Antiquity and Evolution of the MADS-Box Gene Family Controlling Flower Development in Plants

Jongmin Nam, Claude W. dePamphilis, Hong Ma, and Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University

MADS-box genes in plants control various aspects of development and reproductive processes including flower formation. To obtain some insight into the roles of these genes in morphological evolution, we investigated the origin and diversification of floral MADS-box genes by conducting molecular evolutionary genetics analyses. Our results suggest that the most recent common ancestor of today's floral MADS-box genes evolved roughly 650 MYA, much earlier than the Cambrian explosion. They also suggest that the functional classes T (SVP), B (and Bs), C, F (AGL20 or TM3), A, and G (AGL6) of floral MADS-box genes diverged sequentially in this order from the class E gene lineage. The divergence between the class G and E genes apparently occurred around the time of the angiosperm/gymnosperm split. Furthermore, the ancestors of three classes of genes (class T genes, class B/Bs genes, and the common ancestor of the other classes of genes) might have existed at the time of the Cambrian explosion. We also conducted a phylogenetic analysis of MADS-domain sequences from various species of plants and animals and presented a hypothetical scenario of the evolution of MADS-box genes in plants and animals, taking into account paleontological information. Our study supports the idea that there are two main evolutionary lineages (type I and type II) of MADS-box genes in plants and animals.

Introduction

MADS-box genes encode transcription factors and have been found in three eukaryotic kingdoms, plants, animals, and fungi. In plants, MADS-box genes include developmental regulatory genes comparable to homeobox genes in animals. The protein region encoded by the highly conserved MADS-box is called the MADS-domain and is part of the DNA-binding domain. It is composed of approximately 55 amino acids (aa). It has been proposed that there are at least 2 lineages (type I and type II) of MADS-box genes in plants, animals, and fungi (fig. 1; Alvarez-Buylla et al. 2000b). Most of the well-studied plant genes are type II genes and have three more domains than type I genes: intervening (I) domain (~30 codons), keratin-like coiled-coil (K) domain (~70 codons), and C-terminal (C) domain (variable length). These genes are called the MIKC-type and are specific to plants.

The plant-specific MIKC-type MADS-box genes were first discovered in flowering plants (angiosperms). They can be divided into at least nine classes on the basis of their functions and expression patterns (table 1). In angiosperms, several classes of MADS-box genes control flower formation and are often referred to as floral MADS-box genes. In particular, the "ABC" model of flower formation proposes that the four floral components (organs) are controlled by the interactions of three classes of floral MADS-box genes, A, B, and C (Weigel and Meyerowitz 1994; Ma and dePamphilis 2000). More recently, this ABC model was amended to include an interaction with an additional class of genes, called class E genes (Theissen 2001). According to this amended model, called the "quartet model," the combinatorial tetramers of four classes of floral MADS-domain proteins regulate the development of the four floral components (Honma and

Goto 2001; Theissen 2001): sepals by class A genes, petals by class A, B, and E genes, stamens by class B, C, and E genes, and carpels by class C and E genes (table 1). Class A, C, and E genes are also involved in floral meristem development.

Other classes include the class D genes, which are the close relatives of class C genes and control ovule development (Theissen 2001). The recently proposed class B-sister (Bs) genes also appear to control the development of ovule and seed coat, though their protein sequences are quite different from those of D genes (Becker et al. 2002; Nesi et al. 2002). In addition, another group of MADS-box genes that includes *AGL20* (*AGAMOUS-LIKE 20*) in *Arabidopsis thaliana* (thale cress; hereafter called Arabidopsis) plays a pivotal role in flower activation as an integrator of genetic and environmental flowering pathways (Lee et al. 2000). This group of genes will be called "class F genes" instead of the TM3 or orphan group as previously named (Purugganan 1997; Becker et al. 2000). Several genes such as *AGL6* in Arabidopsis seem to be involved in the development of both flowers and vegetative organs (Alvarez-Buylla et al. 2000a). We call these genes "class G genes." Furthermore, there is a group of genes that trigger flowering as an initiator or a repressor. Loss of function of some of these genes resulted in late flowering or early flowering (Hartmann et al. 2000; Michaels et al. 2003). We call these genes "class T genes."

All the above genes are directly involved in flower formation of angiosperms. We therefore call them "floral MADS-box genes" in this article, though this terminology is usually used for the class A, B, C, and E genes. Note that our classification of MADS-box genes is for simplifying the explanation of our study rather than for proposing new terminologies. There are large numbers of other MADS-box genes in angiosperms. Some of them appear to control flowering time or formation of leaves, fruits, roots, etc. (Zhang and Forde 1998; Michaels and Amasino 1999; Sheldon et al. 1999; Alvarez-Buylla et al. 2000a; Hartmann et al. 2000), but the functions of other genes are unknown.
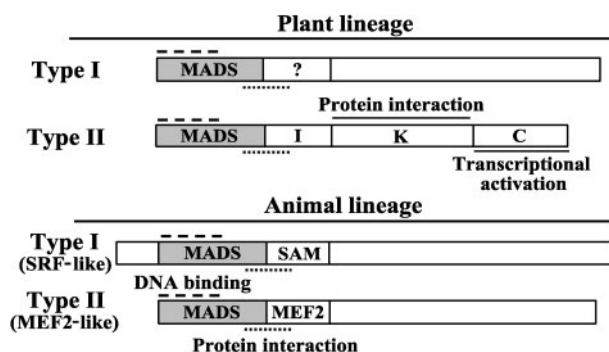
FIG. 1.—Schematic diagram of two types (types I and II) of MADS-box genes in plants and animals. The plant-specific MIKC-type MADS-domain proteins are presented with the name and function of each conserved domain. A broken line indicates the DNA-binding region, and a dotted line the protein-protein interaction region. This figure has been modified from Alvarez-Buylla et al. (2000*b*).

The primary purpose of this article is to investigate the evolutionary relationships and divergence times of floral MADS-box genes. However, because most floral MADS-box genes are known to exist in gymnosperms as well (e.g., Winter et al. 1999; Becker et al. 2000), we consider the genes from both angiosperms and gymnosperms. Previously, Purugganan (1997) studied a similar problem, but this problem should be reexamined because extensive data on MADS-box genes have become available in recent years. Furthermore, to understand the long-term evolution of MADS-box genes, we will also investigate the evolutionary relationships of MADS-domain sequences from plants and animals.

Materials and Methods
Floral MADS-Box Genes Used

At present, MIKC-type MADS-box gene sequences are available from various species of angiosperms, gymnosperms, ferns, club mosses, and mosses (GenBank, TIGR). There are more than 70 MADS-box genes annotated in Arabidopsis (The Arabidopsis Genome Initiative 2000 and our unpublished study). Similarly, we have identified about 70 genes from rice by conducting a TBLASTN search in the Rice Genome Database of China (Yu et al. 2002) and the TIGR Rice Genome Database. From these databases, we compiled 293 full-length MIKC-type MADS-box genes. In the phylogenetic study of floral MADS-box genes, we used 23 reproductive genes, covering all classes of genes shared by angiosperms and gymnosperm species (class B, Bs, C, F, G, and T genes). These genes were chosen from the well-studied eudicot species Arabidopsis, monocot species *Oryza sativa* (rice) and *Zea mays* (maize), and gymnosperm species *Pinus radiata* (Monterey pine), *Picea abies* (Norway spruce), and *Gnetum gnemon* (table 1). We did not include the gymnosperm class E gene (*PrMADS1*) reported from the pine *Pinus radiata*, because this appears to be a contaminated gene from *Eucalyptus grandis* at the time of experimentation (G. Theissen, personal communication). Class A and E genes from angiosperms were also included from our analysis because of their importance

during flower development, though these genes have not been found in gymnosperms. Class D genes were excluded from the analysis, because their protein sequences were close to C gene sequences and the distinction between C and D genes was not always clear.

