Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms

Jongmin Nam[†‡], Joonyul Kim[§¶], Shinyoung Lee[§], Gynheung An[§], Hong Ma[†], and Masatoshi Nei[†]

[†]Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park, PA 16802; and [§]National Research Laboratory of Plant Functional Genomics, Division of Molecular and Life Sciences, Pohang University of Science and Technology, Pohang 790-784, Korea

Contributed by Masatoshi Nei, December 22, 2003

Plant MADS-box genes form a large gene family for transcription factors and are involved in various aspects of developmental processes, including flower development. They are known to be subject to birth-and-death evolution, but the detailed features of this mode of evolution remain unclear. To have a deeper insight into the evolutionary pattern of this gene family, we enumerated all available functional and nonfunctional (pseudogene) MADS-box genes from the *Arabidopsis* and rice genomes. Plant MADS-box genes can be classified into types I and II genes on the basis of phylogenetic analysis. Conducting extensive homology search and phylogenetic analysis, we found 64 presumed functional and 37 nonfunctional type I genes and 43 presumed functional and 4 nonfunctional type II genes in *Arabidopsis*. We also found 24 presumed functional and 6 nonfunctional type I genes and 47 presumed functional and 1 nonfunctional type II genes in rice. Our phylogenetic analysis indicated there were at least about four to eight type I genes and ≈15–20 type II genes in the most recent common ancestor of *Arabidopsis* and rice. It has also been suggested that type I genes have experienced a higher rate of birth-and-death evolution than type II genes in angiosperms. Furthermore, the higher rate of birth-and-death evolution in type I genes appeared partly due to a higher frequency of segmental gene duplication and weaker purifying selection in type I than in type II genes.

Morphological/physiological evolution of organisms has been driven mainly by the evolution of genetic toolkits for developmental/physiological processes such as transcription factors and signaling pathways (1). A large proportion of genetic toolkits are highly conserved even between distantly related organisms. In flowering plants (angiosperms), MADS-box genes are among such toolkits that control various aspects of developmental processes. MADS-box genes are defined by the highly conserved 180-bp-long motif called the MADS-box and are found in animals, fungi, and plants (2). The protein region encoded by the MADS-box is called the MADS-domain (or M-domain) and is part of the DNA-binding domain. It has been proposed that there are at least two evolutionary lineages (types I and II) of MADS-box genes in animals, fungi, and plants (3) (Fig. 1).

There are ≈100 MADS-box genes in *Arabidopsis thaliana* (hereafter called *Arabidopsis*) and >70 MADS-box genes in *Oryza sativa* (hereafter called rice). There are ≈40 clearly identifiable type II MADS-box genes in each of *Arabidopsis* (4, 5) and rice (6). Most of the plant type II genes contain three additional plant-specific domains: intervening (I) domain (≈30 codons), keratin-like coiled-coil (K) domain (≈70 codons), and C-terminal (C) domain (variable length) (7) (Fig. 1). These genes are called MIKC-type genes. The MIKC-type genes can further be divided into two types based on the intron–exon structure: MIKC^c- and MIKC*-type genes (8). The MIKC^c-type genes have been identified in most major evolutionary lineages of green plants such as angiosperms, gymnosperms, ferns, and mosses (9). The MIKC*-type genes were originally found in mosses and clubmosses (8, 10), but these genes are also present in *Arabidopsis* (5). By contrast, the type I MADS-box genes in plants do not encode the K-domain and are sometimes called M-type genes (5).

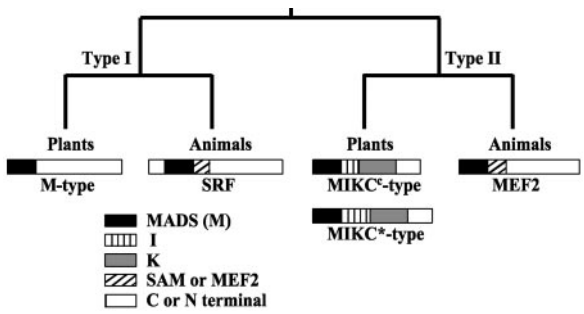

Fig. 1. Domain structures of types I and II MADS-box genes in plants and animals. Adapted from ref. 3 on the structures of types I and II genes and from ref. 8 on the structures of MIKC^c-type and MIKC*-type genes.

It has been shown that at least 11 classes of MADS-box genes are shared between *Arabidopsis* and rice/maize (9, 11). All of them are MIKC^c-type genes, and their expression patterns have been studied intensively in eudicots. Several classes of MIKC^c-type genes, called floral MADS-box genes, are concerned with the development of floral components (organs) such as petals, sepals, stamens, and carpels, as well as regulation of flowering time (12, 13). Other classes of MIKC^c-type genes play diverse roles during vegetative growth (14–16) and fruit development (17). Some of the floral MADS-box genes in monocots have functions equivalent to those of their orthologs in eudicots (18, 19), suggesting an ancient origin of the machinery of flower development. There are also a few other genes that are not shared (lineage-specific) between *Arabidopsis* and rice (9). The functions of MIKC*-type and M-type (or type I) genes are poorly understood.

