Theor Appl Genet (1999) 98: 853–861

M. Iwamoto · H. Nagashima · T. Nagamine
H. Higo · K. Higo

p-SINE1-like intron of the *CatA* catalase homologs and phylogenetic relationships among AA-genome *Oryza* and related species

Received: 29 August 1998 / Accepted: 2 November 1998

Abstract Intron-2 of the *Oryza sativa CatA* catalase gene is similar in nucleotide sequence to *p-SINE1*, a retroposon, and seems to have been added to the ancestral genome of rice. To examine when the *p-SINE1*-like intron was inserted into *CatA* during the evolutionary divergence of *Oryza* species, and to elucidate the evolutionary relationships among *Oryza* species using the sequence of the intron as a marker, we performed polymerase chain reaction (PCR) analyses of 32 accessions of 17 *Oryza* species with various genome types. Agarose-gel electrophoresis of the PCR products revealed that all the *Oryza* species with an AA genome have the *CatA* homolog with the intron, whereas other *Oryza* species have the *CatA* homolog without the intron. These results indicate that intron-2 of *CatA* is a good marker for distinguishing species with an AA genome among *Oryza* species. Sequencing of the PCR products showed that all the introns are similar to *p-SINE1*, though with slight variations in length. We also performed PCR analyses using four accessions of three species in genera related to *Oryza*, and found that there is an intron in the *CatA* homolog of *Leersia perrieri*. On the other hand, the *CatA* homolog of *Porteresia coarctata* has no intron. Sequence data showed that the *L. perrieri* homolog has a *p-SINE1*-like intron similar to that in *Oryza* species with an AA genome. These results suggest that the *p-SINE1*-like intron was already present in the common ancestor of *Oryza* and *L. perrieri* and was then lost in the ancestors of *P. coarctata* and of the *Oryza* species other than those with an AA genome. The phylogenetic tree of *Oryza* species with an AA genome based on the nucleotide sequences of the introns leads us to propose that *Oryza* species with an AA genome evolved from an ancestor of *Oryza longistaminata*.

Key words Catalase · *Oryza* · Rice · Evolution · *p-SINE1*

Communicated by P. M. A. Tigerstedt

M. Iwamoto · H. Nagashima[1] · T. Nagamine
H. Higo · K. Higo (✉)
Department of Genetic Resources, National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305-8602, Japan
Fax: +81-298-38-7408
E-mail: kenhigo@abr.affrc.go.jp

Present address:
[1] Department of Ecological Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi 441-8580, Japan

Introduction

Catalase ($H_2O:H_2O$ oxidoreductase, EC 1.11.1.6; CAT) catalyzes the dismutation of hydrogen peroxide, which is known as a mediator of oxidative damage, into oxygen and water. In *Oryza sativa*, there are three catalase genes, as is the case with the other higher plants so far examined. It is assumed that the three isozyme genes of catalase were generated by consecutive duplication of a primordial catalase gene during evolution (Iwamoto et al. 1998). One of the catalase genes in *O. sativa*, *CatA* (Higo and Higo 1996), has been identified as a rice homolog of the maize *Cat3* gene (Abler and Scandalios 1993). Intron-2 of *O. sativa CatA* is unique in that it is not observed in maize *Cat3*, and shows sequence similarity with a retroposon, *p-SINE1*, which is often seen in the rice genome (Umeda et al. 1991). We previously postulated that the intron was formed by the insertion of a *p-SINE1*-like retroposon into the exon region of *CatA* (Iwamoto et al. 1998).

In our earlier report, we proposed that the intron was gained by the ancestral genome of rice after evolutionary divergence from the ancestor of plants belonging to the Pooideae, which includes cereals such as barley, wheat and oat (Iwamoto et al. 1998). A similar hypothesis was recently presented by Frugoli et al. (1998). The genus *Oryza* consists of species with various genome types, such as AA, BB, CC, EE, FF, BBCC and

Table 1 List of species in *Oryza* and rice-related species used for experiments[a]

