Isolation and Characterization of Rice MADS Box Gene Homologues and Their RFLP Mapping

Yoriko SHINOZUKA, Shoko KOJIMA, Ayahiko SHOMURA, Hideaki Ichimura, Masahiro YANO, Kimiko YAMAMOTO, and Takuji SASAKI*

Rice Genome Research Program, the Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, 446-1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305-0854, Japan and National Institute of Agrobiological Resources, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan

(Received 9 February 1999; revised 6 March 1999)

Abstract

Thirty-five MADS box gene homologues were identified through a large-scale cDNA analysis in rice. Based on the nucleotide sequences of the 3'-untranslated region, these clones were classified into 11 independent species. Seven species were found to be new among the rice MADS box gene family, and the other 4 corresponded to the previously reported OsMADS1, OsMADS2, OsMADS4, and OsMADS5. The full nucleotide sequences of the 7 new species were determined. Each clone encoded a deduced protein of 164–267 amino acids. The K-domain of the MADS protein was conserved in all clones though with lower degree in clone S10304. Reverse transcription PCR analysis showed that clones E31254 and E31864 were expressed mainly in panicles. Dendrogram analysis suggested that E31254 and E31864 are close to Arabidopsis AGL9 and AP1, respectively. Restriction fragment length polymorphism (RFLP) linkage mapping revealed that the rice MADS box gene homologues reported here are not clustered but are located throughout the genome. The locus of E31864 on the RFLP map was closely linked to the long sterile lemma gene, g-I.

Key words: MADS box; RFLP mapping; rice; Oryza sativa L.

1. Introduction

Genetic and molecular analysis of homeotic genes from Arabidopsis and snapdragon (Antirrhinum) has provided the basis for the ABC model of floral development. Almost all of the homeotic genes encode proteins containing a well-conserved 56-amino-acid region termed the MADS box. So far, many MADS box genes have been isolated from dicotyledons such as Arabidopsis, snapdragon, and petunia, and from monocotyledons such as rice, maize, and sorghum. The conserved domain in MADS from many species has been a good target for examining the evolutionary diversification of eukaryotic morphology. However, the function of the MADS box genes of monocotyledons has not been studied in detail yet, as it is difficult to obtain the variety of mutants found in Arabidopsis and snapdragon. Because flower structure differs between monocot and dicot plants, it is not clear whether the ABC model can be also applied to monocot plants. In addition, recently, MADS box genes are known to function as transcription factors in several tissues other than flower meristem.

The Rice Genome Research Program (RGP) has been conducting a large-scale cDNA analysis aimed at cataloguing all expressed genes in rice. This cDNA catalogue is used to mine new family genes or iso-proteins by similarity search. For example, aspartate aminotransferases, chitinases, and zinc finger proteins in rice were first identified as novel members of these families.

In this paper, rice MADS box genes identified by similarity search from our cDNA catalogue are reported with their full sequences. Dendrogram analysis of these gene products was used to evaluate their relative position so as to allow speculation on their function within the ABC model. The loci of these rice MADS genes on the RFLP linkage map were determined for comparison with the loci of mutants of known phenotype.

Communicated by Mituru Takanami
* To whom correspondence should be addressed. Tel. +81-298-38-7015, Fax. +81-298-38-7468, E-mail: tsasaki@abr.affrc.go.jp
Table 1. List of rice MADS box gene homologues fully sequenced in this study.

| clone name | isolated cDNA resource          | amino acid length | accession number (DDBJ) |
|------------|---------------------------------|-------------------|------------------------|
| C50086     | heat shock callus               | 228               | AB003322               |
| E20969     | panicles at meiotic stage       | 224               | AB003323               |
| E31254     | panicles at premeiotic stage    | 239               | AB003324               |
| E31864     | panicles at premeiotic stage    | 267               | AB003325               |
| R3786      | root                            | 222               | AB003326               |
| S10304     | green shoot                     | 164               | AB003327               |
| S11905     | green shoot                     | 230               | AB003328               |

**MADS Domain**

| clone name | sequence |
|------------|----------|
| E20969     | MGRK| |
| E31864     | MGRK| |
| E31254     | MGRK| |
| R3786      | MGRK| |
| S11905     | MGRK| |
| S10304     | MGRK| |
| OsMADS1    | MGRK| |
| OsMADS2    | MGRK| |
| OsMADS5    | MGRK| |
| consensus  | MGRK| |