Protein sequences of these genes were obtained from GenBank or TIGR. The names of the proteins and their GenBank accession numbers or TIGR locus numbers are as follows: AGL9 (At1g24260), AGL6 (At2g45650), AGL20 (At2g45660), APETALA1 (AP1) (At1g69120), APETALA3 (AP3) (At3g54340), PISTILLATA (PI) (At5g20240), AGAMOUS (AG) (At4g18960), SVP (At2g22540), OsMADS3 (S59480), OsMADS4 (T03902), OsMADS8 (AAC49817), OsMADS14 (AAF19047), OsMADS16 (AAD19872), OsMADS17 (AAF21900), OsMADS50 (BAA81886), OsMADS54 (BAA81880), DAL1 (T14846), DAL2 (S51934), DAL3 (T14848), DAL13 (AAF18377), GGM13 (CAB44459), ZMM17 (CAC81053), ABS (At5g23260), and LAMB1 (AAG08991). As is shown in table 1, the protein sequence of a class T gene from *G. gnemon*, GGM12, is available, but it was not used in our analysis because it was a fragmentary sequence. In this article we have used simplified gene notations to make the study understandable for a wide audience.

Phylogenetic Analysis of MIKC-Type Genes

We used protein sequences for our phylogenetic analysis, because the evolutionary pattern of protein sequences appears to be simpler than that of DNA sequences (Nei and Kumar 2000, chapter 2) and protein sequences often give more satisfactory results than DNA sequences in the study of long-term evolution (Hashimoto et al. 1994; Russo, Takezaki, and Nei 1996; Glazko and Nei 2003). In the present case, we could minimize the effect of variation in the GC content at third codon position by using protein sequences.

We aligned 293 protein sequences using the computer program ClustalX (Thompson et al. 1997) with default parameters except the gap opening parameter of 2.0. We then constructed a preliminary Neighbor-Joining (NJ) tree with Poisson-correction (PC) distance using the computer program MEGA2 (version 2.1) (Kumar et al. 2001). (In MEGA2, taxon input orders are randomized for all bootstrap replications.) According to this tree, we divided 293 protein sequences into 18 groups and aligned them separately with the same parameters using ClustalX. These aligned groups were again aligned to each other using the profile alignment option in this program. After elimination of gaps in this alignment, we constructed an initial NJ tree using PC distance. As mentioned above, we selected 24 representative sequences of 142 amino acid sites without gaps, including the MADS-domain, the K-domain, and the conserved region of the I-domain. Using MEGA2, we then constructed NJ trees with *p*-distance (proportion of different amino acids), PC distance, and PC gamma distance (Nei and Kumar 2000, chapter 2). In addition, we constructed maximum-likelihood (ML) trees using the PROTML program with the Poisson and JTT models (Adachi and Hasegawa 1996) and maximum-parsimony

Table 1
Representatives of Different Classes of MADS-Box Genes Considered in This Study

	Gene and Source			
Class	Arabidopsis (Eudicots)	Rice or Maize (Monocots)	Norway Spruce, Monterey Pine, or Gnetum (Gymnosperms)	Function in Arabidopsis
Class A (AP1 or SQUA)	Arabi A (APETALA1 or AP1)	Rice A (OsMADS14)	Unknown	Sepal and petal development, floral meristem development (Weigel and Meyerowitz 1994)
Class B AP3/PI or DEF/GLO	Arabi B-AP3 (AP3) Arabi B-PI (PISTILLATA or PI)	Rice B-AP3 (OsMADS16) Rice B-PI (OsMADS4)	Spruce B (DEFICIENS-AGAMOUS-LIKE 13 or DAL13)	Petal and stamen development (Weigel and Meyerowitz 1994)
Class Bs	Arabi Bs (ABS)	Maize Bs (ZMM17)	Gnetum Bs (GGM13)	Ovule and seed coat development (Nesi et al. 2002)
Class C (AG or PLENA)	Arabi C (AGAMOUS or AG)	Rice C (OsMADS3)	Spruce C (DAL2)	Stamen and carpel development, floral meristem development (Weigel and Meyerowitz 1994)
Class D (not used in this study)	Arabi D (AGL11)	Rice D (OsMADS13)	Unknown	Ovule development (Theissen 2001)
Class E (AGL2/4/9)	Arabi E (AGL9)	Rice E (OsMADS8)	Unknown	Petal, stamen, carpel and floral meristem development (Theissen 2001)
Class F (AGL20 or TM3)	Arabi F (AGL20)	Rice F (OsMADS50)	Spruce F (DAL3)	Flowering activation (integrator of genetic and environmental flowering pathways) (Lee et al. 2000)
Class G (AGL6)	Arabi G (AGL6)	Rice G (OsMADS17)	Spruce G (DAL1)	Expressed in both vegetative and reproductive tissues (Alvarez-Buylla et al 2000a)
Class T (SVP or STMADS11)	Arabi T (SVP)	Rice T (OsMADS54)	Gnetum T (GGM12, partial sequence—not used in this study)	Flowering repression (Hartmann et al. 2000)

NOTE.—We used simplified gene names. Commonly used names and their abbreviations are given in parentheses. The function of each class of genes is based on the studies in Arabidopsis. ''Arabi'' indicates Arabidopsis. All classes of genes are members of floral MADS-box genes.

(MP) trees using the PAUP* program with the stepwise addition and tree-bisection-reconnection (TBR) algorithm with 500 bootstrap resamplings (Swofford 1998). A distantly related MADS-box gene, *LAMB1*, from the club moss *Lycopodium annotinum*, was used as the outgroup in this study. According to our phylogenetic analysis, this gene was closely related to type I genes (see Supplementary Material online at the journal's Web site: http://www.molbiolevol.org). Alvarez-Buylla et al. (2000*b*) have suggested that type I proteins do not have the K-domain (putative coiled-coil structure). However, the LAMB1 protein has a domain similar to the K-domain, including regularly spaced hydrophobic amino acids (e.g., leucine, isoleucine, and valine), which are known to be important for protein-protein interaction (Moon et al. 1999). Therefore, we could align the LAMB1 protein sequence with other MADS-domain protein sequences. Moreover, *LAMB1* has been suggested to be a new MIKC-type MADS-box gene designated as MIKC*-type, whereas the other 23 genes were classical MIKC genes (MIKC^c-type; Henschel et al. 2002). There are two more MIKC*-type genes (*PPM3* and *PPM4*) reported from the moss *Physcomitrella patens* (Henschel et al. 2002). Use of these genes as the outgroups produced essentially the same topology for the floral MADS-box genes.

Once the topology of the phylogenetic tree was determined, we estimated the times of divergence between various types of genes using the linearized tree method (Takezaki, Rzhetsky, and Nei 1995; see program LIN-TREE in http://mep.bio.psu.edu). With the LINTREE method, the time scale constructed does not apply to the outgroup. We also used Yoder and Yang's (2000) likelihood method implemented in the computer program PAML (Yang 2002) with a different evolutionary rate for class B genes of angiosperms from the rate used with the remaining genes. Sanderson's (2003) penalized likelihood method was also used.

Phylogenetic Analysis of MADS-Domains from Plants and Animals

The animal species studied so far seem to have at least one type I gene and one type II MADS-box gene, but the number of the genes is generally very small (Alvarez-Buylla et al. 2000*b*). All of the well-studied plant MADS-box genes are type II genes, and there are many other type II genes in angiosperms and gymnosperms. The existence of plant type I genes has not been well established, except in Arabidopsis, rice, and club moss (Alvarez-Buylla et al. 2000*b* and our unpublished study).

To study the evolutionary relationships of type I and type II MADS-box genes, we used the MADS-domain sequences (~55 aa) of 87 representative genes from plants (Arabidopsis, rice, spruce, pine, gnetum, fern, club moss, and moss) and animals (human, mouse, zebrafish, fruitfly, mosquito, and nematode) (see Supplementary Material

online). In this study we used only MADS-domain sequences, because animal genes do not have the IKC domain. The 87 MADS-domain sequences were aligned by using ClustalX, and the evolutionary relationships of the genes were examined by constructing a NJ tree with *p*-distance for 55 shared amino acids.