MADS-box genes are important regulators of development of angiosperms (and probably nonflowering plants as well), and therefore the study of evolution of MADS-box genes is expected to give important clues for understanding the morphological evolution of plants. In our previous paper (20), we indicated that the MADS-box gene family has been subject to the model of birth-and-death evolution, in which new genes are generated by gene duplication, and some duplicate genes stay in the genome as differentiated genes, whereas others are inactivated into pseudogenes or deleted from the genome (21, 22). Although it

Abbreviations: M-domain, MADS-domain; MRCA, most recent common ancestor.

[‡]To whom correspondence should be addressed. E-mail: jyn101@psu.edu.

[¶]Present address: Michigan State University–Department of Energy Plant Research Laboratory and Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824.

www.pnas.org/cgi/doi/10.1073/pnas.0308430100

has been discussed that types I and II genes might have experienced different modes of gene duplication (23), little is known about the detailed evolutionary process of MADS-box genes. Here we investigate the pattern of birth-and-death evolution in the MADS-box gene family using all available MADS-box functional genes and pseudogenes from *Arabidopsis* and rice.

Materials and Methods

Identification of MADS-Box Functional Genes and Pseudogenes. In this paper, we assume that the annotated genes that encode a complete M-domain are functional genes, and the other genes are pseudogenes. To find functional proteins with M-domain from *Arabidopsis*, we performed the PSI-BLAST search (24) with an *E* value of $\leq 10^{-5}$ against the entire annotated proteins of *Arabidopsis* downloaded from the GenBank database (as of December 2002). We used 149 M-domain sequences from *Arabidopsis*, rice, animals, and fungi as queries. Similarly, we searched for functional MADS-box genes from annotated proteins of rice that are available from The Institute for Genome Research (TIGR). Because annotation of rice gene was still in progress at the time of this study, we ourselves conducted gene annotation by using the computer program FGENESH (www.softberry.com) from the genome sequences obtained from TIGR and the Rice Genome Database (China) (25).

To screen for pseudogenes from *Arabidopsis*, we first masked all annotated MADS-box gene loci (105 loci) in the genome sequence of *Arabidopsis*. We then performed the TBLASTN search with an *E* value of $\leq 10^{-5}$ against every possible reading frame of this MADS-masked genome with all M-domain protein sequences (105 sequences) from *Arabidopsis* as queries.

A similar search as that of *Arabidopsis* genes was performed to screen for MADS-box pseudogenes in rice. However, there could be a number of artificially fragmented genes by assembling errors in BAC/PAC clone sequences, resulting in a high false-positive rate of pseudogenes. For this reason, we used only M-domains to screen for pseudogenes and regarded a gene as a pseudogene if there is at least one stop codon in the MADS-box (see *File 1*, which is published as supporting information on the PNAS web site, for all of the sequences described above and their genomic locations).

Phylogenetic Analysis. Type II proteins generally contain both M- and K-domains that can be used for phylogenetic analysis. For this reason, type II genes were analyzed by using the M- and K-domains data set. When we constructed a tree for the entire types I and II genes, we used the M-domains data set, because type I proteins do not contain the K-domain. The classification of the MIKC-type proteins was made on the basis of the Hidden Markov Model search by using the computer program HMMER (26) and a K-domain matrix (see *File 2*, which is published as supporting information on the PNAS web site).

Protein sequences in each data set were aligned by using the computer program MAFFT (27) with the FFT-NS-i option. We then constructed neighbor-joining trees (28) by using the computer program MEGA2 (Ver. 2.1) (29). In addition, we constructed a maximum-parsimony consensus tree of the sequences in the M- and K-domains data set by using the PAUP* program with tree-bisection-reconnection (TBR) search with 100 bootstrap resamplings (30). Because our data set contained a large number of sequences, we did not use the maximum-likelihood method. We also did not use Bayesian phylogenetics, because this method often gives excessively high posterior probabilities even for wrong topologies (31–33).

Results

MADS-Box Functional Genes and Pseudogenes in *Arabidopsis* and Rice. Our homology search in *Arabidopsis* initially detected 105 functional M-domain protein sequences and 43 MADS-box pseudogenes. Sixteen of these 43 pseudogenes contained the MADS-box. After correcting two possibly misannotated pseudogenes, we finally found 107 functional genes and 41 pseudogenes. Seven of these 41 pseudogenes were annotated pseudogenes in the GenBank database.

In rice, we identified 71 nonredundant functional MADS-box genes. Our preliminary homology search also identified seven pseudogenes that contained stop codons in the MADS-box. Because the complete rice genome sequence is not publicly available at the present time, the numbers of MADS-box functional genes and pseudogenes may increase in the future.

Of 178 functional M-domain proteins from *Arabidopsis* and rice, our Hidden Markov Model search detected K-domains in 39 and 37 sequences from *Arabidopsis* and rice, respectively. These sequences were included in the M- and K-domains data set, and all M-domains of 178 functional MADS-box genes and the M-domains of 21 pseudogenes were included in the M-domains data set.