**K Domain**

| clone name | sequence |
|------------|----------|
| E20969     | KGRL| |
| E31864     | KGRL| |
| E31254     | KGRL| |
| R3786      | KGRL| |
| S11905     | KGRL| |
| S10304     | KGRL| |
| OsMADS1    | KGRL| |
| OsMADS2    | KGRL| |
| OsMADS5    | KGRL| |
| consensus  | KGRL| |

**Figure 1.** Alignment of deduced amino acid sequences of 7 new rice MADS box protein homologues and 4 known rice MADS box proteins for the MADS domain and the K-domain. The consensus amino acids are those identical in more than 6 proteins among 11 at the corresponding positions. These are indicated by boldface letters. The position indicated by asterisks have no dominant consensus amino acids. The gaps in the K-domain were introduced to maximize alignment. Sequences between MADS and K-domains show no homology among the 11 proteins. The sequence portions beyond the K-domain are not shown.

2. Materials and Methods

2.1. Identification and sequencing of rice MADS box cDNA clones

Rice cDNA clones with the MADS box motif sequence were screened in RGP's accumulated partial cDNA sequences with the amino acid sequence of the *Arabidopsis AGAMOUS* MADS box as a query by using the Basic Local Alignment Search Tool (BLAST) algorithm. The picked cDNA clones were classified by the sequences of their 3'-untranslated region (3'-UTR). The full nucleotide sequences of the longest clone in each group were determined on both strands with an automated DNA sequencer (Perkin-Elmer Applied Biosystems Inc., Foster City, USA) with a dye-primer or dye-terminator cycle sequencing kit (Perkin-Elmer).

2.2. Reverse transcription PCR analysis

Total RNA from rice tissues was prepared according to the SDS-phenol method with minor modification. First-strand cDNAs were synthesized from 2 µg of total RNA with a First-Strand cDNA Synthesis Kit (Pharmacia Biotech, Uppsala, Sweden). The concentrations of the first-strand cDNAs were adjusted by referring to the expression levels of actin in each tissue. PCR amplifi-
cation was done with 0.025 U AmpliTaq Gold enzyme (Perkin-Elmer), 200 mM dNTP, 10% glycerol, 400 mM primer designed from each 3'-UTR sequence (Table 1), and 1 ml of the first-strand cDNAs. The reaction mixture was incubated at 95°C for 12 min and then subjected to 35 cycles of PCR (95°C for 30 s, 55°C for 1 min, and 72°C for 1 min).

2.3. Construction of dendrogram

Amino acid sequences of cDNAs analyzed as MADS box gene homologues were compared with known MADS box proteins of Arabidopsis, maize, and rice using the PILEUP program of the GCG package (Genetic Computer Group, Madison, WI, USA),23 and a MADS protein dendrogram was constructed.

2.4. Linkage analysis

The loci of rice MADS box gene homologues were put on the high-density linkage map constructed by Harushima et al.24 The experimental procedures used for RFLP mapping were according to Kurata et al.25 To detect a single polymorphic pattern in Southern blotting, 3'-UTR-specific PCR primers were generated for each cDNA clone.

3. Results and Discussion

3.1. Isolation of rice MADS box gene homologues

Using the whole nucleotide sequence of Arabidopsis AGAMOUS as the query sequence, a BLASTX search among our 30 thousand partial nucleotide sequences identified 35 rice homologues showing significant similarities to the MADS box sequence of the query. The partial sequences came from rice cDNA libraries from callus, green or etiolated shoots, and embryos. The 35 clones were classified into 11 independent species based on their 3'-UTR sequences. Four of the 11 species corresponded to the previously reported rice MADS box genes OsMADS1,7 OsMADS2,8 OsMADS4,8 and OsMADS5.9 The remaining 7 species are newly assigned to the rice MADS box gene family.

