Auxin response factor family in rice

Yutaka Sato1,2, Asuka Nishimura1, Momoyo Ito1, Motoyuki Ashikari1, Hiro-Yuki Hirano3, Makoto Matsuoka1*

1BioScience Center, Nagoya University, Chikusa, Nagoya 464-8601, Japan.
2Present address: Plant Gene Expression Center, 800 Buchanan Street, Albany, CA94710, USA.
3Graduate School of Agricultural and Life Sciences, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan

(Received 10 September 2001, accepted 15 October 2001)

We isolated 11 rice genes homologous to the genes encoding auxin response factors (ARFs) in Arabidopsis. All of the genes encoded a well-conserved amino acid sequence in the N-terminal region, which is considered to be a DNA-binding domain (DBD). Phylogenetic analysis based on comparison of the DBDs indicated that rice has one or two closely related orthologs corresponding to a given respective ARF gene in Arabidopsis. We also analyzed the amino acid sequences of another conserved domain in the C-terminal conserved domain (CTD), which was shared by almost all the rice ARFs, with the exception of OsETTIN1 and OsETTIN2. These results agreed well with the evolutionary relationship deduced from the DBD comparison. In contrast to many ARFs, OsETTIN1 and OsETTIN2 do not contain the conserved C-terminal domain, but do share another consensus motif that is also found in Arabidopsis ETTIN. All of the above observations indicate that rice has functionally diversified ARF genes whose structures and functions correspond to those of various Arabidopsis ARFs, with one or two rice ARFs corresponding to a given Arabidopsis ARF. Thus, auxin signal transduction mechanisms may be well conserved between monocot and dicot plants.

INTRODUCTION

The plant hormone auxin regulates various aspects of plant growth and development such as cell elongation and division, organ and tissue differentiation and morphogenesis, and tropism. These biological events are triggered and controlled by auxin through signal transduction in the cells. One of the well-studied reactions of auxin signal transduction is the activation of a specific group of transcription factors (Abel and Theologis, 1996). The activated transcription factors may enter the nucleus and promote the expression of specific genes. Genes whose expression is stimulated by the activation of preexisting transcription factors are called primary response genes (Abel et al., 1994; Abel and Theologis, 1996; Guilfoyle et al., 1998). This directly means that all of the transacting factors required for the expression of the primary response genes are already present in the cell, and therefore inhibitors of protein biosynthesis such as cycloheximide do not block the expression of primary response genes.

At present, various kinds of primary auxin responsive genes have been identified, including members of the Aux/IAA gene family, SAUR gene family, and GH3 gene family, genes encoding aminocyclopropane-1-carboxylate (ACC) synthase, and genes encoding glutathione S-transferase-like proteins (Abel and Theologis, 1996). Through promoter analyses of these primary auxin responsive genes, auxin responsive elements (AuxREs) in these genes have been identified, and it has been demonstrated that a core motif, TGTCTC, is often present in the AuxREs as direct or inverted repeats (Guilfoyle et al., 1998; Ulmasov et al., 1997a; Ulmasov et al., 1997b).

Using the core motif, TGTCTC, as a bait sequence in a yeast one-hybrid system, Ulmasov et al. (1997a) isolated a transcription factor from Arabidopsis, auxin-response factor 1 (ARF1), which can interact with the core motif. Interaction between ARF1 and the TGTCTC motif in AuxRE has been considered to be important for the auxin-dependent response of AuxRE, since sequences that interact with ARF1 in vitro can promote the expression of a reporter gene in carrot protoplasts while sequences which do not interact with ARF1 in vitro can not induce such expression (Ulmason et al., 1997a).

ARF1 contains two unique domains, namely, a DNA binding domain in the N-terminal portion and a protein-
protein interaction domain in the C-terminal portion (Ulmasov et al., 1999b). The DNA binding domain (DBD) in the N-terminal portion has some sequence similarity to a DNA binding domain of a maize transacting factor, VP1, which has been shown to be able to interact with cis-acting elements of some ABA-responsive promoters (Suzuki et al., 1997). Recently, Ulmasov et al. (1999b) also demonstrated that this domain is essential for the interaction between ARF1 and the TGTCTC AuxRE.