Number of Ancestral MADS-Box Genes in the Most Recent Common Ancestor (MRCA) of *Arabidopsis* and Rice. Because phylogenetic trees of type II genes were more reliable than those of type I genes, we first inferred the number of ancestral type II genes. Fig. 2 shows the evolutionary relationships of 79 type II genes from *Arabidopsis*, rice, mosses, and clubmosses. This tree has low bootstrap supports (<50%) for deep interior branches that determine interclade relationships. To infer the number of ancestral type II genes, however, this phylogenetic tree is quite informative. There are several clades including *Arabidopsis* and rice genes that are supported by a bootstrap value of ≥50%. We will call them shared clades. The genes in a shared clade are likely descendants of an ancestral MADS-box gene in the MRCA of *Arabidopsis* and rice. Thus, the number of shared clades is a minimum estimate of the number of MADS-box genes in the MRCA.

Fig. 2 shows there are 11 such shared clades, and most type II genes are members of these shared clades. A clade for so-called Bs genes was supported by a low bootstrap value (45%). However, we will consider this clade as a shared clade, because in a previous study they appeared to be monophyletic (34). Therefore, 12 shared clades of type II genes were identified. Of these, 11 clades (classes A, B-AP3, B-PI, Bs, C/D, E or AGL2, F or AGL20, G or AGL6, T or SVP, AGL12, and ANR1) were previously reported as different shared clades (9, 11, 20) and belong to the MIKC^c-type. The simplified class names F, G, and T are used according to Nam *et al.* (20). Class S is a shared gene class, and the genes in this class appear to be orthologous to the MIKC*-type genes from the moss. There are also three groups of genes (member genes of class E, F, and ANR1) that are not shared but are sister groups for three "shared clades" belonging to class E, class F, and class ANR1, respectively. Because these sister relationships are reasonably well supported (≥65%), it appears that at least two ancestral genes existed in each of classes E, F, and ANR1 in the MRCA of *Arabidopsis* and rice. Identification of 12 shared and 3 sister clades suggests that there were at least 15 ancestral type II MADS-box genes in the MRCA. A similar result was obtained from the maximum-parsimony tree (data not shown).

There are other type II genes (class FLC genes, *AGL15*, *AGL18*, and *OsMADS32*) that are not members of the above 15 clades. Orthologs of each of these genes might have been lost in either the *Arabidopsis* or the rice lineage, or our phylogenetic analysis could not resolve their evolutionary relationships. If the former is the case, the number of ancestral type II genes can be ≈20 in the MRCA of *Arabidopsis* and rice. Of course, we cannot exclude the possibility that this number is an underestimate because of the incomplete genome sequencing in rice.