Results
Phylogenetic Tree of MIKC-Type Genes

The phylogenetic tree of 24 representative MADS-box genes from eudicots, monocots, and gymnosperms is presented in figure 2. This tree was obtained by the NJ method with PC distance, but very similar trees were obtained by NJ with *p*-distance and PC gamma distance, and by ML and MP methods (see Supplementary Material online). Although the bootstrap values for interior branch a-b, as well as for the B or Bs gene clades of this tree, are very low, the other clades involving class E, G, A, F, and C genes are supported with reasonably high bootstrap values (> 70%). Similar patterns were observed in trees obtained by other tree-building methods. Therefore, the portion of the tree containing the class E, G, A, F, and C genes appears to be reliable.

This tree suggests that after separation of the class T genes from the non-T floral MADS-box genes, class B/Bs genes were the first to diverge from the rest of non-T floral MADS-box genes, although this finding is still provisional. Class C genes then separated from the genes belonging to class F, A, G, and E genes. The next group of genes to diverge was class F genes. Moreover, the taxonomic distribution of functional classes of floral MADS-box genes (table 1) suggests that class E and G genes, which diverged most recently, diverged around the time of angiosperm/gymnosperm split. Several class-specific or taxon-specific amino acids have been reported (e.g., Huang et al. 1995; Kramer, Dorit, and Irish 1998), but we did not find any key features of conserved amino acids supporting any clade of the tree in figure 2. We also compared the positions of introns among all classes of genes, but the positions were too conserved to be informative for inferring the phylogenetic relationships of MADS-box genes (data not shown).

Estimates of Divergence Times

Although molecular estimates of divergence times between genes or species depend on a number of assumptions and are generally very crude (Nei, Xu, and Glazko 2001; Glazko and Nei 2003), they are still useful for obtaining a rough idea of the evolutionary history of genes or species. With this caveat in mind, we estimated the times of divergence between different classes of genes. In the estimation of divergence times, the hypothesis of constant evolutionary rate should first be tested, and then the sequences whose evolutionary rate significantly deviates from constancy should be eliminated (Takezaki, Rzhetsky, and Nei 1995). In this case a number of authors have used Yang's (2002) or Gu and Zhang's (1997) likelihood method for estimating gamma parameter *a*. However, for the purpose of time estimation, these
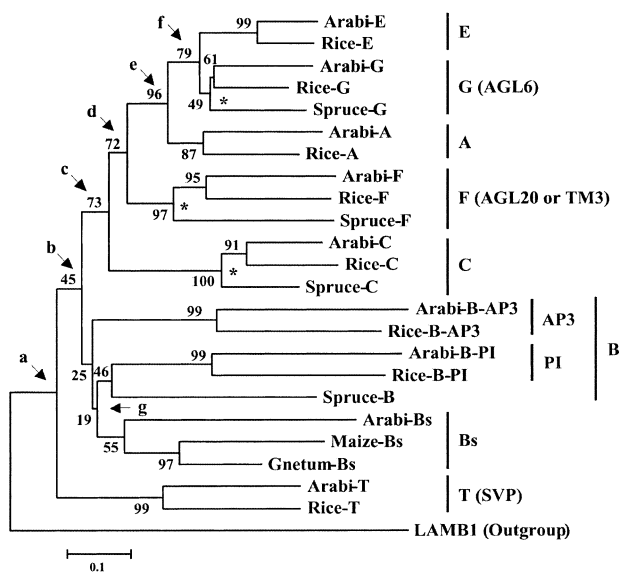

Fig. 2.—Phylogenetic tree of nine classes of MADS-box genes (A, B, Bs, C, D, E, F, G, and T) from monocots, dicots, and gymnosperms with a gene from the club moss *Lycopodium annotinum*, *LAMB1*, used as the outgroup. The number for each interior branch is a percent bootstrap value (500 resamplings). The scale bar indicates the estimated number of amino acid substitutions per site. The number of amino acids used was 142 without gaps per sequence. AP3 and PI are abbreviations of APETALA3 and PISTILLATA, respectively. Gene names were simplified to make the paper understandable to a wide audience (see table 1). Calibration points used for estimating divergence times are marked with an asterisk.

methods, particularly the former method, tend to give underestimates of *a*, and this often leads to overestimation of divergence times when ancient divergence times are estimated (Nei, Xu, and Glazko 2001; Glazko and Nei 2003). This seems to be particularly true for slowly evolving genes such as cytochrome c. Dickerson (1971) showed that in cytochrome c and hemoglobin the number of amino acid substitutions estimated by PC distance ($a = \infty$) is nearly proportional to the time since species divergence up to about 500 MYA. Nei (1987, pp. 47–50) also showed that variation in evolutionary rate among amino acid sites has a relatively small effect on time estimates unless the sequence divergence is very high. We have therefore decided to use primarily PC distance for estimating divergence times. However, we also used Dayhoff's distance to take into account backward and parallel mutations. According to Nei and Kumar (2000, chapter 2), Dayhoff's distance can be computed by a PC gamma distance with $a = 2.25$. We therefore used this method. Note that the use of these distances gives conservative estimates of divergence times compared with those obtained by the PC gamma distance with a likelihood estimate of *a* (see below).

We used the two-cluster test of Takezaki, Rzhetsky, and Nei (1995) to examine the applicability of the molecular clock for the tree in figure 2 and found that the four B genes (2 AP3 genes and 2 PI genes) evolved significantly faster than other genes at the 3% level. We therefore eliminated these four genes and constructed a linearized tree with PC distance for the remaining genes (fig. 3*A*). The two-cluster test also showed that the spruce
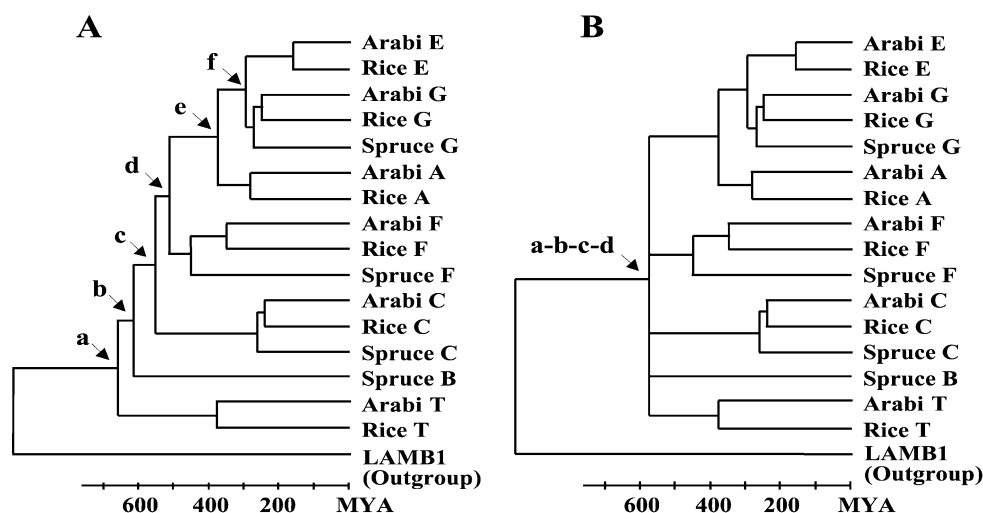
Fig. 3.—Linearized trees used for estimating divergence times. The time scale is based on the results with PC distance. *A*. Topology from figure 2. *B*. Topology when the interior branches between nodes a, b, c, and d are collapsed.