The amino acid sequence located in the C-terminal portion was firstly identified as a sequence similar to the C-terminal regions of Aux/IAA proteins, which are encoded by members of primary auxin-responsive genes (Ulmasov et al., 1997a). There are four conserved domains, domains I, II, III, and IV, in the C-terminal portion of Aux/IAA proteins and ARF1 contains the latter two domains, domains III and IV, which together are called the C-terminal conserved domain (CTD). Recent studies have revealed that these domains facilitate protein-protein interactions among the members of both the ARF and Aux/IAA protein families (Kim et al., 1997; Ulmasov et al., 1997a; Ulmasov et al., 1997b).

Recently, Ulmasov et al. (1999b) isolated more than 10 genes encoding ARF1-homologous proteins from Arabidopsis, and demonstrated that these proteins function as transcriptional regulators of genes containing AuxREs. On the other hand, molecular genetic analyses using morphological mutants of Arabidopsis have revealed that some members of these ARF family proteins are essential for the normal development of Arabidopsis. For example, MONOPTEROS (MP) encodes a typical ARF-type protein, named ARF5 by Ulmasov et al. (1997a), and the loss-of-function of this gene causes defects in the formation of the vascular system in the embryo (Hardtke and Berleth, 1998). Based on these observations, these ARF proteins have been postulated to have important functions in plant developmental processes through the control of auxin signaling.

We have been studying on the molecular mechanism of the development of organs and tissues of Oryza sativa (rice), because rice is one of the most important crops in the world and also is a model for monocot plants. At present, some of the essential tools for molecular biological studies have been established for use in rice, including stable transformation techniques, transposon-tagged mutant libraries, and map-based cloning systems with fine map information and numerous molecular makers (Izawa and Shimamoto, 1996). However, there is not yet any information available on auxin biosynthesis or signaling in rice with the exception of the findings of classical physiological experiments that showed the importance of auxin for the development and growth of rice plants. In the present study, we isolated 11 genes encoding ARF proteins from rice and analyzed the structural features and evolutionary relationships of these rice ARF genes. Based on comparative studies of rice and Arabidopsis ARF genes, we suggest that the divergence of the ARF genes occurred before the divergence of dicots and monocots.

MATERIALS AND METHODS

ARF gene isolation and sequencing. Total RNA was extracted from developing embryos of rice (Oryza sativa CV.Nipponbare)(Sentoku et al., 2000), and poly(A)+-RNA was enriched by two passages through an oligo d(T) cellulose column. The poly(A)+-RNA was used to synthesize double-stranded cDNA, which was cloned into the EcoRI site of λgt11 (Stratagene, La Jolla, CA). Nuclear genomic DNA was isolated from 2-week-old seedlings. The DNA was partially digested with SauAI, and fragments of ~20 kb were enriched by sucrose-gradient centrifugation, and cloned into the BamHI site of EMBL3 (Stratagene). Screening by hybridization was performed in 30% formamide, 6 x SSC (1 x SSC is 0.15 M NaCl and 0.015 M sodium citrate), 5 x Denhardt’s solution (1 x Denhardt’s solution is 0.02% Ficoll, 0.02% PVP, and 0.02% BSA), 0.5% SDS, and 0.1 mg mL⁻¹ salmon sperm DNA at 42°C for 14 hr, using rice EST clones C6147, C11460, E60432 as probes. Nucleotide sequences were determined by the dideoxynucleotide chain-termination method using an automated DNA sequencing system (ABI 373A; Applied Biosystems, Inc., Foster City, CA), according to the manufacturer’s protocol (Perkin-Elmer). The isolated clones were completely sequenced on both strands. The alignment of the amino acid sequences and phylogenetic analysis were performed using the DDBJ analysis server (http://www.ddbj.nig.ac.jp/E-mail/clustalw-j.html).

Mapping of ETTIN1 in rice recombinant inbred lines. To map the ETTIN1 genes, 98 BC1F2 lines (back-cross inbred lines) derived from the cross between a japonica variety, Nipponbare, and an indica variety, Kasalath, were used (Lin et al., 1998). Linkage analysis of ETTIN1 and 245 RFLP markers was performed using MAPMAKER Version 3.0 (Lander et al., 1987).