Complex Species	Chromosome number	Genome group	Geographic distribution: variety/accession
O. sativa complex			
O. sativa	24	AA	Asian cultivated rice: FL134, Kasalath, Canabongbong
O. glaberrima	24	AA	African cultivated rice: C8581, W0025
O. barthii	24	AA	Central and South Africa: W1599
O. nivara	24	AA	Southeast and South Asia: 106148
O. rufipogon	24	AA	Southeast and South Asia, Australia, Central and South America: W0120, W0157, W1944
O. longistaminata	24	AA	Central and South Africa: C101214, C101229, C104075, C104977, C105198, C105204
O. meridionalis	24	AA	Australia: W2078, W2099
O. officinalis complex			
O. officinalis	24, 48	CC, BBCC	Southeast and South Asia: W0002, W0614, NIAS WBC6
O. punctata	24, 48	BB, BBCC	Central and South Africa: W0015, W1564
O. eichingeri	24	CC	East and Central Africa: W0043, W1525
O. australiensis	24	EE	Australia: NIAS WE1
O. minuta	48	BBCC	Philippines: W1213, W1337
O. alta	48	CCDD	Central and South America: W0018
O. grandiglumis	48	CCDD	South America: W0613, W1480B
O. latifolia	48	CCDD	Central and South America: W0048, W1181
O. meyeriana complex			
O. meyeriana	24	GG	Southeast and South Asia: W0022, W0067
O. ridleyi complex			
O. ridleyi	48	HHJJ	Southeast and South Asia: W0001, W0604
O. longiglumis	48	HHJJ	New Guinea: W1220, W1228
Species in genera related to *Oryza*			
Leersia tisseranti	24	Unknown	Africa: W1345, W1620
L. perrieri	24	Unknown	Madagascar: W1529
Porteresia coarctata	24	Unknown	South Asia: W0551

[a] Updated from Vaughan (1994)

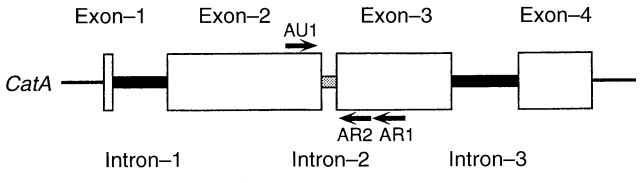
CCDD. *Oryza* species can be taxonomically divided into four complexes: the *Oryza sativa* complex, the *Oryza officinalis* complex, the *Oryza meyeriana* complex and the *Oryza ridleyi* complex (Vaughan 1989; see Table 1). All *Oryza* species with an AA genome belong to the *O. sativa* complex. The *O. officinalis* complex consists of species with BB, CC, EE, FF, BBCC and CCDD genomes. Although the genome types of the species belonging to the *O. meyeriana* and *O. ridleyi* complexes are unknown, Aggarwal et al. (1997) proposed new genomic designations for them, GG for the *O. meyeriana* complex and HHJJ for the *O. ridleyi* complex, based on the results of total genomic DNA hybridization. Since all the rice accessions we analyzed in our previous study have an AA genome, we were not sure whether the intron is also present in the *CatA* homologs of other *Oryza* species with an AA genome, and also in those of *Oryza* species with genomes other than AA.

In the present study, in order to estimate when the *p-SINE1*-like intron was inserted in the *CatA* homolog during the course of evolution, and to elucidate the evolutionary relationships among *Oryza* species using the sequence of the intron as a marker, we performed polymerase chain reaction (PCR) analyses, and sequenced the PCR products which appeared to have the intron. PCR analyses to examine species in genera related to *Oryza* for the *p-SINE1*-like intron were also performed, and the phylogenetic relationships of the *CatA* homologs in *Oryza* species were analyzed.

Materials and methods

Plant materials, template DNA and PCR

Oryza species and species in related genera that were used in this study are listed in Table 1. Both the diploid and the tetraploid rice accessions are included in *O. officinalis* and *Oryza punctata*. In *O. officinalis*, the accessions W0002/W0614 and NIAS WBC6 have the CC and BBCC genomes, respectively. The tetraploid race of *O. officinalis* is also known as *Oryza malampuzhaensis* (Vaughan 1994). *O. punctata* accessions W0015 and W1564 have the BB and BBCC genomes, respectively. Although *Oryza rufipogon* is widely distributed in Southeast and South Asia, Australia, Central and South America (Table 1), we analyzed three accessions collected in Asia for these experiments: W0120 and W0157 from India, and W1944 from China. *O. sativa* cv. FL134 (japonica-type rice), cv. Kasalath (indica-type rice) and cv. Canabongbong (javanica-type rice) were provided by Dr. A. Saito, Kyushu National Agricultural Experimental Station, Japan, and *Oryza glaberrima* (accessions C8581 and W0025), *Oryza barthii* (W1599) and *O. rufipogon* (W1944) by Dr. M. Maekawa, Okayama University. Other *Oryza* species accessions and the four accessions in genera related to *Oryza* are maintained in a domed greenhouse of the National Institute of Agrobiological Resources, Japan. *Oryza nivara* (106148) was originally obtained from the International Rice Research Institute (IRRI), the




```
AU1: 5'-ACACCTACACCTTCGTCAC-3'  (19 mer)
AR1: 5'-GTACGCGAACACCCTGCA-3'   (18 mer)
AR2: 5'-GTTGCGGTTGAGAACGAG-3'   (18 mer)
```

