Theor Appl Genet (1998) 97: 9–19
© Springer-Verlag 1998

M. Iwamoto · M. Maekawa · A. Saito
H. Higo · K. Higo

Evolutionary relationship of plant catalase genes inferred from exon-intron structures: isozyme divergence after the separation of monocots and dicots

Received: 20 November 1997 / Accepted: 5 January 1998

Abstract In order to understand the molecular evolution of catalase genes in higher plants, we compared the exon-intron structures of 12 genomic sequences from six plant species. It was assumed that the putative single primordial catalase gene had seven introns, because only those catalase genes having this structure are found in the monocotyledonae and dicotyledonae classes. After the evolutionary divergence of monocots from dicots, consecutive duplication of the primordial gene followed by the differential loss of introns occurred in each class to form three (or possibly four in dicots) diverse isozyme genes. In monocots, three ancestral isozyme genes were formed before the divergence of ancestral rice and maize. One of the rice genes, *CatA*, has an entirely new short intron which was not found in any other plant catalase gene examined. We have investigated the existence of the intron in the *CatA* homolog in other rice species by polymerase chain reaction (PCR) analysis. One major PCR product was found with the genomic DNAs from *O. sativa* (indica and japonica types), *O. rufipogon* and *O. glaberrima*. DNAs from several accessions of *O. longistaminata* showed variation in both the number and size of the DNA fragments amplified. PCR analyses and sequencing of the PCR products revealed that there are several *CatA* homologs having different sequences in some accessions of *O. longistaminata*. We have extended our study to other species in the Poaceae. The results suggest that the gain of the intron, most likely by insertion of a retroposon, took place in the ancestral genome of rice after its evolutionary divergence from other ancestral cereals such as barley, wheat and oat.

Key words Catalase · Rice · Gene structure · Evolution · Transposon

Communicated by P.M.A. Tigerstedt

M. Iwamoto · H. Higo · K. Higo (✉)
Department of Genetic Resources,
National Institute of Agrobiological Resources,
Tsukuba, Ibaraki 305, Japan
Fax: +81-298-38-7408
e-mail: kenhigo@abr.affrc.go.jp

M. Maekawa
Research Institute for Bioresources,
Okayama University, Kurashiki, Okayama 710, Japan

A. Saito
Department of Crop Improvement,
Kyushu National Agricultural Experimental Station,
Kikuchi, Kumamoto 861-11, Japan

Introduction

Catalase (H_2O_2:H_2O_2 oxidoreductase, EC 1.11.1.6; CAT) catalyzes the dismutation of hydrogen peroxide into oxygen and water, and is found in a wide range of organisms from aerobic bacteria to higher plants and higher animals. Plant catalases are expected to play an important role as one of the antioxidant defense genes that respond to environmental as well as physiological oxidative stress (Scandalios 1990; Scandalios 1994). Plant catalase subunits are encoded by a small gene family, usually three or at most four isozyme genes in one species. In maize (*Zea mays*) (Scandalios 1990; Scandalios 1994) and tobacco (*Nicotiana plumbaginifolia*) (Willekens et al. 1994a,b, 1995), each catalase isozyme gene shows a fairly complicated spatial and temporal pattern of expression throughout the life cycle of the plants. In view of findings indicating the possible roles of hydrogen peroxide in salicylic acid-mediated systemic acquired resistance in plants (Klessig and Malamy 1994), in elicitor-induced plant defense reactions (Mehdy 1994) and also in response to low-temperature stress (Anderson et al. 1994; Prasad et al. 1994a,b), interest in plant catalase has increased considerably in recent years.

So far 17 cDNAs from 11 plant species have been cloned, and their sequences determined (Willekens et al.

1995; Higo and Higo 1996; Guan and Scandalios 1996; and literature cited therein). For a comparison of results obtained from experiments using various plant species, establishment of the correlation of each catalase isozyme and/or its gene is essential. For this purpose, either the amino acid sequences, the cDNA sequences, or both, have been compared among them to draw phylogenetic trees of catalase or its gene (von Ossowski et al. 1993; Willekens et al. 1994a; Frugoli et al. 1996; Guan and Scandalios 1996). However, because of the wide diversity of plant catalase sequences, some discrepancies in the isozyme correlation were recognized among the phylogenetic trees previously reported. Furthermore, these phylogenetic trees do not give information about how these diverse isozymes were formed during evolution.

In this paper, we first compare the positions of introns in 12 plant catalase genes whose sequences are presently available. It appears that the positions of introns are rather restricted and that the introns have generally been lost during the course of evolution. Their hypothetical pathways are in general agreement with the phylogenetic trees of catalase genes, especially those of monocots, deduced previously from the amino acid or nucleotide sequences. Using the polymerase chain reaction (PCR), we then screened 25 species of the Poaceae (Grass) family, as well as wild rice, for the presence of an intron found in one of the rice catalase genes. Some of the PCR products were sequenced for further characterization.

Materials and methods

DNA sequences of plant catalase genes

Catalase genomic sequences of castor bean (*Ricinus communis*), *Arabidopsis thaliana*, soybean (*Glycine max*), potato (*Solanum tuberosum*), maize (*Zea mays*) and rice (*Oryza sativa*) were retrieved from DDBJ/EMBL/GenBank nucleotide sequence databases and the literature: castor bean *Cat1* and *Cat2* (accession number D21161 and D21162, respectively; Suzuki et al. 1994), *Arabidopsis Cat3* (U43147; Zhong and McClung 1996), *Arabidopsis Cat* gene (X94447), soybean *Cat* gene (Z12021), potato *Cat2St* (Z37106; Niebel et al. 1995), maize *Cat1* (X60135; Guan and Scandalios 1993), maize *Cat2* (Z54358; Guan et al. 1996), maize *Cat3* (L05934; Abler and Scandalios 1993) and rice *CatA* (D29966; Higo and Higo 1996). Isolation and characterization of the genomic clones of 2 catalase genes, *CatB* and *CatC*, from *Oryza sativa* will be described elsewhere (Higo et al. in preparation). The nucleotide sequences of *CatB* and *CatC* appear in DDBJ/EMBL/GenBank nucleotide sequence databases under the accession numbers D64013 and D86611, respectively.