C gene evolved significantly more slowly than the Arabidopsis and rice C genes at the 5% level, but we retained this gene because it was important for calibration of the time scale, and because a relatively small deviation of a sequence from rate constancy does not affect time estimates seriously (Nei and Kumar 2000, pp. 200–202). In addition to the four B genes, we also eliminated all Bs genes because of the uncertain phylogenetic position of the genes (fig. 2). To compare our results with previous estimates of divergence times for floral MADS-box genes by Purugganan (1997), we constructed a linearized tree for a simplified Purugganan tree topology. Purugganan studied the phylogenetic tree of many floral MADS-box genes, but the bootstrap values of the interior branches were so low that he merged several interior nodes. If we use only 24 genes, as in our study, the linearized Purugganan tree becomes as given in part B of figure 3. We therefore estimated the divergence time for the merged node (a-b-c-d).

To calibrate the time scale of the linearized tree, a calibration point is necessary. For our data set, the divergence times between "eudicots" and "monocots" and between "gymnosperms" and "angiosperms" may be used as the calibration point. However, there is no good fossil record for the divergence of eudicots and monocots, and other authors have used various values (131–200 MYA) for this divergence (Wolfe et al. 1989; Laroche, Li, and Bousquet 1995; Soltis et al. 2002). This calibration point also gives some unreasonable time estimates for our data set (see below). By contrast, there seems to be a consensus about the divergence time between angiosperms and gymnosperms, which is about 300 MYA. This estimate is supported by both paleontological data and molecular time estimates (Stewart and Rothwell 1993, pp. 505–512; Savard et al. 1994; Goremykin, Hansmann, and Martin 1997; Soltis et al. 2002). In addition, the angiosperm/gymnosperm split calibration will produce smaller standard errors of time estimates than the monocot/ eudicot split calibration, because the former is a more

ancient evolutionary event than the latter (Glazko and Nei 2003). We have therefore decided to use this time as the calibration point.

Figure 3A shows that each of class G, F, and C genes included one gymnosperm gene and two angiosperm genes. We therefore computed the average PC distance (d) between the gymnosperm and angiosperm genes and obtained $d = 0.372$. This gives an estimate of the rate of amino acid substitution (r) to be $r = d/(2 \times 300)$ per million years or $r = 6.2 \times 10^{-10}$ per year. The timescales for trees A and B in figure 3 were obtained by using this rate of amino acid substitution. The times of divergence between different classes of genes can then be estimated from these linearized trees. The results obtained are presented in table 2, which also includes time estimates obtained by using Dayhoff and PC gamma distances. When PC distance is used, the time of divergence between the T and the non-T floral MADS-box genes is estimated to be about 652 MYA. This is well before the time of the Cambrian explosion (about 545 MYA; see fig. 4). Table 2 also suggests that the divergence between class B genes and other non-T floral MADS-box genes (612 MYA) occurred before the Cambrian explosion. The divergence between class C genes and the remaining non-T floral genes (537 MYA) again appears to have occurred around the Cambrian explosion. This might sound strange, because most animal and plant phyla are believed to have evolved no earlier than the time of the Cambrian explosion. However, recent paleontological data (Xiao, Zhang, and Knoll 1998) suggest that, by this time, green algae had already evolved. The fossil record suggests that the first land plants such as bryophytes appeared around 450 MYA. Our estimates in table 2 suggest that class A, G, and E gene lineages originated after the occurrence of land plants. Table 2 also includes an estimate (556 MYA) of the divergence time between B and Bs genes. In the estimation of this divergence time, the class B genes from angiosperms were excluded because of their faster rate of evolution compared to other genes, and the divergence

Table 2
Estimates of Divergence Times (± SE) of Floral MADS-Box Genes

	Linearized Tree Method			Maximum-Likelihood Method		
Node	PC Distance	Dayhoff Distance ($a = 2.25$)	PC Gamma Distance ($a = 1.06$)	Poisson Model ($a = \infty$)	Dayhoff Model	Poisson + Gamma Model ($a = 1.06$)
(a) T/(others)	652 ± 72	721 ± 91	816 ± 120	816	813	836
(b) B/(C/D-F-A-G-E)	612 ± 62	668 ± 77	743 ± 99	749	775	772
(c) (C/D)/ (F-A-G-E)	537 ± 44	573 ± 54	612 ± 68	630	647	631
(d) F/(A-G-E)	502 ± 42	531 ± 50	566 ± 62	564	569	586
(e) A/(G-E)	374 ± 39	380 ± 45	388 ± 51	428	406	422
(f) G/E	289 ± 29	286 ± 31	282 ± 35	341	327	340
(g) B/Bs	556 ± 65	598 ± 78	646 ± 94	656	662	714
Node a-b-c-d (Purugganan tree)	575 ± 49	623 ± 59	684 ± 75	689	701	706

NOTE.—Unit of time estimates is MYA. The gymnosperm/angiosperm split (ca. 300 MYA) in classes C, F, and G was used for calibrating the time scale. Dayhoff distance was computed by using PC gamma distance with $a = 2.25$ (Nei and Kumar 2000, chapter 2). In the linearized tree method, time estimates for nodes (a) ~ (f) were computed by using 16 genes. (Three Bs genes and 4 angiosperm B genes were excluded.) The time of divergence between the B and Bs genes was estimated separately (see text). Because the ML method does not give proper standard errors (Yoder and Yang 2000), those values are not presented.

time was estimated by dividing the distance between the B and Bs genes by $2r$, where $r = 6.2 \times 10^{-10}$ per year. This estimate suggests that the gymnosperm B and Bs genes diverged a long time ago, if they are clearly definable separate gene groups.

Because many of the above estimates of divergence times far exceed the times of first appearance of land plants in the fossil record (450 MYA), they might be over-estimates. However, if we use Dayhoff distance or PC gamma distance with an ML estimate (1.06) of a obtained by Gu and Zhang's method, the divergence time estimates become even greater (table 2). This was especially so when PC gamma distance was used. In this case branch points a and b were estimated to be 816 and 743 MYA, respectively. We also used Yoder and Yang's method without eliminating B genes but with the assumption that these genes evolved faster than the other genes (two rates model). This method also gave greater estimates than those obtained by PC distance even when the Poisson model ($a = \infty$), Dayhoff model, or Poisson gamma model ($a = 1.06$) was used (table 2). Sanderson's penalized likelihood method gave even greater estimates than other methods (see Supplementary Material online). Therefore, our estimates obtained from the linearized tree method with PC distance are most conservative.

One might wonder whether we used most closely related copies (orthologous genes) of the class G, F, and C genes between angiosperms and gymnosperms for computing the time scale. Actually we tried to do so, but there is no guarantee for the use of real orthologous genes, in part because no complete genome sequence is yet available from any gymnosperm species and in part because it is not easy to determine orthologous genes even in the presence of complete genome sequences (Theissen 2002). However, if we had used nonorthologous genes for any of these gene classes, our estimates would have been lower than unbiased estimates, because the rate of amino acid substitution should have been overestimated. This factor also tends to make our estimates conservative.

As already mentioned, some authors have used the monocot/eudicot divergence (200 MYA) as the calibration point. In our data set, however, the use of this calibration point gave a divergence time estimate of 251 MYA between the angiosperms and the gymnosperms. (The average distance of the angiosperm and gymnosperm genes from class C, F, and G genes was used.) When we used a calibration point of 150 MYA for the monocot/eudicot divergence, we obtained an estimate of divergence of 188 MYA for the angiosperm and gymnosperm split. These estimates are clearly unreasonable, because angiosperms and gymnosperms are believed to have diverged about 300 MYA. We therefore decided not to use the monocot/eudicot calibration point. Incidentally, if we use the angiosperm/gymnosperm divergence (300 MYA) as the calibration point, we obtain an expected divergence time of 239 MYA between monocots and eudicots.

In figure 3B, we have Purugganan's topology. If we estimate the branch point (a-b-c-d) of this topology, we obtain 575 MYA. This is considerably greater than Purugganan's estimate (476 MYA). This difference has occurred in part because Purugganan used the monocot/eudicot divergence (200 MYA) as the calibration point and in part because he used paralogous genes of E genes between monocots and eudicots.

