEDP-
DeAP3     IEELRGLEQN MDNSLKIVRD RKYHVITTQT ETYRKKLRNL HETHNNLLRE FE-GRDEDT-
CpAP3     IKELRGLEQD LRDTEKVVRQ RKFGLLSSQG ETQRKKIRNL AEINGNLWQE YQ--ERMED-
RbAP3     IHELRSLEQD LRNSAKVVRL RKFGLLSSQG ETQKKKIKNL AGINGSLWQE YQ--ERVEE-
PhAP3     FEQLHGLEQK VERASNIVRE RKEKAISTKV DTLNKKVKGY EKQHEGYRQK LEMME-----
MfAP3     IEELRGLEQN LESSIKVVRE RKYHVIQTQT ETYKKKLRSL NDEQAKLIRV LEGQAENG--

GLO       YKELMVLEDA LENGTSALKN KQMEFVRMM- ---RKHNEMV EEENQSLQFK LRQMHL-DP-
PI        LKNLMAVEHA IEHGLDKVRD HQMEILISK- ---RRNEKMM AEEQRQLTFQ LQQQ------
DaPI      PKELIPIEAA LQNGLTEVRA KQAEVWKMM- ---KKNDRLL EEENKHLNYV LQKQQQ-MH-
PhPI      PEELIEIESI LDDGLTNIRN KQDKGFG--- ---RK----- -EP--ALGLH LEQQMG--V-
LtPI-1    PRELLPIEEA FQNGLACVRS KQMEYLKML- ---KKNERTL EEENKRLSYI LHQQQ--LA-
OsMADS2   PKELIMIEEA LDNGIVNVND KLMDHWERH- ---VRTDKML EDENKLLAFK LHQQD--IA-

          181                                                                 240
DEFA      ---HF-GLVD NE--G----D YNSV------ LGFPNG---G PRIIALRLPT NHH---PT--
SvAP3     ---QY-GLVD NE--G----D YNSV------ LGFPNG---G PRIIALRLPS NHH---PN--
pMADS1    ----Y-GLVE QE--G----D YNSV------ LGFPNG---G HRILALRLQP NHH--QPNHH
NTDEF     ----Y-GLVE QE--G----D YNSV------ LGFPNG---G PRILALRLQP NH---QPNHH
LeAP3     ----YGGLVE HD--G----D YNSV------ LGFPTG---G PRILDLRLQP NNN--YHNH-
StDEF     ----YGGLVE QE--G----D YNSV------ LGFPTG---G HHILALGLQP NNNHHHH---
AP3       ---HY-GLVD NG--G----D YDSV------ LGYQIE---G SRAYALRFHQ NHHHYYPNHG
BobAP3    ---HY-GLVD NG--G----D YDSV------ LGYQL----- ------RFHQ NHHHHYPNHA
Boi1AP3   ---HY-GLVD NG--G----D YDSV------ LGYQIE---G SRAYALRYHQ NHHHHYPNHA
Boi2AP3   ---HY-GLVD NG--G----D YDSV------ LGYQL----- ------RFHQ NHHHHYPNHA
RAD1      ---HC-GLVY NG--G----E YDHV------ M----R---- SHLVGLHF-P REAH-IP---
RAD2      -------L-- N--------E Y--------- ---------- ----GAD--P ------PT--
SLM3      ---PY-GLVD N---G----D YSNV------ MGYNDA---- SRVLALRLQP C----QPN--
NMH7      ---RF-EMMD NG--G----E YESV------ IGFSN---LG PRMFALSLQP TH----PN--
PD2       ---KY-GVVE NE--G---H- YNSA------ VAFANG---V HNLYAFRLQP LH----PN--
TM6       ---KY-GVVE NE--G---H- YHSA------ VAFANG---V HNLYAFRLQP LH----PN--
CMB2      ---QF-AIGE DD--P---RN YEAAAAAA-- V-YGNDVAAA -NLFALSRHP IT--------
PtAP31a   ---QY-GLED D---G---VD DEPA------ IALTNGN--- SHIFAFRLHP N----QPN--
PtAP31b   ---QY-GLED D---G---GD NESA------ IALTNGN--- SHLFAFRLHP N----QPN--
PcAP3     EFLFY-AIPA DRE-GDHNRD YLSSSSSMRT LSI-NGGDCE NSQITFQLQP S----QPN--
PnAP31    EFHFY-AIPA DTE-GDHHRD YLSSSSSMRR LSISGGGDCE NSQITFQLQP S----QPN--
PnAP32    ---HY-VIAH QE------ED YESA------ IELAHGG--- PNIFAFRLQP S----QPN--
DeAP3     ---HY-ALA- NE--G----D YETA------ LEMANGG--- QNIFAFRLQP S----QPN--
CpAP3     ---EY-ALA- N---G----- -MST------ LELGNG---- --VFSFRLQP S----QPN--
RbAP3     ---EY--IA- S---G----- -MSE------ LELGNAV--A PQVFSFRLQP S----QPN--
PhAP3     --------VD YG--G---TQ YKGI------ PMCG------ PHFLGYNMQG N-----PYHE
MfAP3     ---AY-GLVD NG--G---PD YESA------ LVLANGG--- AHI------- ----------