RESULTS AND DISCUSSION

Isolation of ARF family in rice. Ulmasov et al. (1997a, 1999a, 1999b) reported the cloning of more than 10 genes encoding ARF proteins (ARF1–10) in Arabidopsis. Some of them were independently isolated through the molecular genetic approach using morphological mutants, these were MONOPTEROS (MP=ARF5, Hardtke and Berleth, 1998), ETTIN (=ARF3, Sessions et al., 1997) and NPH4 (=ARF7, Harper et al., 2000). Based on the characterization of these mutants, the biological functions of these genes are relatively well characterized in comparison to those of other ARF genes. Therefore, we attempted to isolate as many ARF-like genes, including ARF3, 5, and
7, as possible from rice. For this purpose, we used two libraries: a cDNA library from developing embryos, and a rice genomic library. These libraries were screened with EST clones C61470, C11460, and Au030910, released from the Rice Genome Project (RGP); each of these clones encoded a conserved domain of ARF. More than 50 positive clones were isolated and classified according to their restriction-enzyme digestion patterns and partial sequences. Finally, we focused on 11 clones, including clones for genes homologous to MP, ETTIN, and ARF7, and determined the entire sequences of these clones.

All of the predicted proteins contained highly conserved region of about 320 amino acid residues in their N-terminal portion, which corresponds to DBD of the Arabidopsis ARF family, and most of the OsARF proteins contained CTD, (motifs III and IV) although OsETTIN1 and OsETTIN2 did not (Figs. 1 and 2). The lack of CTD in OsETTIN1 and OsETTIN2 as well as in of Arabidopsis ETTIN/ARF3 suggest that these rice proteins are orthologs of Arabidopsis ETTIN (see below). Between DBD at the N-terminus and CTD at the C-terminus, there is non-conserved region called the middle region (Ulmasov et al., 1999b). Although this middle region does not contain any consensus motif in ARF proteins, it also shows a unique feature, that is, a high frequency of one or a few amino acid residues (Fig. 1). Indeed, the middle regions of OsMP (OsARF5), OsARF6, OsARF7, and OsARF8 are rich in glutamine (Q), leucine (L), and serine (S), while OsARF2 contains a high proportion of S and proline (P). OsARF16 also shows a biased amino acid composition in this region (Fig. 1). Such biased distribution of particular amino acid residues is also observed in Arabidopsis ARF proteins, for example, ARF5–8 contains high proportions of Q, L, and S (Ulmasov et al., 1999a). In some Arabidopsis ARFs, these non-conserved regions with biased amino acid compositions work as transcriptional activation or repression domains, depending on the kind of biased amino acid residues (Ulmasov et al.,

![Fig. 1. Schematic diagrams of rice OsARF proteins.](image)

The shaded oval of each OsARF protein indicates the DNA binding domain (DBD) located in the N-terminal portion. The box present in the C-terminal portion of each OsARF, with the exception of OsETTIN1 and OsETTIN2, represents the dimerization domain that corresponds to motifs III and IV found in Aux/IAA proteins. The central region of each OsARF protein except OsETTIN1 and OsETTIN2 contains a biased amino acid composition, as indicated. The rhombuses in ETTIN1 and OsETTIN2 indicate the direct repeat sequences specifically observed in proteins of the ETTIN subfamily.
Therefore, the middle regions of OsARF proteins may also be involved in the transcriptional activity of these proteins.

**DNA binding domain in the N-terminal portion.**

For more precise comparison of these rice OsARF proteins, we aligned the amino acid sequences of the DBDs of the OsARF proteins (Fig. 2). The positions of introns are also indicated along with the amino acid sequences. The positions of intron are well conserved among these genes with a few exceptions. OsARF2, OsARF6, OsARF7, OsARF8, and OsMP have 9 introns in this region, designated A to I and indicated by closed red arrowheads. In contrast, OsETTIN1 and OsETTIN2 lack the intron position A, as indicated by an open arrowhead. OsARF10 is an exceptional case, because it has only one intron (at position 1) in this region. The exon-intron structure of OsARF10 is distinct from that of the other OsARF genes. Accession numbers are as follows: OsETTIN1 (AB071290), OsETTIN2 (AB071291), OsMP (AB071292), OsARF2 (AB071293), OsARF6a (AB071294), OsARF6b (AB071295), OsARF7a (AB071296), OsARF7b (AB071297), OsARF8 (AB071298), OsARF10 (AB071299), and OsARF16 (AB071300).
and this distinction between OsARF10 and other ARFs agrees with the phylogenetic relationships deduced from amino acid comparison (Fig. 3 and Fig. 4C). The positions of introns are well conserved among the genes of rice and Arabidopsis (Ulmasov et al., 1999b), which can be classified into common clades (Fig 3), suggesting again that these genes have orthologous relationships.