Phylogenetic Tree of 87 MADS-Domains from Plants and Animals

Figure 5 shows a NJ tree of type I and type II MADS-domain sequences from plant and animal species. Type I and type II genes form their own clades, and these clades are quite well supported by the bootstrap test. Type II genes are further divided into plant and animal genes. The monophyletic cluster of animal type II genes is well supported. Plant type II genes also form a monophyletic cluster, although the bootstrap support is rather weak (51%). Animal type I genes form a monophyletic group. In contrast, plant type I genes do not form a monophyletic cluster, although genes from Arabidopsis and rice form a well-supported cluster. This failure of plant type I genes to form a monophyletic cluster could be due to the small number of amino acids used.

Although our results are somewhat ambiguous, they generally support the view put forth by Alvarez-Buylla
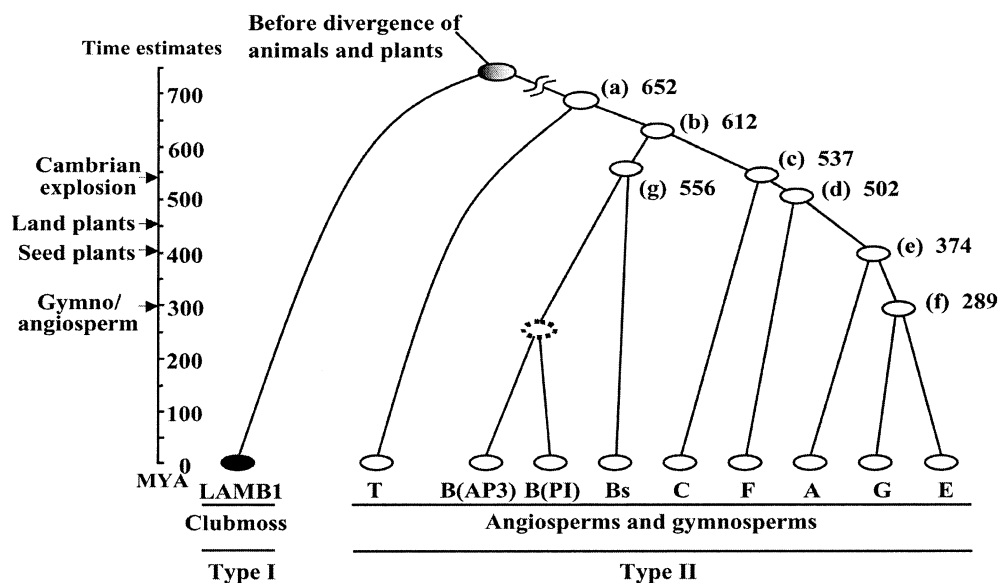
Fig. 4.—Schematic representation of the evolution of floral MADS-box genes. Divergence time estimates (MYA) are indicated for each node of the tree in figure 3*A*. The divergence time for node g was estimated separately (see text). Several important events in plant evolution are indicated to the left of the time scale. The time estimates of these major events are taken from Stewart and Rothwell (1993, pp. 505–512).

et al. (2000*b*), that the type I and type II genes were generated by a gene duplication that occurred before the plant/animal divergence. Animal type I genes control very basic transcription processes concerned with various aspects of cell growth and differentiation and neuronal transmission, etc., whereas type II genes are responsible for muscle development (Shore and Sharrocks 1995). The function of plant type I genes is not well understood, and these genes have only been identified by genomic sequencing of Arabidopsis and rice, although the *LAMB1* gene in the club moss has been suspected to be a type I gene. Many plant type II genes in figure 5 belong to one of the nine classes of MIKC-type MADS-box genes considered in figure 2. However, there are additional MADS-box genes that control various developmental processes such as root formation.

Plant type II genes form many clades of a few genes, and many of these clades are statistically supported relatively well. However, their inter-clade relationships are poorly supported. In particular, B/Bs genes are no longer monophyletic. Nevertheless, the relationships of the genes belonging to floral MADS-box gene classes A, C, E, F, G, and T are virtually the same as those in figure 2. Therefore, the tree in figure 5 may reflect the evolutionary history of MADS-box domains to some extent. The low bootstrap values for these relationships occurred primarily because we used many sequences with only 55 aa, and because there are many other MADS-box genes which are closely related to but are distinct from floral MADS-box genes in plant genomes. It is possible that the nine classes of floral MADS-box genes were derived from some of these distinct MADS-box genes nearly independently. In the present case it is not meaningful to try to estimate the divergence times of these genes, because the number of amino acids per sequence is very small.

Discussion
Reliability of Estimates of Divergence Times

The fact that nonflowering gymnosperms have most classes (B, Bs, C, G, and T) of floral MADS-box genes indicates that the gene duplications that generated these genes occurred long before their angiosperm-specific functions were established. It is not clear what kinds of function these floral MADS-box genes had before their functional diversification, but they were probably involved in the regulation of broad developmental and reproductive processes, as was suggested by Becker et al. (2000). This evolutionary pattern is similar to that of homeobox genes that control segmentation of animal body structure (Zhang and Nei 1996; Purugganan 1998). Cnidarian species such as jellyfish do not have a segmented body structure, yet they have hox genes (Ferrier and Holland 2001). Actually, similar evolutionary patterns are observed with several other gene families controlling development (e.g., Burglin 1997; Meyerowitz 2002), and it appears that the occurrence of gene duplication before functional diversification is a generalized phenomenon with gene families controlling development.