Template DNA and PCR

Total DNAs as templates for PCR were isolated, as described previously (Murray and Thompson 1980), from leaves of *Oryza sativa* cv 'FL134' (japonica-type rice), *O. sativa* cv 'Kasalath' (indica-type rice), *Hordeum vulgare* (barley), *H. spontaneum* (wild barley), *Triticum aestivum* cv 'Chinese Spring' (wheat, hexaploid; AABBDD genome), *T. monococcum* (wheat, diploid; AA genome), *T. boeoticum* (wheat, diploid; AA genome), *Secale cereale* (rye), *Avena sativa* (oat), *Lolium multiflorum* (italian ryegrass), *L. perenne* (perenial ryegrass), *Festuca arundinacea* (tall fescue), *Pennisetum americanum* (pearl millet), *Panicum miliaceum* (broom-corn millet), *P. maximum* (guinea grass), *Echinochloa utilis* (japanese millet), *Brachiaria ruziziensis* (congo grass), *Arundinella hirta*, *Zea mays* (maize), *Saccharum officinarum* (sugarcane), *Sorghum bicolor* (sorghum), *Coix lacryma-jobi* var 'mayuen', *Arundo donax* (giant reed), *Phragmites australis* (common reed) and *Phyllostachys heterocycla* (bamboo). DNAs from *O. rufipogon* (Asian wild rice), *O. glaberrima* (African rice) and six accessions (C101214, C101229, C104075, C104977, C105198 and C105204) of *O. longistaminata* (African wild rice) were isolated essentially as described (Edwards et al. 1991) except that leaves in a microcentrifuge tube were ground with a pipette tip in liquid nitrogen. *O. longistaminata* accessions were originally obtained from International Rice Research Institute (IRRI), the Philippines. PCR amplification reaction mixtures contained 10 m*M* *Tris*-HCl (pH 8.8), 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.1% Triton X-100, 200 μ*M* of each dNTP, 0.2 μ*M* of each primer and 1.25 U of Gene Taq (Wako, Japan) in a total volume of 50 μl. PCR was done with primers designed to flank the intron-2 of rice *CatA* (Higo and Higo 1996). The first PCR was performed using 100 ng of genomic DNA as a template, and AU1 (5′-ACACCTACACCTTCGTCAC-3′; nucleotide 942–960 of *CatA*) and AR1 (5′-GTACGCGAACACCCTGCA-3′; nucleotide 1449–1432 of *CatA*, complementary sequence) as primers. The second PCR was done with 1 μl of a 100-fold dilution of the first PCR mixture, AU1 and AR2 (5′-GTTGCGGTTGAGAA CGAG-3′; nucleotide 1338–1321 of *CatA*, complementary sequence) primers. The sequence between the AU1 and AR2 primer is located within the region amplified by the first PCR using AU1 and AR1 primers. Amplifications were performed with a Perkin-Elmer/Cetus DNA thermal cycler model PJ1000NPT programmed for 30 cycles of 1 min at 94°C 2 min at 50°C 1 min at 70°C and ending with 5 min at 72°C. The amplified PCR products were fractionated both on 2% agarose gels and on 3% MetaPhor™ agarose gels (FMC BioProducts, ME) in 0.5 × TBE buffer.

DNA sequencing and phylogenetic analysis

PCR products of the accessions of *O. longistaminata* in which only one major PCR product was detected were directly sequenced with a DNA sequencer model 373A and an ABI PRISM™ dye terminator cycle sequencing ready reaction kit (Perkin Elmer, Calif.) with either AU1 or AR2 as a primer. PCR products of the accessions of *O. longistaminata* in which several PCR products were detected were cloned using a TA cloning kit (Invitrogen Corp, Calif.) and sequenced. Each PCR product was sequenced in both directions. All the sequences were deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases with accession numbers AB004768 (*O. longistaminata* accession: C101229; gene: *CatA1*), AB004769 (C101229; *CatA2*), AB004770 (C104977; *CatA1*), AB004771 (C104977; *CatA2*), AB004772 (C101214; *CatA*), AB004773 (C104075; *CatA*), AB004774 (C105198; *CatA*) and AB004775 (C105204; *CatA*).

Sequences obtained from *O. longistaminata* accessions were aligned with the "malign" multiple-sequence alignment program, then subjected to phylogenetic analysis by the maximum-parsimony method using a computer at DNA Data Bank of Japan, National Institute of Genetics, Japan.

Results

Exon-intron structures of plant catalase genes

There are 12 plant catalase genes whose genomic sequences have been determined from six species (rice,

maize, castor bean, *Arabidopsis*, potato and soybean). Exon-intron structures of 12 genes are graphically presented in Fig. 1. The sizes of the exons, with the exception of the first and the last ones, are generally conserved among plant genes. Although the first and the last exons differ in size among plant genes, the lengths of the amino acid coding regions within these two exons are highly conserved. The number of exons in the rice (Os) *CatB*/castor bean (Rc) *Cat1*/*Arabidopsis* (At) *Cat* gene/potato (St) *Cat2St* group is eight, the highest number among plant catalase genes identified to date. Although Os *CatC* and maize (Zm) *Cat2* have six exons, their exon-3 corresponds to exons-3, -4 and -5 of Os *CatB*/Rc *Cat1*/At *Cat* gene/St *Cat2St*. On the other hand, the exon-6 of Rc *Cat2*/At *Cat3* corresponds to exons-6 and -7 of the Os *CatB* group.

Comparison of the exon-intron structures of rice catalase genes with those of maize genes established a counterpart relationship; that is, Os *CatB*, *CatC* and *CatA* are structurally similar to Zm *Cat1*, *Cat2* and *Cat3*, respectively. This is in complete agreement with the isozyme correspondence based on similarities both in nucleotide and in deduced amino acid sequences. Os *CatA* and Zm *Cat3* are thought to be monocot-specific catalases (see Discussion). Os *CatA* includes, however, intron-2 (86 bp), which is not seen in Zm *Cat3* or in other genes.

Evolutional pathways of gene divergence inferred from exon-intron structures

Presumable pathways of the evolutional divergence of plant catalase genes are shown in Fig. 2. Since the gene structure consisting of eight exons and seven introns (i.e. that of Os *CatB*/Rc *Cat1*/At *Cat* gene/St *Cat2St*) is the only one which is shared in monocots and dicots, we assume that the primordial catalase gene in a common ancestor of monocots and dicots had eight exons (A–H in Fig. 2) and seven introns (1–7 in Fig. 2). We presume that the Os *CatB*, Rc *Cat1*, At *Cat* gene and St *Cat2St* still maintain the exon-intron structure of the primordial catalase gene. This hypothetical relationship is indicated by thick lines in Fig. 2. The other genes have fewer than seven introns and differ between monocots and dicots. Therefore, after the evolutionary divergence of monocots from dicots, consecutive duplication of the primordial gene followed by a differential loss of introns occurred independently in monocots and in dicots to form isozyme genes. In monocots, since the exon-intron structures of rice catalase genes are similar to those of their maize counterparts, three ancestral isozyme genes were probably formed before the divergence of ancestral maize from ancestral rice. Introns corresponding to introns-3 and -4 in Os *CatB* are not seen in either of the two rice genes (Os *CatA* and *CatC*) or in the two maize genes (Zm *Cat2* and *Cat3*). This suggests that these two introns have been lost in the common ancestor of Os *CatA*, Os *CatC*, Zm *Cat2* and Zm *Cat3* after the first duplication of the primordial catalase gene. The additional loss of introns-2, -5 and -6 is thought to have taken place to form the ancestor of Os *CatA*/Zm *Cat3*. Os *CatA*, however, requires a gain of an entirely new intron according to this pathway (see below). On the other hand, Zm *Cat1* probably lost the intron corresponding to intron-4 of Os *CatB* after the separation of rice and maize. In dicots, there are three groups according to the exon-intron structures of the genes identified to date. Both Rc *Cat2* and At *Cat3* are composed of seven exons and six introns. Since they have no intron corresponding to intron-6 of Rc *Cat1*/At *Cat* gene/St *Cat2St*, this intron seems to have been lost in the common ancestor of Rc *Cat2*/At *Cat3* before the separation of castor bean and *Arabidopsis*. The Gm *Cat* gene has lost intron-3 which is present in other dicot genes identified to date.