Based on the alignment of the amino acid sequences of rice ARF proteins and of Arabidopsis ARF proteins (Ulmasov et al. 1999b), we analyzed the evolutionary relationships among rice and Arabidopsis ARF proteins. Dendrograms were obtained by the neighbor-joining method (Fig. 3), and the maximum-parsimony method gave similar relationship (data not shown). This phylogeny shows that one or two rice ARF proteins correspond to given Arabidopsis ARF protein. For example, OsETTIN1 and OsETTIN2 show high similarity, and these two proteins belong to the same clade as Arabidopsis ETTIN/ARF3. As we could not find another protein grouped into this clade in the database of the complete Arabidopsis genome sequence, the Arabidopsis genome appears to have only one gene in this subfamily while the rice genome has at least two genes. This suggests that OsETTINs function redundantly in rice, in contrast to ETTIN in Arabidopsis. In fact, a single mutation in the Arabidopsis ETTIN gene causes the production of abnormal flowers with increased perianth organ number, decreased stamen number, and abnormal apical-basal patterning of the gynoecium (Sessions et al., 1997). In rice, on the other hand, no morphological mutant has been reported so far around the map position of OsETTIN1, which is located on the long arm of chromosome 5 near the centromere (tightly linked with a molecular marker, R521, released by RGP).

One-to-two relationships between rice and Arabidopsis are also observed in the clades of OsARF6/ARF6 and OsARF7/NPH4. The high redundancy of rice ARF proteins may suggest that rice needs more ARF proteins than Arabidopsis, and it is possible that the functions of ARF proteins are more precisely specialized in rice for more restricted organs or tissues or for specified stages. In contrast to the above cases, OsMP, OsARF2, OsARF8, OsARF10, and OsARF16 each showed a one-to-one relationship with the corresponding Arabidopsis ARF protein. This suggests that these proteins function in a similar manner to the corresponding Arabidopsis ARF proteins. It remains possible that there are other related rice ARF protein, which we could not identify in this work. Indeed, we detected no rice counterpart proteins corresponding to ARF4, ARF9, or ARF1. More extensive analyses will be necessary to fully clarify the relationships among the members of the ARF family in rice and Arabidopsis.

**Protein-protein interaction domain at the C-terminus.** We also aligned the amino acid sequences of the CTDs of the OsARF proteins (Fig. 4A). The insertion sites of introns are indicated by red arrowheads, as they were in Figure 2. All these OsARF proteins have two introns at the same sites, with the one exception of position X in OsARF10. OsARF10 lacks this intron, as does Arabidopsis ARF10, again suggesting that these proteins are orthologs of each other.

Kim et al. (1997) predicted that the secondary structure of motif III of IAA proteins has a βαα structure, and pointed out that this structure is similar to the βαα DNA-binding domain of prokaryotic repressor proteins such as Arc. As it has been pointed that the second α-helix of the βαα structure is important for dimerization, we analyzed
Fig. 4. Comparison and secondary structure of the protein-protein interaction domain (the C-terminal domain (CTD)). (A) Sequence alignment of CTDs of OsARF proteins. White letters on a black background or black letters on a gray background indicate the same things as in Figure 2. The predicted leucine zipper structure with specifically arranged leucine and methionine residues is bracketed, and marked by arrows at the top of the alignment. Motifs III and IV are consensus sequences shared by Aux/IAA proteins. (B) Helical wheel representation of the α-helix structure of OsMP. (C) Phylogenetic analysis of rice and Arabidopsis ARF proteins based on the CTD alignment. The tree was generated by the neighbor-joining method. The Arabidopsis sequences were taken from Ulmasov et al. (1999b).