3.2. Characteristics of sequences of new rice MADS proteins

Complete nucleotide sequences of the longest cDNAs from each of the 7 species were determined. Table 1 lists the clone numbers, encoded lengths of amino acids, and the DDBJ accession numbers. As shown in Fig. 1, the 7 novel proteins have a well-conserved amino acid sequence from the 4 known rice MADS proteins in the MADS domain, located immediately after the initial methionine residue.7-9 Only 1 exception, clone S10304 which had the shortest amino acid sequence among the 7 clones (164 residues), had 1 extra amino acid residue in the MADS region. The K-domain, which is ubiquitously observed in plant MADS box proteins and is thought to form a coiled-coil structure to participate in protein-protein interactions,26,27 was also present in all 7 MADS proteins (Fig. 1). The secondary structure calculated by Chou and Fasman's method28 predicted a helical structure in the K-domains of all 7 proteins in spite of the weak similarity of amino acid sequences in this domain (data not shown). This could indicate the ability of these 7 MADS proteins, including S10304, to function like the ordinary MADS proteins so far known among other plant species.

Figure 2. Expression level analysis of rice MADS box gene homologues. Expression levels in various tissues were analyzed by RT-PCR. Numbers at the top indicate tissues examined by RT-PCR: 1, root; 2, green shoot; 3, etiolated shoot; 4, panicles at premeiotic stage; 5, panicles at meiotic stage; 6, panicles at flowering stage; 7, panicles at ripening stage; 8, plasmid with each corresponding MADS box gene homologue.
3.3. Expression profile of rice MADS box gene homologues

RT-PCR analysis was done with various tissues to analyze the expression profile of the 7 MADS box gene homologues. As shown in Fig. 2, each MADS gene was expressed mainly in the tissue from which it was isolated, though the level and specificity of expression differed in each case. For example, S10304 from green shoots was ubiquitously expressed among the 7 tissues examined, but E31254 from panicles at the premeiotic stage was observed only in panicles from premeiosis to flowering. E31864 was also expressed in panicles, but at a low level.

Other rice MADS box proteins, such as OsMADS 1-8, 24, and 45 have all been isolated from immature floral tissues and are suggested to be involved in flower development.7-10,14 Our RT-PCR analysis of OsMADS1-5 revealed a high degree of expression in panicles (data not shown). The similar profiles of E31254 and E31864 suggests that they may function in floral organ development in rice (see below).

3.4. Dendrogram analysis of rice MADS box genes

MADS box proteins related to flower development were classified mainly into 4 groups, AP3, PI, AG, and AP1/AGL9, based on characteristics of their amino acid sequences.15 To elucidate the function of the rice MADS box gene homologues reported in this study, a dendrogram was constructed using the MADS domain for the 7 new rice MADS proteins and known MADS box proteins of Arabidopsis, maize, and rice. As shown in Fig. 3, the
7 rice homologues were scattered among the 4 representative MADS protein groups. For example, E20969 and E31864 are closely related to AP3 and AP1, respectively. The expression profile of E31864 in Fig. 2 supports its relatedness to floral development in which AP1 is involved. In contrast, E20969 was found in all tissues examined, although the degree of expression was slightly higher in the panicle at meiotic and flowering stages (Fig. 2). E20969 might act mainly but not specifically in pistil and whorl 2 (petal) development.

E31254 is closely related to OsMADS1 and OsMADS5, which cause dwarfism and early flowering in transgenic tobacco by ectopic expression. This suggests the possibility of E31254 participating in flower development as a constituent of the ABC model. A distinct expression of E31254 in the panicle, highest at the premeiotic stage and gradually decreasing to flowering stage, supports this idea (Fig. 2).

R3786 and S11905 are related most closely to AGL12 and AGL14, respectively. AGL12 and AGL14 from Arabidopsis are expressed only in root tissues and are phylogenetically isolated from other MADS box genes related to floral development; they are termed "orphan". R3786 and S11905 are the first relatives of these orphan genes to be found in rice. However, R3786 and S11905 were transcribed not only in root tissues but also in green and etiolated shoots and panicles throughout their developmental stages. These 2 rice MADS homologues could work in a general fashion to support development in rice plants.

C50086 is almost an orphan but shows weak relatedness to Arabidopsis AGL15, which is highly expressed in developing embryos. In contrast, C50086 was highly transcribed in green and etiolated shoots and ripening panicles. This might indicate a role for C50086 in the development of leaves and spikelets. C50086 is the first relative of AGL15 to be found.

The most distantly separated rice homologue, S10304, was expressed at almost the same level throughout embryo development and among root and shoot tissues. S10304 is a unique MADS protein because of its shortest amino acid sequence among known MADS proteins and the weak sequence similarity of its K-domain to others. Elucidation of its physiological function could add new insights into the function of MADS box genes.
3.5. RFLP linkage analysis of rice MADS box gene homologues

The loci of the 11 homologues of the rice MADS box gene (7 newly identified and 4 corresponding to OsMADS 1, 2, 4, and 5) on an RFLP linkage map were determined. The homologues were dispersed throughout the rice genome (Fig. 5). It is noteworthy that several of the homologues were mapped closely to the loci of traits related to floral morphogenesis.