Intron corresponding to the *CatA* intron-2 in other rice species

While Fig. 2 shows the progressive loss of introns during evolution, Os *CatA* seems to have acquired an intron during evolution. To examine the presence of the intron corresponding to the *CatA* intron-2 in other rice species, we performed PCR analyses to amplify the intron-2 region. Primers used for the PCR amplification were designed to amplify only the fragment of the *CatA* homolog, not that of the *CatB* or *CatC* homolog. Agarose gel electrophoresis of the PCR products revealed one major band in *O. sativa* cv 'FL134' (japonica type), *O. sativa* cv 'Kasalath' (indica type), *O. glaberrima* (cultivated rice in Africa) and *O. rufipogon* (wild rice in Asia) (Fig. 3A, lanes 2). In *O. longistaminata* (C104977; wild rice in Africa), two major bands, one of which is longer than those in other rice species, were detected in the second PCR products (lane 2). The sizes of all the major bands suggest the existence of the intron.

We then performed PCR analyses of six accessions of *O. longistaminata* collected from various regions of Africa. When the PCR products of *O. longistaminata* C104977 were electrophoresed in MetaPhor™ agarose gel, an agarose gel with fine resolution capabilities, the lower major band of C104977 in Fig. 3A was separated into two (Fig. 3B, lane 2). The results of agarose gel electrophoresis indicated that these six accessions can be separated into three types: Type 1, one major band of about 400 bp (C101214, C104075, C105198, C105204); Type 2, four major bands between 600 and 750 bp (C101229); Type 3, three major bands between 380 and 450 bp (C104977) (Fig. 3B, lanes 2). Type 1 seems to be most common among *O. longistaminata* accessions. The sizes of the major bands suggest that all of them contain the intron. We also performed PCR analyses of several accessions of *O. rufipogon* and *O. glaberrima* besides those shown in Fig. 3A. Agarose gel

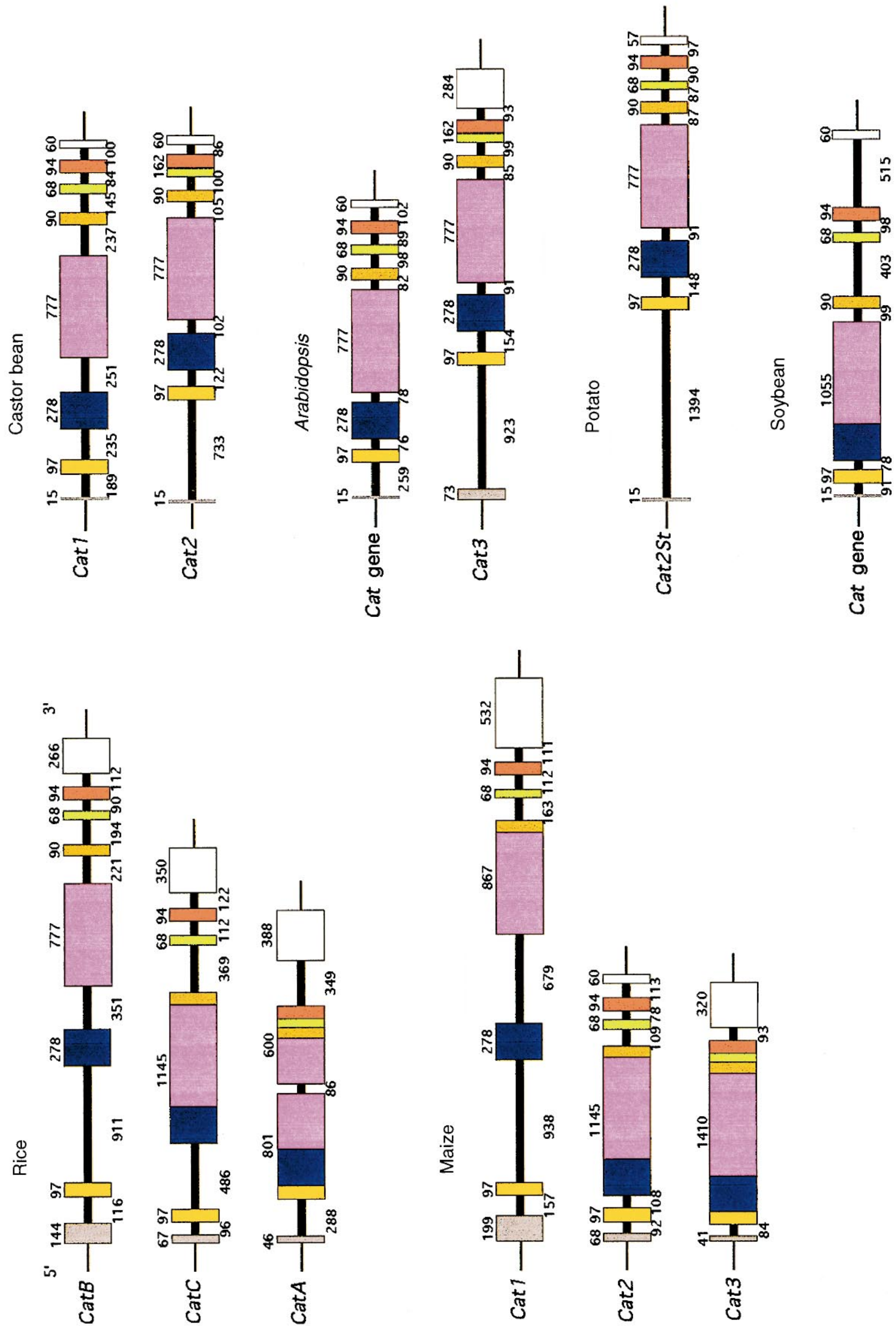


Fig. 1 Exon-intron structures of catalase genes in rice, maize, castor bean, *Arabidopsis*, potato and soybean. Exons are indicated by *colored boxes*, with the same color representing corresponding regions among genes. The *thick lines* between exons represent introns. The *numbers above colored boxes* and *under thick lines* are the sizes of exons and introns in base pairs, respectively