GLO       ---MN-DNVM ES-QAVYDHH HHQN-----I ADYEAQM--- P--FAFRVQP M----QPN--
PI        ---------- E--MAIASNA RGMM-----M RDHDGQ---- ---FGYRVQP I----QPN--
DaPI      ---MN-GNAR D-----LENE YHQN-----G RDYPSHM--- P--FTFRAQP M----QPN--
PhPI      ---YS-G--- -------ETK YLLN-----N FAPKSS---- P--IAFHVQP LH----PN--
LtPI-1    ---MD-GNAR E-----MDHG YNQR-----E REYHQQL--- P--FTFRLQP I----QPN--
OsMADS2   ---LS-GSMR D-----LELG YHP------D RDFAAQM--- P--ITFRVQP SH----PN--
```

(*Continued*)

APPENDIX 2 (Continued)

```
            241                                                  283
DEFA        --LH--SGGG S-----DLTT FALLE----- ---------- ---
SvAP3       --LH--SGGG S-----DLTT FALLE----- ---------- ---
pMADS1      HHLH--SGGG S-----DITT FALLE----- ---------- ---
NTDEF        --LH--SGGG S-----DITT FALA------ ---------- ---
LeAP3       --LH--SGGG S-----DITT FALG------ ---------- ---
StDEF       --LH--SGGG S-----DITT FALG------ ---------- ---
AP3         --LHAPSAS- ------DIIT FHLLE----- ---------- ---
BobAP3      --LHEASAS- ------DIIT FHLLE----- ---------- ---
Boi1AP3     --LHAPSAS- ------DIIT FHLLE----- ---------- ---
Boi2AP3     --LHEASAS- ------DIIT FHLLE----- ---------- ---
RAD1        ------SAGG S-----CLTT YTYLE----- ---------- ---
RAD2        -------AAG S-----YLTT YTYLE----- ---------- ---
SLM3        --LHAGAGSG S-----CVTT YALL------ ---------- ---
NMH7        --PHN--GGA SAAS--DLTT YPLLFSHFSL RIRTTNTTIT FQQ
PD2         --LQNE-GG- ------FGSR DLRLS----- ---------- ---
TM6         --LQNE-GG- ------FGSR DLRLS----- ---------- ---
CMB2        ---------- ---------- ---------- ---------- ---
PtAP31a     --LHIN-GGG ------YGFH NLHLA----- ---------- ---
PtAP31b     --LHIN-GGG ------YGFH NLHLA----- ---------- ---
PcAP3       --LHHAAGGG ------YLYN -QHYV----- ---------- ---
PnAP31      --LHHAAGGG ------YFYS ~QHYA----- ---------- ---
PnAP32      --LHN--GGG ------YNCH DLRLA----- ---------- ---
DeAP3       --LHD--GGG ------YGSH DLRLA----- ---------- ---
CpAP3       --LHND--EE ------YEIH DLRLA----- ---------- ---
RbAP3       --LHDD--EE ------YEIH DLRLV----- ---------- ---
PhAP3       S-AHSDVTTA NNISSAYGIY DLRLA----- ---------- ---
MfAP3       --LHDT-G-- ------FGIH DLRLA----- ---------- ---

GLO         --LQ------ ------ERF- ---------- ---------- ---
PI          --LQ------ ------EKIM SLVID----- ---------- ---
DaPI        --LQ------ ------ENQ- ---------- ---------- ---
PhPI        --LQ------ ------EMK- ---------- ---------- ---
LtPI-1      --LH------ ------QNQ- ---------- ---------- ---
OsMADS2     --LQ------ ------ENN- ---------- ---------- ---
```

Incomplete sequence of *LtAP3* and sequences of unpublished genes *AsAP3* and *SILKY-1* are not shown (see Table 3 for references). Amino acids 25–283 were used in the phylogenetic analysis.