the predicted secondary structure of the corresponding region of OsMP as one example of this family (Fig. 4B). Other OsARF proteins, except OsARF10 and OsARF16, also form a similar secondary structure (data not shown). According to the algorithm of Chou-Fasman (Chou and Fasman, 1978) and Robson (Garniar et al., 1978), this region forms an α-helix, and leucine or methionine residues (indicated by arrows in Figure 4A) are periodically aligned in this helical region to form a leucine zipper structure (Fig. 4B). In OsARF10 and 16, however, the residue at the third position does not correspond to suitable amino acids such as leucine, isoleucine or methionine, but is rather replaced by lysine in OsARF16 and glutamine in OsARF10. These exchanges should affect the function of the leucine zipper structure, and it is possible that such differences in the amino acid residues in this region are involved in the recognition of partner proteins.

Using the alignment of this region, we also evaluated the evolutionary relationships among rice and Arabidopsis ARF proteins (Fig. 4C). The phylogenetic tree constructed using this region is essentially the same as that constructed using DBDs (Fig. 3). This supports the conclusion that these phylogenetic trees reflect the real relationships of these ARF proteins. However, there is one small contradiction between these trees, that is, the position of OsARF16. OsARF16 has an almost equal distance from three Arabidopsis ARF proteins, including ARF16, in the CTD tree (Fig. 4C), while this protein is closely related to ARF6 in the DBD tree (Fig. 3). Besides these conserved domains, in the case of OsARF16 and ARF16, clear similarity is observed in the region between DBD and CTD, in which amino acid sequences are not generally con-
served among ARF proteins. Taking all these results together, we conclude that OsARF16 is a rice ortholog of Arabidopsis ARF16.

Comparison between primary structures of rice OsETTIN and Arabidopsis ETTIN proteins. In contrast to other ARF proteins, neither ETTIN nor OsETTIN retains a CTD in the C-terminal portion. This unique feature leads us to speculate that ETTIN may have a unique function(s). Based on this consideration, we suspected that ETTIN might have a unique structure(s) in its C-terminal region, in addition to the well-conserved DBD

Fig. 5. Comparison of the amino acid sequences of OsETTIN1, OsETTIN2 and ETTIN/ARF3. (A) Dashes indicate gaps introduced to maximize matching. Asterisks or dots indicate the same residues in all three proteins or in two proteins, respectively. The direct repeats of 19 amino acids are boxed and indicated as (1) and (2). (B) Alignment of amino acid sequence in the direct repeats specific to ETTIN proteins. White letters on a black background or black letters on a gray background indicate the same things in Figure 2.
shared with other ARF proteins. To test this possibility, we compared the entire amino acid sequences of OsETTIN1, OsETTIN2 and Arabidopsis ETTIN (Fig. 5A). The N-terminal halves corresponding to the DBD were well conserved, while such strong similarity was not seen in the C-terminal halves. However, there were two conserved motifs in each ETTIN protein (Fig. 5A). Interestingly, these two motifs found in each ETTIN protein were similar to each other, and thus ETTIN proteins contain direct repeats of amino acid sequence in this region (Fig. 5B). These direct repeats are composed of 19 amino acid residues, 8 of which are completely conserved in all these repeats. These consensus residues are flanked by the same or similar amino acid residues in these repeats. According to computer-assisted predictions of the secondary structure, this region forms an α-helix adjacent to two random coil structures (data not shown). Therefore, these α-helices with conserved amino acids may be able to move easily. It is possible that these α-helix structures, which are not shared with other ARF proteins, are involved in the function(s) specific to proteins in the ETTIN subfamily.

CONCLUSIONS

Rice has more than 11 genes encoding auxin response factors (ARFs), which are considered to be involved in auxin signal transduction. A comparative study of the primary structures of rice and Arabidopsis ARFs revealed that rice contains one or two closely related orthologous gene(s) corresponding to each respective Arabidopsis ARF gene. This one-to-one or one-to-two relationship between rice and Arabidopsis ARF genes suggests that the function(s) of the corresponding ARF proteins in these plants may be similar to each other. We also showed that ARF proteins in the ETTIN subfamily share a short direct repeat of amino acid sequence in the C-terminal portion.

This work was supported by grant-in-aid for Center of Excellence (COE) research to M. M.

REFERENCES