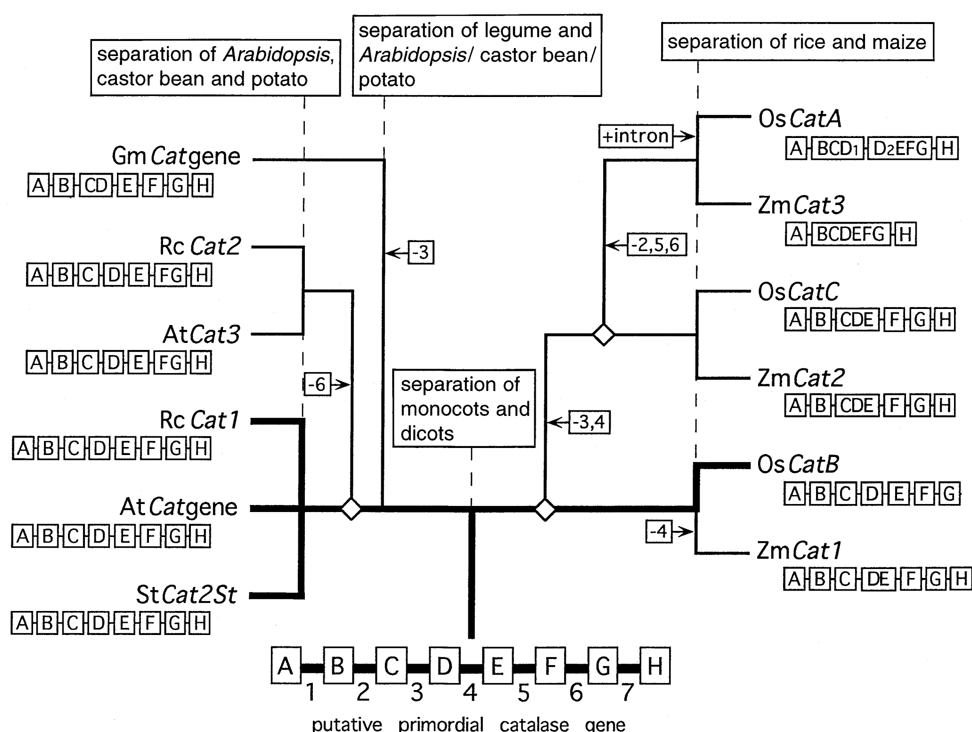


Fig. 2 Presumable pathways of the evolutional divergence of plant catalase genes among rice (*Os*), maize (*Zm*), soybean (*Gm*), castor bean (*Rc*), *Arabidopsis* (*At*) and potato (*St*) deduced from the exon-intron structures. The putative primordial catalase gene is shown at the *bottom*. Exons 1–8 are indicated by the *letters A–H*, respectively, and introns by *numbers 1–7*. *Opened lozenges* at the *intersection* of *lines* represent events in which gene duplication is thought to have occurred. The *numbers* with a *minus sign* in a *square* represent the name of the intron lost in the course of divergence. *Thick lines* connecting the Rc *Cat1*, At *Cat* gene, St *Cat2St*, Os *CatB* and the primordial gene indicate the close similarity in the exon-intron structure among them

electrophoresis of the PCR products revealed one major band, the size of which suggests the existence of the intron, in these accessions (data not shown), and there was no variation among the PCR products of the accessions of each species.

Nucleotide sequences of PCR products of *O. longistaminata*

To obtain information in detail about the structures of the introns of these *O. longistaminata* accessions, we cloned and sequenced the PCR products. Two longer products of C101229 (700 and 750 bp in length) and the 450-bp product of C104977 (see Fig. 3B, lanes 2) failed to clone into plasmid vectors. To distinguish the two cloned gene fragments in accessions C101229 and C104977, we named the shorter fragment *CatA1* and the longer fragment *CatA2*. The nucleotide sequences of eight PCR products were determined (Fig. 4A). The results showed that all the products contained, in addition to the intron region, parts of the exon regions which were nearly identical to the corresponding regions of the exon-2 and -3 of Os *CatA*. Therefore, all of the PCR products seem to have originated from the *CatA* homolog in the genome of each accession. The lengths of the introns are 69 bp (C104075 *CatA*, C104977 *CatA1*, C105198 *CatA*, C105204 *CatA*), 78 bp (C104977 *CatA2*), 82 bp (C101214 *CatA*), 289 bp (C101229 *CatA1*) and 331 bp (C101229 *CatA2*).

Recently, it has been found that the intron-2 of Os *CatA* contains a region similar to *p-SINE1* (H. Ohtsubo, personal communication). *p-SINE1* is a retroposon often seen in the rice genome (Umeda et al. 1991). The Os *CatA* intron-2 contains, however, only a part of the *p-SINE1* sequence: the flanking direct repeats called target site duplications (TSDs) and B-box are present, but the A-box is missing (Fig. 4A, B). The A-box and B-box are the promoter sequences for RNA polymerase III (Gali et al. 1981). TSD is believed to have been generated by duplication of the target site sequence when *p-SINE1* was inserted into the target sequence (Umeda et al. 1991). Furthermore, two products from C101229 (*CatA1* and *CatA2*) contained not only an additional *p-SINE1*-like sequence which contains a B-box, but also an AT-rich region (77% AT) and the inverted repeat region which are not found in the PCR products from other accessions (Fig. 4B). C101229 *CatA2* additionally contains tandem repeats of 41 bp which includes the B-box. The lengths of presumed TSDs in *O. longistaminata* accessions are 4 bp (C101229) and 8 bp (other accessions) (Fig. 4A). These TSDs are shorter than those of normal *p-SINE1*s

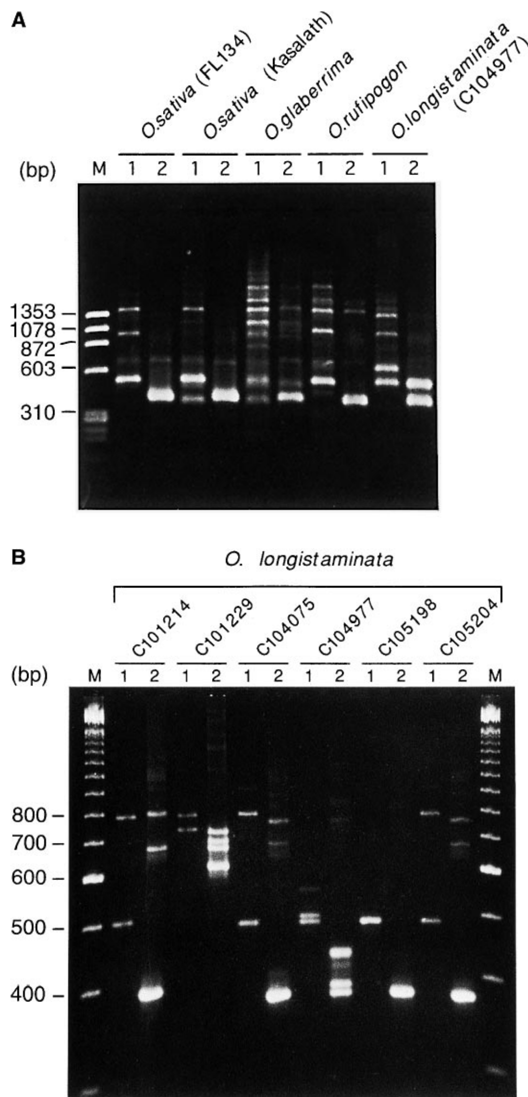


Fig. 3 A Agarose gel electrophoresis of PCR products of rice species. The PCR products of the first amplification with *O. sativa* ('FL134'), *O. sativa* ('Kasalath'), *O. glaberrima*, *O. rufipogon* and *O. longistaminata* (C104977) using AU1 and AR1 as primers were loaded in *lane 1*, and those of the second amplification using AU1 and AR2 as primers were loaded in *lane 2*. ϕX174 DNA/*Hae*III markers were loaded in *lane M*. **B** Agarose (MataPhor™) gel electrophoresis of PCR products of *O. longistaminata* accessions. The PCR products of the first amplification using AU1 and AR1 as primers were loaded in *lane 1*, and those of the second amplification using AU1 and AR2 as primers were loaded in *lane 2*. *lane M* 100-bp ladder markers

(about 14 bp; Motohashi et al. 1997). The TSD flanking *p-SINE1*-like sequence in Os *CatA* is also short (8 bp in length).