Our conservative estimates suggest that class A and B floral genes diverged about 612 MYA, which is two times earlier than the paleontological estimates of divergence time between gymnosperms and angiosperms. It also far exceeds the paleontological estimate of the time of first land plants (mosses) (ca. 450 MYA). However, mosses are known to have at least two genes that are homologous to classical MIKC-type genes (Henschel et al. 2002). It should also be noted that classical MIKC-type genes have been identified even in green algae such as *Chara*, *Coleochaete*, and *Closterium* (M. Hasebe, personal communication), all of which evolved earlier than land plants. Note that the oldest fossil record of green algae is 700–750 Myr old (Chen and Xiao 1991; Butterfield 2000),
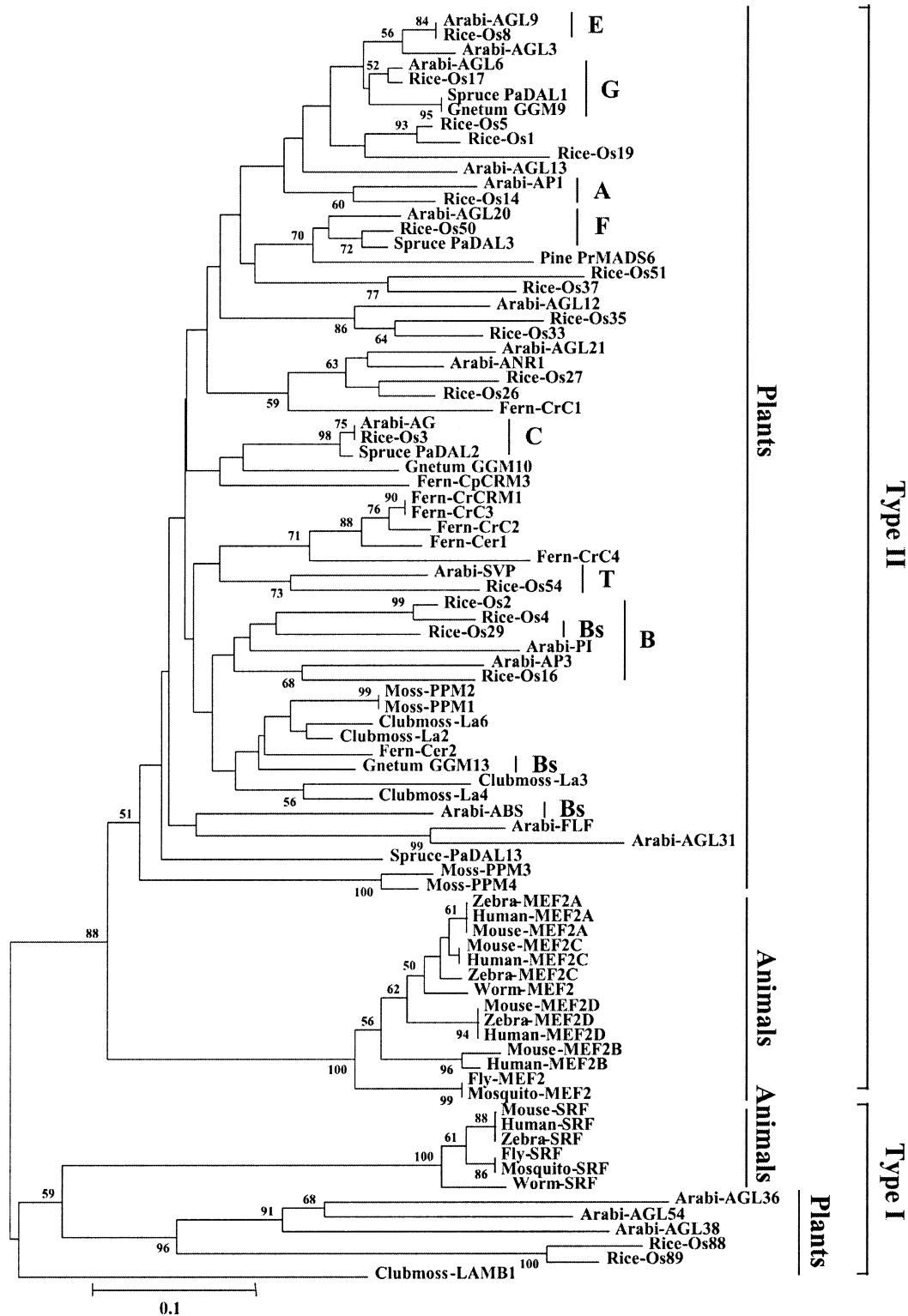
FIG. 5.—Phylogenetic tree of 87 MADS-domain sequences from Arabidopsis, rice, gymnosperms, ferns, club mosses, mosses, and animals. This tree was constructed by the NJ method with *p*-distance for a 55-aa domain. The number for each interior branch is the percent bootstrap value (500 resamplings), and only values greater than 50% are shown. The names of plant species used are the same as those in figure 2, except for ferns and mosses. Those of the remaining species are as follows: fern, *Ceratopteris richardii*; moss, *Physcomitrella patens*; human, *Homo sapiens*; mouse, *Mus musculus*; zebrafish, *Danio rerio*; nematode, *Caenorhabditis elegans*; mosquito, *Anopheles gambiae*; fly, *Drosophila melanogaster*.

although green algae do not appear to be monophyletic. These observations suggest that our estimate of the time of origin of floral MADS-box genes may not be too early.

In this discussion we have used the most conservative estimates of divergence times obtained by PC distance. If we use PC gamma distance or Yoder and Yang's method, estimates of the time of origin of floral MADS-box genes become greater than 800 MYA. These estimates appear to be too early if we consider the fossil record of land plants and green algae, but we cannot rule out this possibility because the fossil record is notoriously incomplete. It is worth noting that, until recently, all or most orders of placental mammals were believed to have diverged only about 65 MYA. At present, however, we know of the fossil remains of a placental mammal that is about 125 Myr old (Ji et al. 2002). The notion of the Cambrian explosion, in which most visible eukaryotic organisms are believed to have been absent before 545 MYA, is also slowly changing. We now know 570 Myr-old fossils of animal eggs (Xiao, Zhang, and Knoll 1998), 900–1,200 Myr-old fossils of red algae (Butterfield 2000), and 1,100–1,200 Myr-old trace fossils of worm (Seilacher, Bose, and Pfluger 1998; Rasmussen et al. 2002), although the authenticity of these trace fossils has been questioned (Conway Morris 2002).

Nevertheless, it is not clear what kind of function the MIKC-type genes had in ancestral non-seed plants. In recent years an intensive study has been made to identify genes orthologous to floral MADS-box genes in non-seed plants, but that study has not been very successful (e.g., Münster et al. 1997; Hasebe et al. 1998; Hohe et al. 2002; Svensson and Engström 2002). What are the possible reasons for these negative results? There seem to be at least five: First, the orthologs of floral MADS-box genes in non-seed plants so far studied might have been lost in the course of evolution. Second, the orthologs of floral MADS-box genes in non-seed plants are so different from the floral MADS-box gene that it is difficult to identify orthologs now. Third, our molecular time estimates are too old, even though we used the most conservative method. This may happen if the rate of amino acid substitution was faster in the early stage of evolution of floral MADS-box genes than in the later stage. Fourth, the current fossil record is incomplete and land plants might have evolved earlier than currently believed. Fifth, the genes so far studied may be incomplete, and a complete genome search may find the genes. At present, however, it is difficult to resolve the discrepancy between the theoretical and experimental studies.

Long-term Evolution of MADS-Box Genes

As mentioned, MADS-box genes are highly conserved, and the MADS-domain sequences are shared by plants, animals, and fungi, indicating that MADS-box genes have an ancient history. Therefore, studying the history of MADS-box genes, we should be able to obtain some insight into the evolution of morphological characters in eukaryotes. Unfortunately, our knowledge about the MADS-box genes and their function in early eukaryotes is quite limited. Nevertheless, it would be interesting to speculate about the evolution of MADS-box genes in eukaryotes, taking into account both paleontological information and molecular dating. Having a plausible scenario may give some useful information for future experimental studies. Here we consider only the evolution of plant and animal genes, because MADS-box genes in fungi other than the budding yeast are not well studied.

We can see from figure 5 that both plants and animals have two different types of MADS-box genes, type I and type II genes. As indicated by Alvarez-Buylla et al. (2000b), this suggests that these two types of genes diverged by a gene duplication that occurred before the plant/animal divergence (fig. 6). The oldest geological evidence of eukaryotes is given by a lipid biomarker, which has been dated 2,700 MYA (Brocks et al. 1999). There are also eukaryotic fossils that have been dated 2,100 MYA (Han and Runnegar 1992). There is no fossil record that indicates the time of divergence between plants and animals, but molecular data suggest that the divergence time is about 1,400 MYA (Feng, Cho, and Doolittle 1997; Wang, Kumar, and Hedges 1999; Nei, Xu, and Glazko 2001). If these estimates are reliable, the gene duplication must have occurred some time between 1,400 MYA and 2,700 MYA (fig. 6). Because yeast, *Caenorhabditis elegans*, and *Drosophila melanogaster* all have a small number of type I and type II genes (two type I genes and two type II genes in yeast; one type I gene and one type II gene in *C. elegans* and *D. melanogaster*), it is likely that the early plants (possibly red and brown algae, Cavalier-Smith 2002; note that the monophyly of plants and these algae is still controversial) also had a small number of type I and type II genes. This hypothesis may be tested by examining the genomes of extant red and brown algae. Because these early plants have quite complex morphological characters and life cycles, this would help us to understand the ancient function of MADS-box genes during plant evolution. According to the conservative estimates of divergence times of MADS-box genes we present in table 2, a group of green algae which are believed to have evolved 700–750 MYA (fig. 6) is expected to have at most one gene that is ancestral to all the floral MADS-box genes currently present in angiosperms and gymnosperms. However, if our estimates from gamma distance are correct, green algae may have three genes that are ancestral to the current T, B (and Bs), and E (or A, C, F, G) classes of genes.