C30192 (OsMADS1) was mapped closely to leafy hull sterile-1 (lhs-1)\textsuperscript{30} on chromosome 3. C10491 (OsMADS2) and S10304 were mapped to the area near Lax panicle (lax)\textsuperscript{31} and extra glume (eg)\textsuperscript{31} on chromosome 1. E23291 (OsMADS4) was closely linked to open hull sterile (ops)\textsuperscript{31} on chromosome 5. Although recent studies showed that neither C10491 nor S10304 is eg, the possibility remains that C30192, C10491, and E23291 might have some relation to lhs-1, lax, and ops, respectively. The localization of C10491 (OsMADS2) is consistent with that reported by Kang et al.\textsuperscript{10}

C1032 (OsMADS5) was mapped on chromosome 6. The localization is consistent with the result of Kang et al.\textsuperscript{10} From this localization, we suspected that C1032 might be related to depressed palea-1 (dp-1),\textsuperscript{32} and did more precise mapping to determine the loci of dp-1 and C1032. Genomic DNA from 90 F\textsubscript{2} plants with dp-1 was prepared, and the polymorphism in the dp-1 phenotype was analyzed by using 3'-specific probes from C1032. However, dp-1 segregated from C1032 and thus does not correspond to C1032 for its phenotype (data not shown).

E31864, which belongs to the API group in the dendrogram analysis, was mapped closely to the long sterile lemma (g-1)\textsuperscript{33} on chromosome 7. API is a class A MADS protein and is reported to act in whorls 1 (sepal) and 2 (petal).\textsuperscript{1} Mutations in API cause homeotic conversion of whorls 4 and 2 to whorl 3.\textsuperscript{3} Rice flowers lack sepals and petals, and instead have two small organs, called lodicules. Thus further analysis of the relationship between E31864 and g-1 could provide valuable insights for identifying the rice organs that correspond to whorls 1 and 2 in dicots. The locus of E31864 was also close to SbmADS2 from sorghum on the RGP map. A similarity search against public databases suggested that E31864 encodes a protein that is most similar to ZAP1, the maize homologue of Arabidopsis floral homeotic gene AGAMOUS.\textsuperscript{34}

Acknowledgements: This work was supported by the Ministry of Agriculture, Forestry, and Fisheries, Japan, and by the Japan Racing Association.

References

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

2. Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. 1990, The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors, Nature, 346, 35–39.

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

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

5. Tröhner, W., Ramirez, L., Motte, P. et al. 1992, GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis, EMBO J., 11, 4693–4704.

6. Angenent, G. C., Busscher, M., Franken, J., Mol, J. N. M., and van Tunen, A. 1993, Differential expression of two MADS box genes in wild-type and mutant petunia flowers, Plant Cell, 4, 983–993.

7. Chung, Y.-Y., Kim, S.-R., Finkel, D., Yanofsky, M. F., and An, G. 1994, Early flowering and reduced apical dominance result from ectopic expression of a rice MADS box gene, Plant Mol. Biol., 26, 657–665.

8. Chung, Y.-Y., Kim, S.-R., Kang, H.-G. et al. 1995, Characterization of two rice MADS box genes homologous to GLOBOSA, Plant Science, 109, 45–56.

9. Kang, H.-G. and An, G. 1997, Isolation and characterization of a rice MADS box gene belonging to the AGL2 gene family, Mol. Cells, 7, 45–51.

10. Kang, H.-G., Jang, S., Chung, J.-E., Cho, Y.-G., and An, G. 1997, Characterization of two rice MADS box genes that control flowering time, Mol. Cells, 7, 559–566.

11. Schmidt, R. J., Velt, B., Mandel, M. A., Mena, M., Hake, S., and Yanofsky, M. F. 1993, Identification and molecular characterization of ZAG1, the maize homologue of the Arabidopsis floral homeotic gene AGAMOUS, Plant Cell, 5, 729–737.

12. Theißen, G., Strater, T., Fischer, A., and Saedler, H. 1995, Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of AGAMOUS-like MADS-box genes from maize, Gene, 156, 155–166.