Fig. 3 shows the phylogenetic tree of all MADS-box genes

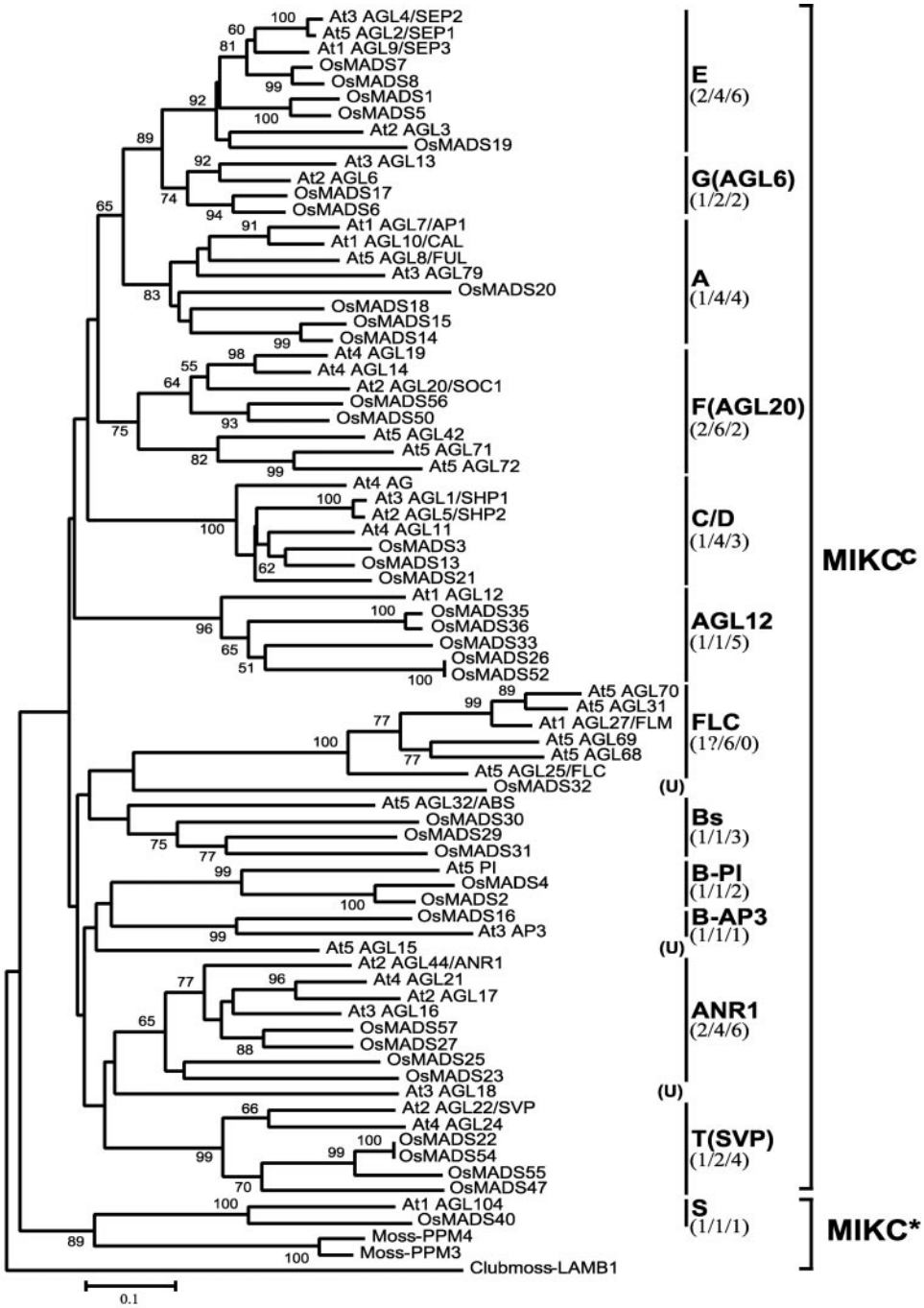

Fig. 2. Phylogenetic tree of 79 MIKC-type (type II) genes from *Arabidopsis*, rice, mosses, and clubmosses. This tree was constructed by the neighbor-joining method with Poisson-correction (PC) distance. One hundred five amino acids were used after all alignment gaps were eliminated. The number for each interior branch is the percent bootstrap value (500 resamplings), and only values >50% are shown. The scale bar indicates the number of amino acid substitutions per site. We generally followed ref. 4 on the notations of *Arabidopsis* genes. Simplified class names following ref. 20 were used except for classes A, B-AP3, B-PI, C/D, E, FLC, and S. The genes marked with (U) are unassigned genes for any classes. The three numbers in parentheses below each class name refer to the numbers of ancestral MADS-box genes, MADS-box genes in *Arabidopsis*, and MADS-box genes in rice, respectively. Two MIKC*-type genes (*PPM3* and *PPM4*) from the moss *Physcomitrella patens* (10) and one MIKC*-type gene (*LAMB1*) from the clubmoss *Lycopodium annotinum* (8) were used as reference sequences.

from *Arabidopsis* and rice and some of types I and II MADS-box genes from animals (223 sequences). According to this tree, all MADS-box genes that encode detectable K-domains form a clade together with animal type II genes, although they are not statistically well supported. Interestingly, there are also 13 MADS-box genes that do not encode an intact K-domain (at least in their predicted ORFs, genomic sequences, and adjacent genomic sequences) but are apparently very similar to MIKC-type (type II) genes. These genes might have lost the K-box during the evolution, or the absence or fragmentation of the K-box could be due to an assembling error of genome sequences. In this paper, these genes will be included in the type II genes on the basis of their close evolutionary relationships with other type II genes. The remaining MADS-box genes appear to have