Figure 6 shows several evolutionary events in both animal and plant lineages. Molecular estimates of divergence times of early metazoan animals are almost always considerably earlier than paleontological estimates. For example, molecular data have suggested that the nematode lineage diverged from the vertebrate lineage 800–1,100 MYA (e.g., Feng, Cho, and Doolittle 1997; Wang, Kumar, and Hedges 1999; Nei, Xu, and Glazko 2001), which is about two times earlier than the times of the Cambrian explosion. The nematode *C. elegans* is known to have one type I gene and one type II MADS-box gene (Alvarez-Buylla et al. 2000b; our unpublished data). The type I and type II MADS-box genes in animals have not been studied very well, but the zebrafish has several type I and type II genes (our unpublished results). These findings
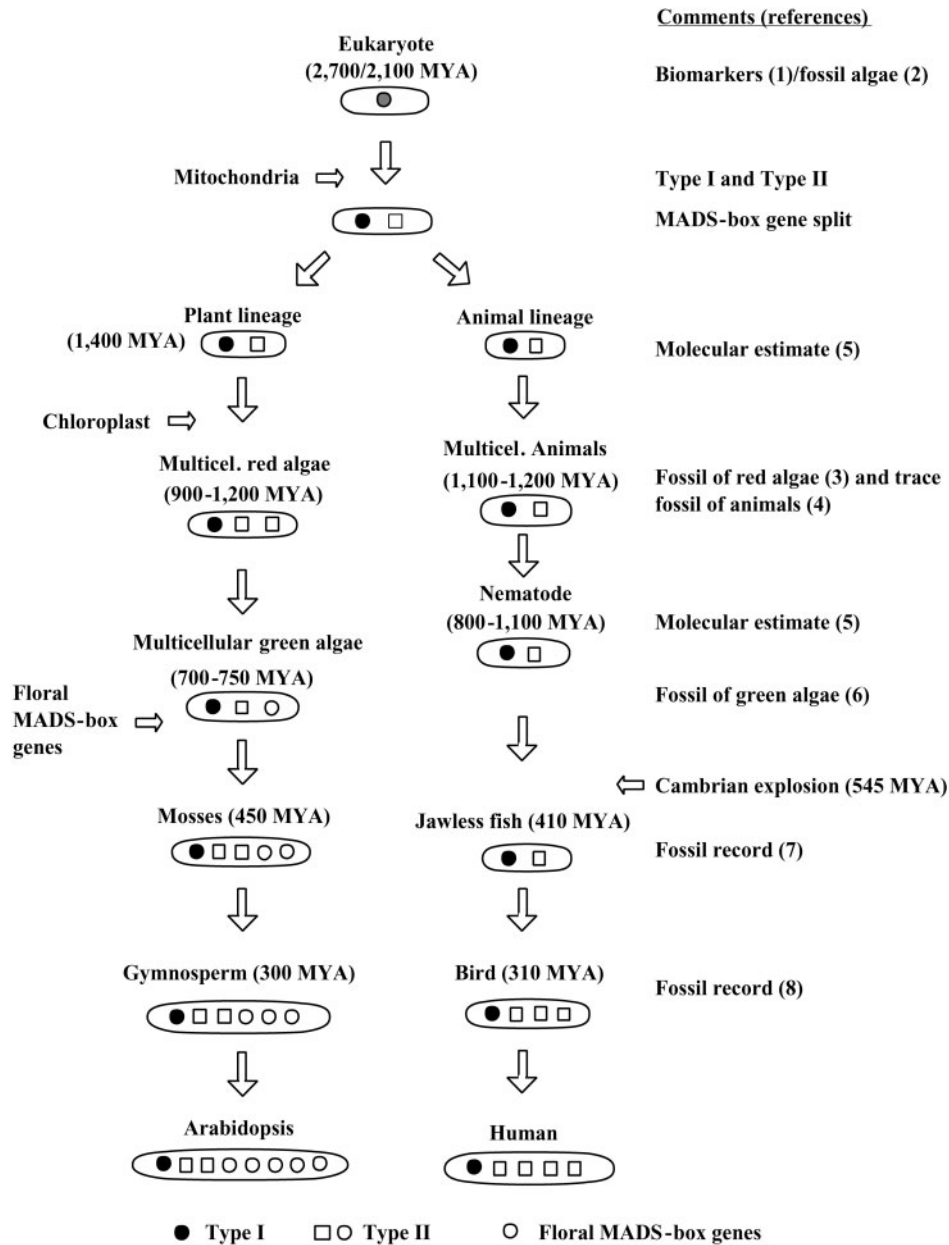
Fig. 6.—A scenario of the evolution of MADS-box genes in plant and animal lineages. Important events of plant and animal evolution (divergence from the lineage leading to Arabidopsis or human) are presented with their estimated times. The references for these estimates are as follows: (1) time of the oldest biomarkers of eukaryotes (Brocks et al. 1999), (2) oldest fossil record of eukaryotic algae (Han and Runnegar 1992), (3) fossil record of some forms of red algae (Butterfield 2000), (4) trace fossil of animals (Seilacher, Bose, and Pfluger 1998; Rasmussen et al. 2002), (5) molecular time estimates of the animal/plant split and nematode evolution (Feng, Cho, and Doolittle 1997; Wang, Kumar, and Hedges 1999; Nei, Xu, and Glazko 2001), (6) fossil record of green algae (Chen and Xiao 1991), (7) fossil record of jawless fish (Maisey 1996, pp. 52–55), and (8) fossil record of the bird/mammal split (Benton 1993, pp. 717–771). The number of circles and squares does not represent the real gene number in each organism. The estimated numbers of MADS-box genes in the species of available genome sequences are as follows: Arabidopsis (>70 genes), rice (>70 genes), human (5 genes), fly (2 genes), nematode (2 genes), and budding yeast (4 genes).

suggest that MADS-box genes are very ancient and evolved gradually in the long history of plants and animals.

Previously we indicated that the MADS-box gene family is an important gene family comparable to the animal homeobox gene family. In this regard, it is interesting to note that the homeobox gene family also exists in plants, animals, and fungi (Burglin 1997; Kappen 2000), and that there are at least two lineages of homeobox genes that diverged before the plant/animal/fungal split. It

would be interesting to investigate how these two different multigene families controlling development coevolved.

Gene Family Expansion or Birth-and-Death Evolution?

Figure 2 shows a pattern of functional diversification of major groups of MADS-box genes. This figure suggests that the number of genes of this multigene family has steadily increased as the reproductive system became more

complex. However, although the gene number must have increased from the time of early plants, this tree does not give the entire picture of evolution of MADS-box genes, because we did not include many genes that are not directly related to flower formation. Our tree in figure 5 is not very reliable, but if it represents a general pattern of evolution of MADS-box genes, it is possible that different floral MADS-box genes were derived from other floral MADS-box genes, which have already been lost, or even from other reproductive MADS-box genes. Furthermore, the Arabidopsis genome is known to contain several MADS-box pseudogenes or truncated genes (our unpublished data), indicating that some MADS-box genes died out in the evolutionary process. These observations suggest that the MADS-box gene family might have been subjected to the birth-and-death model of evolution, in which some genes generate duplicate genes with new functions but others become nonfunctional or are deleted from the genome (Nei, Gu, and Sitnikova 1997). If this is the case, it is possible that the genome of gymnosperms or ferns contains nearly as many MADS-box genes as the angiosperm genomes and that the genes in these plants merely exert the different functions required for the different forms of reproduction. Of course, it is also possible that the phylogenetic tree of current angiosperm genes in figure 2 in large part reflects the history of the increase of member genes of the MADS-box gene family in gymnosperms and angiosperms. At present, we cannot distinguish between the two alternative hypotheses, but this could be done rather easily if the genomic sequences of gymnosperms and ferns were determined. It is also important to note that the two hypotheses are not mutually exclusive and we are interested only in the relative importance of the two possibilities.

Acknowledgments

Literature Cited

Adachi, J., and M. Hasegawa. 1996. MOLPHY, a computer program package for molecular phylogenetics. Version 2.3. The Institute of Statistical Mathematics, Tokyo.