Phylogenetic tree of *O. longistaminata* accessions

A phylogenetic tree of the *O. longistaminata* accessions based on the nucleotide sequences of the PCR products was constructed (Fig. 5A). The accessions having a 69-bp intron (C104075 *CatA*, C105204 *CatA*, C105198 *CatA*, C104977 *CatA1*) formed a cluster, indicating a very close relationship among them. Although C104977 *CatA1* and *CatA2* were obtained from the same accession, C104977 *CatA2* did not belong to the same cluster as C104977 *CatA1*. Based on a comparison among the nucleotide sequences of *O. sativa* and *O. longistaminata* accessions, there are six positions where nucleotides are deleted or inserted (Fig. 5B). The phylogenetic tree also shows when nucleotides were deleted or inserted during the evolutionary divergence of *O. sativa* and *O. longistaminata*. Two C101229 sequences (*CatA1* and *CatA2*) have more complicated structures (see Fig. 4B), and were therefore excluded from the multiple alignment of the sequences used to construct the phylogenetic tree.

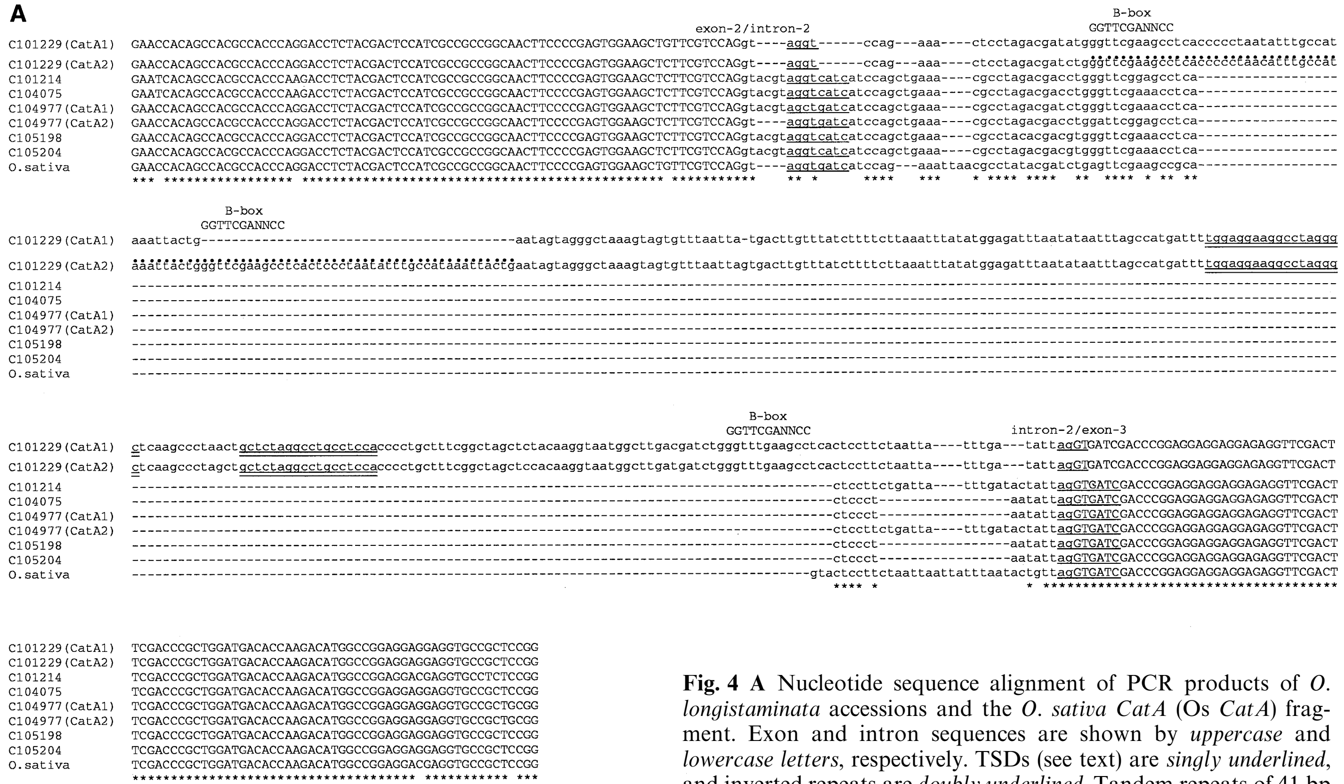
Intron corresponding to *CatA* intron-2 in grass species

We examined the presence of the intron corresponding to the intron-2 of Os *CatA* in other grass species. PCR analyses of grass species using the rice primers (see Materials and methods) showed that plants in the Pooideae subfamily had one major band, the size of which was equal to that of the *CatA* homolog without intron (Fig. 6A). However, no major band was obtained from other grasses including bamboo which had been reported to be closely related to rice (Tzvelev 1989). Although there was no major band in 10 species belonging to the Panicoideae subfamily such as maize, the maize *Cat3*, a counterpart of Os *CatA*, has no intron corresponding to the intron-2 of Os *CatA* (Fig. 1). Taken together, we hypothesize that the Os *CatA* intron-2 was gained in the ancestral genome of rice after evolutionary divergence from the ancestor of Pooideae such as barley, wheat and oat, based on the dendrogram by Watson et al. (1985) (Fig. 6B).

Discussion

Evolutionary divergence of plant catalase genes

Although the genetics, biochemistry and molecular biology of plant catalase genes have been investigated for some years (Scandalios 1990; 1994 for reviews), little is known about catalase gene evolution. In this paper, we compared the exon-intron structures of 12 catalase genes from six plant species. As the sizes of the exons are generally conserved, the formation of larger exons through the loss of introns could be traced. However, there are minor differences in the sizes of the exons of several genes compared to other genes. The pink-colored region of exon-3 in Os *CatA* (Fig. 1) is 3 bp shorter than corresponding regions of other genes. On the other hand, each of the pink- and brown-colored



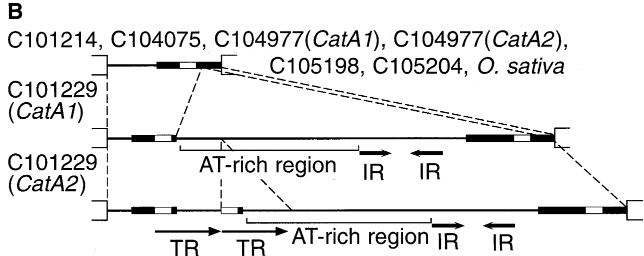


Fig. 4 A Nucleotide sequence alignment of PCR products of *O. longistaminata* accessions and the *O. sativa CatA* (Os *CatA*) fragment. Exon and intron sequences are shown by *uppercase* and *lowercase letters*, respectively. TSDs (see text) are *singly underlined*, and inverted repeats are *doubly underlined*. Tandem repeats of 41 bp in C101229 *CatA2* are indicated by *dots*. Gaps, represented by *dashes*, are introduced to maximize the alignment. Conserved nucleotides among sequences are indicated by *asterisks* at the *bottom*. B-box consensus sequences (see text) are shown *above* the alignment. **B** Structures of the introns in *O. longistaminata* accessions and *O. sativa*. *Opened thick bars* indicate B-boxes. Regions that showed similarities to the *p-SINE1* family are indicated by *closed thick bars*. Two *unidirectional thin arrows with TR* represent tandem repeats, and *thick arrows with IR* represent inverted repeats. AT-rich regions are indicated by *brackets*. The similar regions among accessions are linked with *dotted lines*

regions in exon-2 of Zm *Cat3* is 3 bp longer than those of other genes and, therefore, the length of exon-2 (1410 bp) is 6 bp longer than the total of exons-2, -3, -4, -5 and -6 of Zm *Cat1* (1404 bp). Thus, Os CAT-A catalase lacks one amino acid residue and Zm CAT-3 has two additional residues compared with other plant catalases. In Zm *Cat2*, the blue-colored region and the pink-colored region of exon-3 are 3 bp shorter and longer, respectively, than the corresponding regions of other genes. Therefore, the length of the amino acid sequence of Zm CAT-2 is identical to those of other catalases. Because these nucleotide deletion and insertion events occurred within the exon – not at the boundary between the exon and intron – the possibility of intron sliding can be excluded.