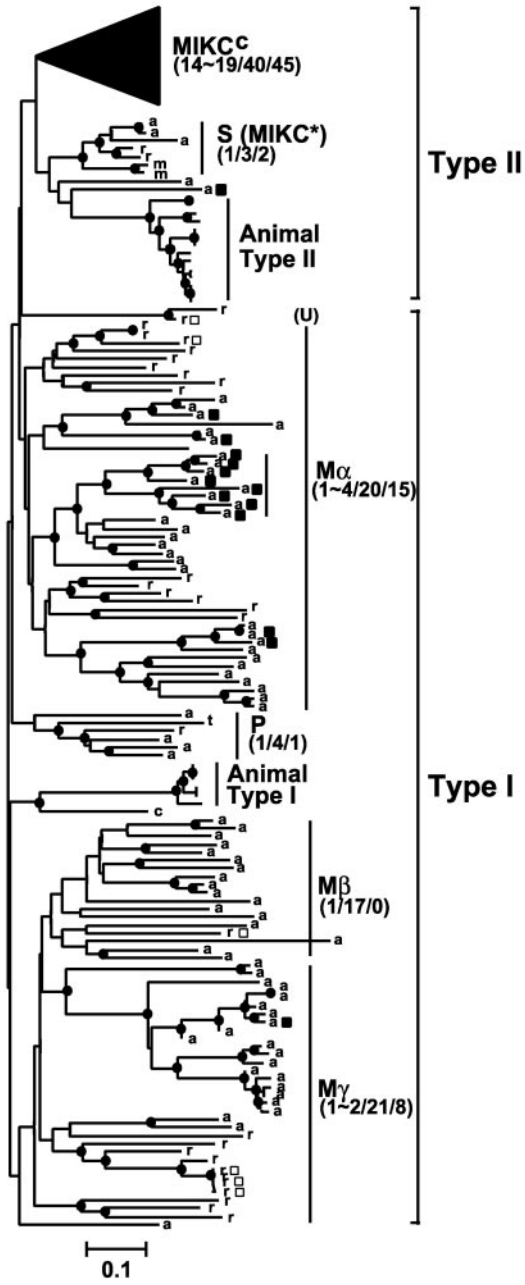

Fig. 3. Phylogenetic tree of 223 M-domain sequences from *Arabidopsis*, rice, mosses, clubmosses, and animals. This tree was constructed by the neighbor-joining method with *p*-distance and the pairwise deletion option (29) of ≈55 aa. *p*-distance is known to be more efficient in obtaining the correct topology when the sequence length is short (42). The genes from *Arabidopsis* and rice are labeled with ''a'' and ''r,'' respectively. The reference sequences from the moss *P. patens* and the clubmoss *L. annotinum* are labeled with ''m'' and ''c,'' respectively. Genes labeled with black squares (■) are pseudogenes from *Arabidopsis*, and those with open squares (□) are pseudogenes from rice. Interior branches with bootstrap values (500 bootstraps) >50% are indicated by black dots (●). The portion of the tree corresponding to the MIKC^c-type genes is compressed, because it is essentially the same as that in Fig. 2. The numbers in parentheses below each class name are the numbers of ancestral MADS-box genes, MADS-box genes in *Arabidopsis*, and MADS-box genes in rice, in this order.

diverged from type II genes before the animal/plant split, although the bootstrap support is weak. These genes correspond to the type I genes proposed by Alvarez-Buylla *et al.* (3). Fig. 3 also suggests that at least one type I gene existed in the MRCA of animals and plants. This observation is consistent with that of other researchers (3, 9, 23).

Most of the shared classes observed in Fig. 2 remain unchanged in the original tree of Fig. 3 (data not shown). We also identified another shared clade (P) supported by a bootstrap value of 79%. Class P genes from *Arabidopsis* were previously classified as type I by Alvarez-Buylla *et al.* (3). However, we also observed that MIKC*-type genes from mosses and class S and class P genes from *Arabidopsis* and rice formed a clade, when we used different sets of genes (data not shown). It is therefore possible that class P genes are also closely related to the MIKC*-type (type II) genes from mosses as proposed by other researchers (4, 5, 23), although they are not orthologous to the latter genes. On the basis of the tree shown in Fig. 3, the remaining type I genes can further be subdivided into classes Mα, Mβ, and Mγ, in agreement with Parenicova *et al.*'s (4) classification, although bootstrap supports of these classes are very low, and class Mγ genes are not monophyletic. Although our classification of type I genes is very crude, it suggests that at least about four to eight ancestral type I genes existed in the MRCA of *Arabidopsis* and rice. The numbers of functional genes and ancestral genes estimated in this way for each type of MADS-box genes in *Arabidopsis* and rice are shown in Fig. 3 (see numbers in parentheses).

Our study of ancestral MADS-box genes therefore leads to the hypothesis that there were at least ≈15–20 type II genes and at least about four to eight type I genes in the MRCA of *Arabidopsis* and rice. Because there are 43 type II genes and 64 type I genes in *Arabidopsis*, the results of the present study suggest that type I genes have experienced a higher birth rate than type II genes in the *Arabidopsis* lineage. A similar pattern was also observed in rice, although it is preliminary. In addition, this pattern is quite general across most gene classes except class FLC in *Arabidopsis* and class AGL12 in rice (see numbers in parentheses in Figs. 2 and 3). One may argue that if we use more stringent criteria for estimating the number of ancestral type I genes, the number may change, and therefore the rate of gene birth would change. However, this does not affect our conclusion that type I genes have experienced a higher birth rate than type II genes. This is because many type I genes in each of classes Mα, Mβ, and Mγ from either *Arabidopsis* or rice appear to be monophyletic, suggesting that they were duplicated after the *Arabidopsis* and rice split.

Classification of MADS-Box Pseudogenes. Existence of pseudogenes means that functional genes die sometimes in the evolutionary process. To examine whether there are differences in the death rate among different types of MADS-box genes, we classified pseudogenes on the basis of sequence similarity to functional MADS-box genes. In *Arabidopsis*, four pseudogenes were most similar to the type II genes (see Table 1, which is published as supporting information on the PNAS web site), and none of these pseudogenes had the MADS-box. The remaining 37 pseudogenes were most similar to the type I genes. Fourteen of these 37 pseudogenes had the MADS-box. In the case of rice, only one of the seven pseudogenes belonged to type II, and the remainder were type I genes. These results show that the proportion of pseudogenes is significantly different between types II and I genes in both *Arabidopsis* and rice. When we applied the same criterion of pseudogenes as that of rice pseudogenes (existence of stop codons in the MADS-box), we detected nine type I pseudogenes and no type II pseudogenes in *Arabidopsis*. Our homology search and phylogenetic analysis also showed that several pseudogenes belonging to class Mα are monophyletic (see Fig. 3), suggesting that the number of pseudogenes has increased recently in this lineage. Even if we exclude such lineage-specific pseudogenes, the difference in the proportion of