Alvarez-Buylla, E. R., S. J. Liljegren, S. Pelaz, S. E. Gold, C. Burgeff, G. S. Ditta, F. Vergara-Silva, and M. F. Yanofsky. 2000a. MADS gene evolution beyond flowers, expression in pollen, endosperm, guard cells, roots, and trichomes. Plant J. 24:457–466.

Alvarez-Buylla, E. R., S. Pelaz, S. J. Liljegren, S. E. Gold, C. Burgeff, G. S. Ditta, L. Ribas de Pouplana, L. Martinez-Castilla, and M. F. Yanofsky. 2000b. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA 97:5328–5333.

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. Mol. Genet. Genomics 266:942–950.

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

Benton, M. J. 1993. The fossil records 2. Chapman and Hall, New York.

Brocks, J. J., G. A. Logan, R. Buick, and R. E. Summons. 1999. Archean molecular fossils and the early rise of eukaryotes. Science 285:1033–1036.

Burglin, T. R. 1997. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. 25:4173–4180.

Butterfield, N. J. 2000. Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. Paleobiology 26:386–404.

Cavalier-Smith, T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int. J. Syst. Evol. Microbiol. 52:297–354.

Chen, M., and Z. Xiao. 1991. Discovery of the macrofossils in the Upper Sinain Doushantuo Formation at Miaohe, eastern Yangtze Gorges. Sci. Geol. Sinica 4:317–324.

Conway Morris, S. 2002. Ancient animals or something else entirely? Science 298:57–58.

Dickerson, R. E. 1971. The structures of cytochrome c and the rates of molecular evolution. J. Mol. Evol. 1:26–45.

Feng, D. F., G. Cho, and R. F. Doolittle. 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA 94:13028–13033.

Ferrier, D. E., and P. W. Holland. 2001. Ancient origin of the Hox gene cluster. Nat. Rev. Genet. 2:33–38.

Glazko, G. V., and M. Nei. 2003. Estimation of divergence times for major lineages of primate species. Mol. Biol. Evol. 20:424–434.

Goremykin, V. V., S. Hansmann, and W. F. Martin. 1997. Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised molecular estimates of two seed plant divergence times. Plant Syst. Evol. 206:337–351.

Gu, X., and J. Zhang. 1997. A simple method for estimating the parameter of substitution rate variation among sites. Mol. Biol. Evol. 14:1106–1113.

Han, T. M., and B. Runnegar. 1992. Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee Iron Formation, Michigan. Science 257:232–235.

Hartmann, U., S. Hohmann, K. Nettesheim, E. Wisman, H. Saedler, and P. Huijser. 2000. Molecular cloning of SVP, a negative regulator of the floral transition in Arabidopsis. Plant J. 21:351–360.

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

Hashimoto, T., Y. Nakamura, F. Nakamura, T. Shirakura, J. Adachi, N. Goto, K. Okamoto, and M. Hasegawa. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, Giardia lamblia. Mol. Biol. Evol. 11:65–71.

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

Hohe, A., S. A. Rensing, M. Mildner, and R. Reski. 2002. Day length and temperature strongly influence sexual reproduction and expression of a novel MADS-box gene in the moss Physcomitrella patens. Plant Biol. 4:595–602.

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

Huang, H., M. Tudor, C. A. Weiss, Y. Hu, and H. Ma. 1995. The Arabidopsis MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol. Biol. **28**:549–567.

Ji, Q., Z. X. Luo, C. X. Yuan, J. R. Wible, J. P. Zhang, and J. A. Georgi. 2002. The earliest known eutherian mammal. Nature **416**:816–822.

Kappen, C. 2000. Analysis of a complete homeobox gene repertoire: implications for the evolution of diversity. Proc. Natl. Acad. Sci. USA **97**:4481–4486.

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.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

Laroche, J., P. Li, and J. Bousquet. 1995. Mitochondrial DNA and monocot-dicot divergence time. Mol. Biol. Evol. **12**:1151–1156.

Lee, H., S. S. Suh, E. Park, E. Cho, J. H. Ahn, S. G. Kim, J. S. Lee, Y. M. Kwon, and I. Lee. 2000. The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. Genes Dev. **14**:2366–2376.

Ma, H., and C. dePamphilis. 2000. The ABCs of floral evolution. Cell **101**:5–8.

Maisey, J. G. 1996. Discovering fossil fishes. Henry Holt and Co., New York.

Meyerowitz, E. M. 2002. Plants compared to animals: the broadest comparative study of development. Science **295**:1482–1485.

Michaels, S. D., and R. M. Amasino. 1999. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell **11**:949–956.

Michaels, S. D., G. Ditta, C. Gustafson-Brown, S. Pelaz, M. F. Yanofsky, and R. M. Amasino. 2003. *AGL24* acts as a promoter of flowering in Arabidopsis and is positively regulated by vernalization. Plant J. **33**:867–874.

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

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

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. USA **94**:7799–7806.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Nei, M., P. Xu, and G. Glazko. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc. Natl. Acad. Sci. USA **98**:2497–2502.

Nesi, N., I. Debeaujon, C. Jond, A. J. Stewart, G. I. Jenkins, M. Caboche, and L. Lepiniec. 2002. The *TRANSPARENT TESTA16* locus encodes the Arabidopsis Bsister MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell **14**:2463–2479.

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.

———. 1998. The molecular evolution of development. Bioessays **20**:700–711.

Rasmussen, B., S. Bengtson, I. R. Fletcher, and N. J. McNaughton. 2002. Discoidal impressions and trace-like fossils more than 1200 million years old. Science **296**:1112–1115.

Russo, C. A., N. Takezaki, and M. Nei. 1996. Efficiencies of different genes and different tree-building methods in re-covering a known vertebrate phylogeny. Mol. Biol. Evol. **13**:525–536.

Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19**:301–302.

Savard, L., P. Li, S. H. Strauss, M. W. Chase, M. Michaud, and J. Bousquet. 1994. Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extant seed plants. Proc. Natl. Acad. Sci. USA **91**:5163–5167.

Seilacher, A., P. K. Bose, and F. Pfluger. 1998. Triploblastic animals more than 1 billion years ago: trace fossil evidence from India. Science **282**:80–83.

Sheldon, C. C., P. P. Perez, J. Metzger, J. A. Edwards, W. J. Peacock, and E. S. Dennis. 1999. The *FLF* MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell **11**:445–458.

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

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. Proc. Natl. Acad. Sci. USA **99**:4430–4435.

Stewart, W. N., and G. W. Rothwell. 1993. Paleobotany and the evolution of plants. Cambridge University Press, New York.

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

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12**:823–833.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**:796–815.

Theissen, G. 2001. Development of floral organ identity, stories from the MADS house. Curr. Opin. Plant Biol. **4**:75–85.

———. 2002. Secret life of genes. Nature **415**:741.

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 Res. **24**:4876–4882.

Wang, D. Y., S. Kumar, and S. B. Hedges. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals, and fungi. Proc. R. Soc. Lond. Ser. B. **266**:163–171.

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

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

Wolfe, K. H., M. Gouy, Y. W. Yang, P. M. Sharp, and W. H. Li. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86**:6201–6205.

Xiao, S., Y. Zhang, and A. H. Knoll. 1998. Three-dimensional preservation of algae and animal embryos in a *Neoproterozoic phosphorite*. Nature **391**:553–558.

Yang, Z. 2002. Phylogenetic analysis by maximum likelihood (PAML). Version 3.13. University College London, London.

Yoder, A. D., and Z. Yang. 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. **17**:1081–1090.

Yu, J., S. Hu, J. Wang et al. (100 co-authors). 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296**:79–92.

Zhang, H., and B. G. Forde. 1998. An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture. Science **279**:407–409.

Zhang, J., and M. Nei. 1996. Evolution of Antennapedia-class homeobox genes. Genetics **142**:295–303.