We then hypothesized the pathways of plant catalase gene divergence (Fig. 2). Because the gene containing seven introns is present in both monocots and dicots (i.e. Os *CatB*, Rc *Cat1*, At *Cat* gene and St *Cat2St*), we assume that this structure is close to the primitive form of the catalase gene and that the primordial catalase gene in a common ancestor of monocots and dicots had the same structure. Catalase isozyme genes are thought to have been formed by the duplication of an ancestral gene followed by the loss of introns. A similar suggestion, based on the comparison of the intron position of catalase genes, was recently made by Guan and Scandalios (1996). There are three nuclear-encoded isozyme genes in rice and maize, and counterpart relationships based on the deduced amino acid sequence are: Os *CatB*-Zm *Cat1*, Os *CatC*-Zm *Cat2* and Os *CatA*-Zm *Cat3*. This suggests that two duplications of the ancestral catalase gene and the formation of three isozyme genes occurred before the separation of rice and maize. Along with the formation of three structurally diverse isozyme genes, the functions of the catalase gene were also diversified. Os *CatA*, Zm *Cat3* and barley (Hv) *Cat2* (Skadsen et al. 1995) show a high similarity both in amino acid and in cDNA sequences. This type of catalase appears to be monocot-specific

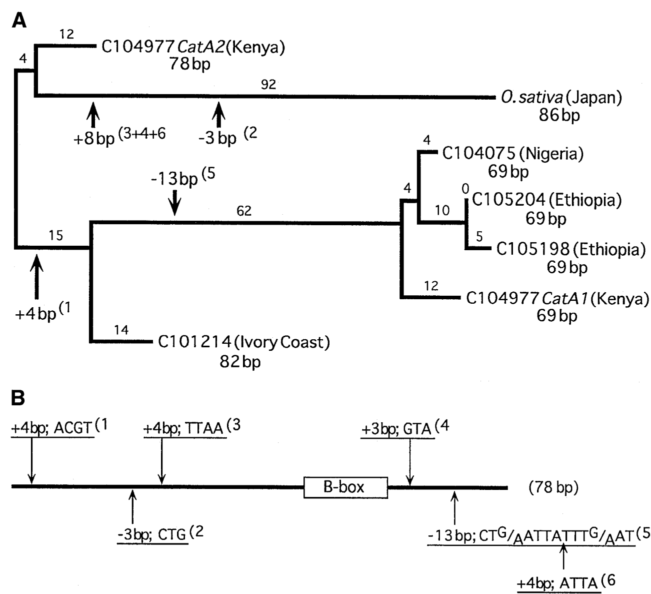


Fig. 5 A Unrooted phylogenetic tree, constructed by the parsimony method, based on the nucleotide sequences of *O. sativa CatA* (Os *CatA*) and PCR products of *O. longistaminata* accessions. Countries where the accessions have been isolated are indicated in *parentheses*. Size of the intron in base pairs is shown *below* the *accession number*. Events of presumed addition (+) or deletion (−) of nucleotides during evolution are indicated with *arrows*, the location in the intron (Fig. 5B) being indicated with *superscripts*. Branch lengths are shown *above* the *branches*. **B** Locations in the intron of presumed addition or deletion of nucleotides (number 1–6) shown in Fig. 5A during evolution

because no counterpart has been found in dicots. These three catalases lack a SRL motif (Ser-Arg-Leu), a putative peroxisomal localization signal near the carboxyl terminus that is found in most known plant catalases (Willekens et al. 1995; Frugoli et al. 1996; Guan and Scandalios 1996; and literature cited therein). Zm CAT-3 has been reported to be associated with mitochondria (Scandalios 1990, 1994). Therefore, we presume that Os CAT-A and Hv CAT-2 are also associated with mitochondria.

Possible mechanism of the gain of intron-2 in rice catalase genes

In the rice varieties investigated, no rice was found without the intron corresponding to the Os *CatA* intron-2 (Fig. 3A, B), and all of the introns contained the *p-SINE1*-like sequence (Fig. 4A, B). Because maize (Fig. 1) and the Pooideae species (Fig. 6A, B) do not have the intron, we presume that the insertion event took place after the separation of rice from maize and the Pooideae species. The similarity between the nucleotide sequence of *CatA* intron-2 and that of *p-SINE1* (H. Ohtsubo, personal communication) suggests either that the intron was formed by the insertion of a retroposon similar to *p-SINE1* into an exon or that a retroposon was inserted into a pre-existing intron. Since TSDs, which are generally found in both the 5′- and 3′-flanking regions of *p-SINE1*, are seen at both ends of intron-2 of the rice varieties (Fig. 4A), intron-2

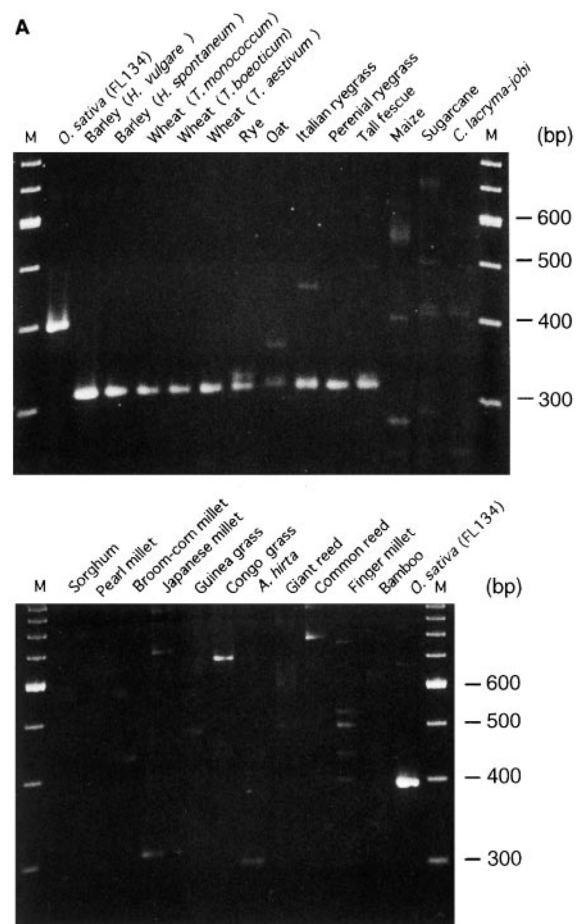


Fig. 6 A Agarose gel electrophoresis of PCR products of 25 grass species. PCR products of the second amplification were loaded in each lane. *Lane M* 100-bp ladder markers. **B** Presence (*circle*) or absence (*triangle*) of the intron in Os *CatA* and *CatA* homologs of 25 grass species. Because of the lack of the PCR products, no conclusion could be drawn for the other plants shown here. Dendrogram of grass genera in the Poaceae is according to Watson et al. (1985)