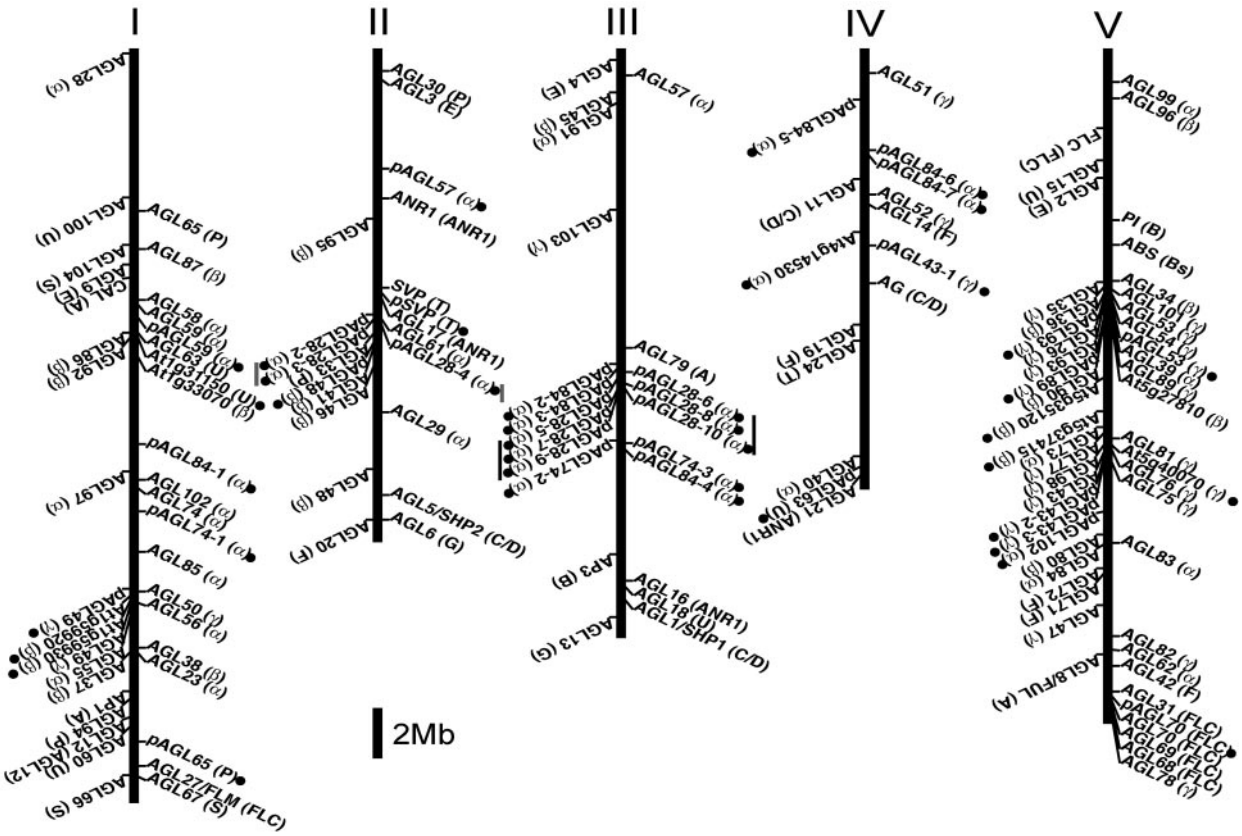

Fig. 4. Genomic organization of MADS-box genes in *Arabidopsis*. Genes with black dots (●) are pseudogenes. For seven annotated pseudogenes, we used their gene codes from GenBank. Unannotated pseudogenes are indicated by ''p'' in front of the name of the functional gene that is most similar to the pseudogene. For example, a pseudogene that is most similar to ''*SVP*'' is designated as ''pSVP.'' The class name of each gene is given in parentheses at the end of the gene name. Of these class names, ''(α),'' ''(β),'' ''(γ),'' and ''(U)'' refer to ''Mα,'' ''Mβ,'' ''Mγ,'' and ''Unassigned,'' respectively. I, II, III, IV, and V represent chromosome numbers. The scale bar below chromosome II is for 2 megabase pairs (Mb).

pseudogenes between types I and II genes is still substantial. Although type I genes are expected to include more pseudogenes than type II genes because of their higher birth rate, this factor alone does not explain the difference in pseudogenes between types I and II genes. Therefore, type I genes should have had a higher death rate than type II genes.

It is not easy to have an unambiguous definition of pseudogenes, because even a fragmentary gene can be functional (35, 36), and young pseudogenes may not be distinguishable from functional genes. Therefore, different criteria for pseudogenes may change our conclusion about the death rates of types I and II genes. As mentioned above, however, our conclusions about the death rates based on two different criteria in *Arabidopsis* are essentially the same. Note also that our searches for pseudogenes are apparently biased for pseudogenes similar to more conserved functional genes (type II genes in this study) than for less conserved functional genes. Therefore, our conclusion about the difference in death rate between types I and II genes is conservative.