13. Montag, K., Salamini, F., and Thompson, R. D. 1995, ZEMa, a member of a novel group of MADS box genes, is alternatively spliced in maize endosperm, Nucl. Acids Res., 23, 2168–2177.

14. Greco, R., Stagi, L., Colombo, L., Angenent, G. C., Sari-Gorla, M., and Pé M. E. 1997, MADS box genes expressed in developing inflorescences of rice and sorghum, Mol. Gen. Genet., 253, 615–623.

15. Purugganan, M. D., Rounsley, S. D., Schmidt, R. F., and Yanofsky, M. F. 1995, Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family, Genetics, 140, 345–356.

16. Rounsley, S. D., Ditta, G. S., and Yanofsky, M. F. 1995, Diverse roles for MADS box gene in Arabidopsis development, Plant Cell, 7, 1259–1269.

17. Yamamoto, K. and Sasaki, T. 1997, Large-scale EST sequencing in rice, Plant Mol. Biol., 35, 135–144.

18. Song, J., Yamamoto, K., Shimura, A., Yano, M., Minobe, Y., and Sasaki, T. 1996, Characterization and
mapping of cDNA encoding aspartate aminotransferase in rice, *Oryza sativa*, *DNA Res.*, 3, 303-310.

19. Nagasaki, H., Yamamoto, K., Shomura, A. et al. 1997, Rice class III chitinase homologues isolated by a random cloning strategy of rice cDNAs, *DNA Res.*, 4, 379-385.

20. Song, J., Yamamoto, K., Shomura, A. et al. 1998, Isolation and mapping of a family of putative zinc-finger protein cDNAs from rice, *DNA Res.*, 5, 95-101.

21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, 215, 403-410.

22. Hattori, T., Iwasaki, Y., Sakajo, S., and Asahi, T. 1983, Cell-free synthesis of succinate dehydrogenase and mitochondrial adenosine triphosphatase of sweet potato, *Biochem. Biophys. Res. Commun.*, 113, 235-240.

23. PILEUP Version 9.0 Users Manual, Genetics Computer Group, Madison, Wisconsin, USA.

24. Harushima, Y., Yano, M., Shomura, A. et al. 1998, A high density rice genetic linkage map with 2275 markers using a single F2 population, *Genetics*, 148, 479-494.

25. Kurata, N., Nagamura, Y., Yamamoto, K. et al. 1994, A 300 kilobase interval genetic map of rice including 883 expressed sequences, *Nature Genet.*, 8, 365-372.

26. Ma, H., Yanofsky, M. F., and Meyerowitz, E. M. 1991, AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes, *Gene & Develop.*, 5, 484-495.

27. Davies, B., Egea-Cortines, M., Silva, E. A., Saedler, H., and Sommer, H. 1996, Multiple interactions amongst floral homeotic MADS box proteins, *EMBO J.*, 15, 4330-4343.

28. Chou, P. Y. and Fasman, G. D. 1974, Prediction of protein conformations, *Biochemistry*, 13, 222-245.

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

30. Kinoshita, T., Hidano, Y., and Takahashi, M. 1977, A mutant "long hull sterile" found in the rice variety "So-rachi" — Genetic studies on rice plant. LXVII, *Mem. Fac. Agr. Hokkaido Univ.*, 10, 247-268 (in Japanese with English summary).

31. Iwata, N. and Omura, T. 1971, Linkage analysis by reciprocal translocation method in rice plants (*Oryza sativa* L.) 1. Linkage groups corresponding to the chromosomes 1, 2, 3 and 4, *Japan. J. Breed.*, 21, 19-28 (in Japanese with English summary).

32. Nagamatsu, T. and Omura, T. 1962, Linkage study of the genes belonging to the first chromosome in rice, *Japan. J. Breed.*, 12, 231-236.

33. Nagao, S. and Takahashi, M. 1963, Trial construction of twelve linkage groups in Japanese rice. Genetic studies on rice plant. XXVII, *J. Fac. Agr. Hokkaido Univ.*, 53, 72-130.

34. Devos, K. M. and Gale, M. D. 1997, Comparative genetics in the grasses, *Plant Mol. Biol.*, 35, 3-15.

35. Yoshimura, A., Ideta, O., and Iwata, N. 1997, Linkage map of phenotype and RFLP markers in rice, *Plant Mol. Biol.*, 35, 49-60.