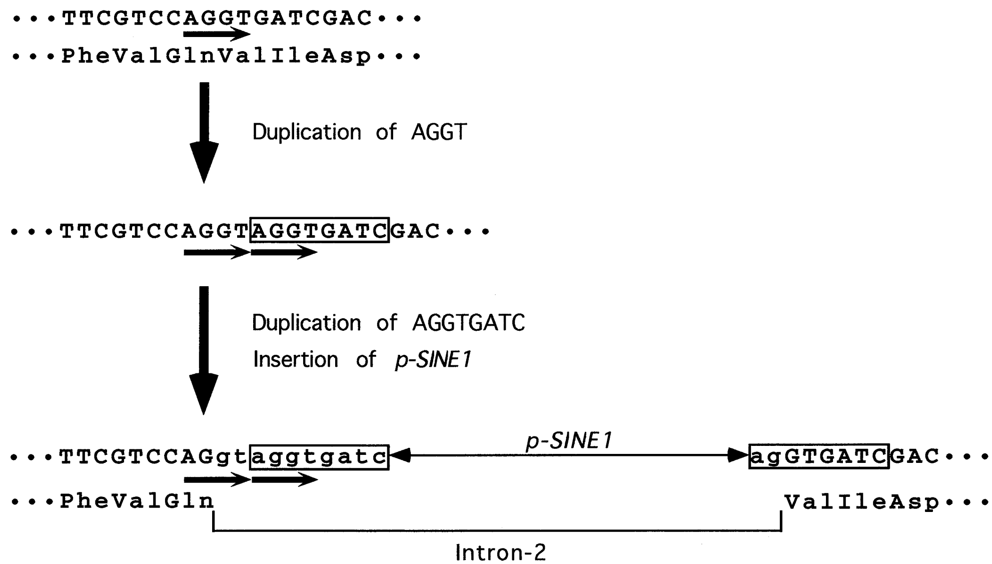


Fig. 7 Possible insertion mechanism of the intron-2 in Os *CatA* and *CatA* homologs of *O. longistaminata* accessions. AGGT sequences are indicated by *arrows*, and the sequence in a *square* represents TSD. Nucleotide sequence, which is deleted as an intron, is shown by *lowercase letters*

seems to be derived from the insertion of *p-SINE1* into an exon. Comparison of the sequences of Os *CatA* and their counterparts in *O. longistaminata* accessions (all of them contain the intron) with those of Zm *Cat3* and Hv *Cat2* (neither contain the intron) suggests a possible insertion mechanism (Fig. 7) in which tandem duplication of the AGGT sequence in the exon occurred, followed by duplication of the target site sequence (AGGTGATC), and *p-SINE1* was then inserted between the target site sequences. This two-step duplication resulted in the formation of the splicing site at both ends of the *p-SINE1* element. Thus, the *p-SINE1* element can be removed as an intron after transcription, and functional catalase was produced in spite of insertion of the *p-SINE1* element into an exon. We also detected a similar two-step duplication in a rice sequence previously reported as *p-SINE1*-r7 (Mochizuki et al. 1992). In this sequence, AGGATT was first duplicated in tandem, then GGATTCAGT was duplicated at both ends of the *p-SINE1*-r7 sequence (AGGATTAGGATTCAGT–*p-SINE1*-r7–GGATTCAGT). In C101214 *CatA*, C104075 *CatA*, C104977 *CatA1*, C105198 *CatA* and C105204 *CatA*, ACGT were inserted later between the tandem AGGT sequences (see the nucleotide sequences around the exon-2/intron-2 junction in Fig. 4A). Because C104977 *CatA1* contains AGCT instead of AGGT near the exon-2/intron-2 junction, AGCT is thought to have been generated by nucleotide substitution, from G to C, at the third position of AGGT. Because intron-2 has a structural defect as *p-SINE1* due to the lack of an A-box, one of two conserved promoter elements for RNA polymerase III, we suppose that this region no longer functions as a retroposon.

Evolutionary divergence of *O. longistaminata* accessions

Sequence analyses of the PCR products were very useful for the classification of *O. longistaminata* accessions. In the phylogenetic tree (Fig. 5A), C105204 and C105198 are the closest among the accessions examined. It has been reported that both C105204 and C105198 have the ability to mate with *O. sativa* in spite of the generally strong reproductive isolation between *O. sativa* and *O. longistaminata* (Maekawa et al. 1996). All the rice species we examined have the AA genome (Vaughan 1994). While the usefulness of *p-SINE1* members for classification of various rice strains with the AA genome had been reported previously (Mochizuki et al. 1993), almost all the *p-SINE1* sequences reported had not been correlated to specific genes. To our knowledge, this is the first report on a phylogenetic tree based on the nucleotide sequences of the introns containing the *p-SINE1* element in presumably functional catalase genes.

The introns of all of the *O. longistaminata* accessions examined lack an A-box and have short TSDs. Motohashi et al. (1997) have also found that two *p-SINE1* members (*p-SINE1*-r7 and -r31) lack an A-box and have either short TSDs (r7; 9 bp in length) or no recognizable TSD (r31). These results suggest that the 5′-region of *p-SINE1* where an A-box and 5′-TSD sequence are located tends to be deleted. The TSDs of two C101229 sequences are 4 bp long (Fig. 4A). Because the alignment in Fig. 4A suggests that the nucleotides corresponding to a part of the 5′-TSD sequence in other accessions are deleted in two C101229 sequences,

the original TSDs in C101229 might have been longer than those currently observed.

The two similar but not identical sequences (*CatA1* and *CatA2*) obtained from C101229 and C104977 (Fig. 4A) suggest the presence of at least two *CatA* homologs in each of these accessions. This implies either that one of these *CatA* homologs survived in *O. sativa*, and also in other accessions of *O. longistaminata*, or that duplication of the *CatA* homolog occurred in C101229 and C104977. There is no evidence to support either one of those speculations at present.

Two C101229 sequences (*CatA1* and *CatA2*) are more similar to each other than to sequences of other accessions, and there are five nucleotide substitutions and two insertions or deletions (1 and 41 bp in length) between the two sequences (Fig. 4A). Because the 41-bp sequence (TR in Fig. 4B) is almost identical to the adjacent 41-bp sequence, this sequence must have been generated by duplicating the sequence tandem in ancestral C101229 *CatA2* after the divergence of C101229 *CatA1* and *CatA2*. Tandem repeats are also seen in seven *p-SINE1* members (Motohashi et al. 1997) and one of the TS (Tobacco SINE) families in tomato (Yoshioka et al. 1993). Although two C101229 sequences have another *p-SINE1* at the 3′ end region of the intron, there is no recognizable TSD at either end of the *p-SINE1*.

A comparison of the nucleotide sequences among *O. sativa* and *O. longistaminata* accessions revealed five nucleotide substitutions in exons-2 and -3 (Fig. 4A). All of the nucleotide substitutions occurred at the third position of the codons, giving rise to either no amino acid substitution or the substitution of a chemically similar one.