Genomic Organization of MADS-Box Genes in *Arabidopsis*. The genomic locations of all MADS-box genes in *Arabidopsis* are shown in Fig. 4. In general, MADS-box genes are scattered all over the chromosomes. However, we also observed a number of clusters of closely located MADS-box genes in *Arabidopsis*. Most of these genes belonged to type I genes, and in general the genes in each cluster are evolutionarily closely related. These closely related MADS-box genes were probably generated by recent segmental duplication. The genomic locations of pseudogenes are also shown in Fig. 4. Most pseudogenes are closely located to each other as well as to their closely related functional MADS-box gene, although there are several exceptions. We also found a genomic cluster of type I pseudogenes without any functional MADS-box genes (but there are other genes) on chromosome 3 (genes with vertical bars in Fig. 4). The genomic locations and the phylogenetic tree of M-domain sequences (Fig. 3; see also Fig. 5, which is published as supporting information on the PNAS web site) suggest that this gene cluster was formed by segmental duplication of an ancestral pseudogene cluster, which was in turn duplicated from another pseudogene cluster on chromosome 2 (genes with gray bars in Fig. 4).

Rice MADS-box genes are also scattered all over the chromosomes, and more clusters of type I genes were found than those of type II genes (J.N., unpublished work).

Discussion

We have seen that type I genes have experienced faster birth-and-death evolution than type II genes in the *Arabidopsis* and rice lineages. The higher birth rate of type I genes is apparently caused by a higher rate of gene duplication, because duplicate genes generally do not cause harmful effects. In fact, type I genes are associated with a higher frequency of segmental duplications than type II genes in *Arabidopsis* (see Fig. 4). (We do not think the genome duplication is responsible for the different birth rates of types I and II genes, because in this case the birth rate should be the same for all genes. Therefore, we will not discuss this factor.) By contrast, the death of functional genes may have harmful effects, and therefore the death rate may be influenced

by functional requirements of duplicate genes as well as genomic events and fixation by genetic drift. Our estimates of the numbers of nonsynonymous nucleotide substitutions per nonsynonymous site (d_N) and synonymous nucleotide substitutions per synonymous site (d_S) suggested that type I genes have been under weaker purifying selection than type II genes (see Fig. 6, which is published as supporting information on the PNAS web site). This observation may explain why type I genes have experienced a higher death rate than type II genes, because the death of type I genes could be less harmful than that of type II genes. It is possible that, after duplication, type II genes became functionally differentiated in a relatively short time and therefore have been maintained as functional genes in the genome. This might be related to the extensive morphological diversification of angiosperms.

Although type I genes are apparently under weaker purifying selections than type II genes, they still might have played some important roles, because most of the recently duplicated type I gene pairs show significantly lower d_N and d_S (Fig. 6). Recently, it has been proposed that the expression of a type I MADS-box gene, *PHERES1*, in *Arabidopsis* is associated with seed abortion in a certain mutant background (37). However, the functions of type I genes are not well understood. If there are functionally redundant duplicate genes, it would be difficult to study their functions by mutagenesis experiments. Moreover, if type I genes are involved in a short period of developmental processes, it may also be difficult to study their functions. At the present time, gaining insights into the functional constraints of type I genes by evolutionary analysis may be of some help for future experimentation. Our study suggests that type I genes may be more variable among different angiosperm species than type II genes because of faster birth-and-death evolution than that of type II genes. In addition, type I genes are generally less conserved than type II genes.

There are a substantial number of type II duplicate genes, although the birth rate of type II genes is lower than that of type I genes. Therefore, some extent of functional redundancy or differentiation is expected to be observed among highly similar type II genes. For example, three class E genes (*AGL2/4/9* or *SEP1/2/3*) in *Arabidopsis* are known to be functionally redundant, because single gene mutations showed only subtle phenotypic changes, whereas triple mutants showed significant phenotypic changes in flowers of *Arabidopsis* (38). Nevertheless, our d_N and d_S analysis suggests that these genes are generally subject to strong purifying selection (see *File 3*, which is published as supporting information on the PNAS web site). Therefore, more careful study of single gene mutations may reveal some unrecognized phenotypic effects in plants. Moreover, there is substantial conservation or differentiation in gene expressions (4, 5) and in protein coding region (39–41) among paralogous MADS-box genes. By combining experimental studies with evolutionary analyses, we may be able to have a better insight into gene functions.

We thank Hongzhi Kong, Yoshi Niimura, Nikos Nikolaidis, Li Hao, Jim Leebens-Mack, Claude dePamphilis, Kerstin Kaufmann, Mitsuyasu Hasebe, Guenter Theissen, Doug Soltis, Mike Purugganan, and Lucia Colombo for useful comments. This work was supported, in part, by National Institutes of Health Grant GM20293 (to M.N.) and grants from the Crop Functional Genomic Center, the 21st Century Frontier Program, Korea (CG1111), and the Biogreen 21 Program, Rural Development Administration, Korea (to G.A.). J.N. was partially supported by a scholarship from the Rotary Foundation.

1. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. (2001) *From DNA to Diversity* (Blackwell Science, Malden, MA).
2. Shore, P. & Sharrocks, A. D. (1995) *Eur. J. Biochem.* **229,** 1–13.
3. Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Ribas de Pouplana, L., Martinez-Castilla, L. & Yanofsky, M. F. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5328–5333.
4. Parenicova, L., De Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., *et al.* (2003) *Plant Cell* **15,** 1538–1551.
5. Kofuji, R., Sumikawa, N., Yamasaki, M., Kondo, K., Ueda, K., Ito, M. & Hasebe, M. (2003) *Mol. Biol. Evol.* **20,** 1963–1977.
6. Lee, S., Kim, J., Son, J. S., Nam, J., Jeong, D. H., Jang, S., Lee, J., Lee, D. Y., Kang, H. G. & An, G. (2003) *Plant Cell Physiol.* **44,** 1403–1411.
7. Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. (1991) *Genes* **5,** 484–495.
8. Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Munster, T. & Theissen, G. (2002) *Mol. Biol. Evol.* **19,** 801–814.
9. Becker, A. & Theissen, G. (2003) *Mol. Phylogenet. Evol.* **29,** 464–489.
10. Svensson, M. E., Johannesson, H. & Engstrom, P. (2000) *Gene* **253,** 31–43.
11. Munster, T., Deleu, W., Wingen, L. U., Ouzunova, M., Cacharron, J., Faigl, W., Werth, S., Kim, J. T. T., Saedler, H. & Theissen, G. (2002) *Maydica* **47,** 287–301.
12. Weigel, D. & Meyerowitz, E. M. (1994) *Cell* **78,** 203–209.
13. Theissen, G. (2001) *Curr. Opin. Plant Biol.* **4,** 75–85.
14. Zhang, H. & Forde, B. G. (1998) *Science* **279,** 407–409.
15. Michaels, S. D. & Amasino, R. M. (1999) *Plant Cell* **11,** 949–956.
16. Alvarez-Buylla, E. R., Liljegren, S. J., Pelaz, S., Gold, S. E., Burgeff, C., Ditta, G. S., Vergara-Silva, F. & Yanofsky, M. F. (2000) *Plant J.* **24,** 457–466.
17. Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L. & Yanofsky, M. F. (2000) *Nature* **404,** 766–770.
18. Ma, H. & dePamphilis, C. (2000) *Cell* **101,** 5–8.
19. Kang, H. G., Jeon, J. S., Lee, S. & An, G. (1998) *Plant Mol. Biol.* **38,** 1021–1029.
20. Nam, J., dePamphilis, C. W., Ma, H. & Nei, M. (2003) *Mol. Biol. Evol.* **20,** 1435–1447.
21. Nei, M. (1969) *Nature* **221,** 40–42.
22. Nei, M., Gu, X. & Sitnikova, T. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 7799–7806.
23. De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouze, P., Theissen, G. & Van de Peer, Y. (2003) *J. Mol. Evol.* **56,** 573–586.
24. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
25. Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002) *Science* **296,** 79–92.
26. Eddy, S. R. (2001) HMMER: Profile Hidden Markov Models for Biological Sequence Analysis (Ver, 2.3.2), http://hmmer.wustl.edu.
27. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) *Nucleic Acids Res.* **15,** 3059–3066.
28. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
29. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001) *Bioinformatics* **17,** 1244–1245.
30. Swofford, D. L. (1998) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) (Sinauer, Sunderland, MA), Ver. 4.
31. Suzuki, Y., Glazko, G. V. & Nei, M. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 16138–16143.
32. Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A. & Winka, K. (2003) *Syst. Biol.* **52,** 477–487.
33. Misawa, K. & Nei, M. (2003) *J. Mol. Evol.* **57,** S290–S296.
34. Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M. A., Saedler, H. & Theissen, G. (2002) *Mol. Genet. Genomics* **266,** 942–950.
35. Chen, F., Kook, H., Milewski, R., Gitler, A. D., Lu, M. M., Li, J., Nazarian, R., Schnepp, R., Jen, K., Biben, C., *et al.* (2002) *Cell* **110,** 713–723.
36. Shin, C. H., Liu, Z. P., Passier, R., Zhang, C. L., Wang, D. Z., Harris, T. M., Yamagishi, H., Richardson, J. A., Childs, G. & Olson, E. (2002) *Cell* **110,** 725–735.
37. Kohler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W. & Grossniklaus, U. (2003) *Genes Dev.* **12,** 1540–1553.
38. Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. & Yanofsky, M. F. (2000) *Nature* **405,** 200–203.
39. Lamb, R. S. & Irish, V. F. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 6558–6563.
40. Immink, R. G., Ferrario, S., Busscher-Lange, J., Kooiker, M., Busscher, M. & Angenent, G. C. (2003) *Mol. Genet. Genomics* **268,** 598–606.
41. Vandenbussche, M., Theissen, G., Van de Peer, Y. & Gerats, T. (2003) *Nucleic Acids Res.* **31,** 4401–4409.
42. Takahashi, K. & Nei, M. (2000) *Mol. Biol. Evol.* **17,** 1251–1258.