CatA homologs in grasses

PCR analyses of grass species using the rice primers suggest that all of the species in the Pooideae examined have the *CatA* homologs. The cDNA sequence of Hv *Cat2*, the barley counterpart of Os *CatA*, had already been reported (Skadsen et al. 1995), but the genomic sequence for Hv *Cat2* was not available, so that it was unknown whether the intron corresponding to the Os *CatA* intron-2 was present in Hv *Cat2*. PCR analysis with the rice primers revealed the absence of the intron in Hv *Cat2* (Fig. 6A). The maize counterpart, *Cat3*, on the other hand, did not amplify any fragments with the rice primers. The nucleotide sequences of three annealing sites for the AU1, AR1 and AR2 primers in maize and barley DNAs have 78.9%, 100%, 66.7% and 94.7%, 89.5%, 88.9% identities to those of corresponding rice DNA, respectively. The rice annealing site sequences are more similar on a whole to the barley sequences than to those of maize. Moreover, the 3′ end of AU1 cannot anneal with the corresponding maize sequence. Annealing of the 3′ end of primers with template DNA is essential in order to amplify fragments. More suitable primers, other than the rice primers used here, would be required for accurate PCR analyses of *CatA* homologs of plants distantly related to rice.

Acknowledgments We are grateful to Dr. H. Ohtsubo of University of Tokyo for providing us with the information about retroposon in the rice catalase gene. This work was supported by a grant from the Ministry of Agriculture, Forestry and Fisheries, Japan.

References

Abler ML, Scandalios JG (1993) Isolation and characterization of a genomic sequence encoding the maize *Cat3* catalase gene. Plant Mol Biol 22:1031–1038

Anderson MD, Prasad TK, Martin BA, Stewart CR (1994) Differential gene expression in chilling-acclimated maize seedlings and evidence for the involvement of abscisic acid in chilling tolerance. Plant Physiol 105:331–339

Chen Z, Iyer S, Caplan A, Klessig DF, Fan B (1997) Differential accumulation of salicylic acid and salicylic acid-sensitive catalase in different rice tissues. Plant Physiol 114:193–201

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res 19:1349

Frugoli JA, Zhong HH, Nucccio ML, McCourt P, McPeek MA, Thomas TL, McClung CR (1996) Catalase is encoded by a multigene family in *Arabidopsis thaliana* (L.) Heynh. Plant Physiol 112:327–336

Gali G, Hofstetter H, Birnstiel ML (1981) Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature 294:626–631

Guan L, Scandalios JG (1993) Characterization of the catalase antioxidant defense gene *Cat1* of maize, and its developmentally regulated expression in transgenic tobacco. Plant J 3: 527–536

Guan L, Scandalios JG (1996) Molecular evolution of maize catalases and their relationship to other eukaryotic and prokaryotic catalases. J Mol Evol 42:570–579

Guan L, Polidoros AN, Scandalios JG (1996) Isolation, characterization and expression of the maize *Cat2* catalase gene. Plant Mol Biol 30:913–924

Higo K, Higo H (1996) Cloning and characterization of the rice *CatA* catalase gene, a homologue of the maize *Cat3* gene. Plant Mol Biol 30:505–521

Klessig DF, Malamy J (1994) The salicylic acid signals in plants. Plant Mol Biol 26:1439–1458

Maekawa M, Rikiishi K, Matsuura K, Noda K (1996) Genic analysis of rhizomatous trait of wild species (*Oryza longistaminata*) in rice (in Japanese). Breed Sci 46 [Suppl 1]:323

Mehdy MC (1994) Active oxygen species in plant defense against pathogens. Plant Physiol 105:467–472

Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E (1992) Characterization of a plant SINE, *p-SINE1*, in rice genomes. Jpn J Genet 57:155–166

Mochizuki K, Ohtsubo H, Hirano H-Y, Sano Y, Ohtsubo E (1993) Classification and relationships of rice strains with AA genome by identification of transposable elements at nine loci. Jpn J Genet 68:205–217

Motohashi R, Mochizuki K, Ohtsubo H, Ohtsubo E (1997) Structures and distribution of *p-SINE1* members in rice genomes. Theor Appl Genet 95:359–368

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Niebel A, Heungens K, Barthels N, Inze D, Van Montagu M, Gheysen G (1995) Characterization of a pathogen-induced potato catalase and its systemic expression upon nematode and bacterial infection. Mol Plant-Microbe Interac 8: 371–378

Prasad TK, Anderson MD, Martin BA, Stewart CR (1994a) Evidence for chilling-induced oxidative stress in maize seedlings and a regulatory role for hydrogen peroxide. Plant Cell 6 : 65–74

Prasad TK, Anderson MD, Stewart CR (1994b) Acclimation, hydrogen peroxide, and abscisic acid protect mitochondria against irreversible chilling injury in maize seedlings. Plant Physiol 105 : 619–627

Scandalios JG (1990) Response of plant antioxidant defense genes to environmental stress. Adv Genet 28 : 1–41

Scandalios JG (1994) Regulation and properties of plant catalases. In: Foyer CH, Mullineaux PM (eds) Causes of photoxidative stress and amelioration of defense systems in plants. CRC Press, Boca Raton, Fla., pp 275–315

Skadsen RW, Schulze-Lefert P, Herbst JM (1995) Molecular cloning, characterization and expression analysis of two catalase isozyme genes in barley. Plant Mol Biol 29 : 1005–1014

Suzuki M, Ario T, Hattori T, Nakamura K, Asahi T (1994) Isolation and characterization of two tightly linked catalase genes from castor bean that are differentially regulated. Plant Mol Biol 25 : 507–516

Tzvelev NN (1989) The system of grasses (Poaceae) and their evolution. Bot Rev 55 : 141–204

Umeda M, Ohtsubo H, Ohtsubo E (1991) Diversification of the rice *Waxy* gene by insertion of mobile DNA elements into introns. Jpn J Genet 66 : 569–586

Vaughan DA (1994) The wild relatives of rice: a genetic resources handbook. International Rice Research Institute, Manila, Philippines

von Ossowski I, Hausner G, Loewen PC (1993) Molecular evolutionary analysis based on the amino acid sequence of catalase. J Mol Evol 37 : 71–76

Watson L, Clifford HT, Dallwitz MJ (1985) The classification of Poaceae: subfamilies and supertribes. Aust J Bot 33 : 433–484

Willekens H, Langebartels C, Tir C, Van Montagu M, Inz D, Van Camp W (1994a) Differential expression of catalase genes in *Nicotiana plumbaginifolia* (L.). Proc Natl Acad Sci USA 91 : 10450–10454

Willekens H, Villarroel R, Van Montagu M, Inz D, Van Camp W (1994b) Molecular identification of catalases from *Nicotiana plumbaginifolia* (L.). FEBS Lett 352 : 79–83

Willekens H, Inze D, Van Montagu M, van Camp W (1995) Catalases in plants. Mol Breed 1 : 207–228

Yoshioka Y, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc Natl Acad Sci USA 90 : 6562–6566

Zhong HH, McClung CR (1996) The circadian clock gates expression of two *Arabidopsis* catalase genes to distinct and opposite circadian phases. Mol Gen Genet 251 : 196–203