Fig. 1 Exon-intron structure of *CatA*. Exons and introns are indicated by *boxes* and *thick lines*, respectively. The *p-SINE1*-like intron is shown by the *shaded thick line*. Primers (AU1, AR1 and AR2) used for PCR amplification are indicated by *arrows*. Nucleotide sequences of the primers are shown below

Philippines, except for three cultivars of *O. sativa*, *O. officinalis* (NIAS WBC6) and *Oryza australiensis* (NIAS WE1) and all other species and accessions were from the National Institute of Genetics, Japan. Total DNAs used as templates for PCR were isolated from leaves, as previously described (Edwards et al. 1991; Iwamoto et al. 1998). PCR-amplification was performed according to the method of Iwamoto et al. (1998) using AU1 and AR1 primers for the first PCR and then AU1 and AR2 for the second PCR (see Fig. 1).

DNA sequencing and phylogenetic analysis

All PCR products with the intron were directly sequenced with a model 373 A DNA sequencer and an ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer, Calif.), with either AU1 or AR2 (see Fig. 1) as a primer. PCR products of *O. rufipogon* W0120 were cloned using a TA Cloning Kit (Invitrogen, Calif.) and then sequenced. Each PCR product was sequenced in both directions. All the sequences were deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases with DNA accession numbers AB014441 (species, rice accession and gene: *sativa* FL134 *CatA*), AB014442 (*sativa* Kasalath *CatA*), AB014443 (*sativa* Canabongbong *CatA*), AB014444 (*rufipogon* W0120 *CatA1*), AB014445 (*rufipogon* W0120 *CatA2*), AB014446 (*rufipogon* W0157 *CatA*), AB014447 (*rufipogon* W1944 *CatA*), AB014448 (*nivara* 106148 *CatA*), AB014449 (*meridionalis* W2078 *CatA*), AB014450 (*meridionalis* W2099 *CatA*), AB014451 (*barthii* W1599 *CatA*), AB014452 (*glaberrima* C8581 *CatA*), AB014453 (*glaberrima* W0025 *CatA*), AB014454 (*perrieri* W1529 *CatA*) and AB014455 (*coarctata* W0551 *CatA*). The intron sequences of AB004768 (*longistaminata* C101229 *CatA1*), AB004769 (C101229 *CatA2*), AB004770 (C104977 *CatA1*), AB004771 (C104977 *CatA2*), AB004772 (C101214 *CatA*), AB004773 (C104075 *CatA*), AB004774 (C105198 *CatA*) and AB004775 (C105204 *CatA*) were reported previously (Iwamoto et al. 1998).

Sequences of the PCR products with the intron were aligned with the "malign" multiple-sequence alignment program, and then subjected to phylogenetic analysis by the maximum-parsimony method using a computer at the DNA Data Bank of Japan, National Institute of Genetics, Japan.

Results and discussion

Distribution of the *p-SINE1*-like intron among *Oryza* species

To examine whether *CatA* homologs in *Oryza* species other than those examined previously (Iwamoto et al. 1998) have the *p-SINE1*-like intron, we performed PCR analyses of 32 accessions from 17 *Oryza* species. PCR analyses of *O. sativa* cv. FL134, *O. sativa* cv. Kasalath and six accessions of *Oryza longistaminata* (C101214, C101229, C104075, C104977, C105198 and C105204) were performed previously and showed the presence of the intron (Iwamoto et al. 1998). *O. longistaminata* is a wild species distributed in Africa (Vaughan 1994). Primers which were designed to amplify the fragment including intron-2 of *O. sativa* *CatA* were used for the PCR amplification (Fig. 1). Agarose-gel electrophoresis of the PCR products revealed that all AA-genome *Oryza* species had a major band with a size of about 400 bp, indicating the presence of the intron (Fig. 2). In *O. rufipogon* W0120 there are two major bands of about 400 and 450 bp. In the *Oryza* species with genomes other than the AA genome, one major band of about 300 bp, the size of which is equal to the PCR product of the *CatA* homolog without the intron, was detected irrespective of genome type. The allotetraploid *Oryza* species/accessions with a BBCC or a CCDD genome are generally thought to have more genetic variation than the diploid rice species with a BB or CC genome. Agarose-gel electrophoresis indicated that they contain only the *CatA* homolog without the intron, as do the diploid rice species. Agarose-gel electrophoresis of the PCR products of the *O. meyeriana*, *O. ridleyi* and *Oryza longiglumis* accessions showed the absence of the intron in the *CatA* homolog.

Nucleotide sequences of PCR products of *Oryza* species and accessions with an AA genome

To obtain detailed information about the structure of the introns, we sequenced the PCR products of *Oryza* species and accessions with an AA genome. The nucleotide sequences of 13 PCR products were determined (Fig. 3). The results show that all the products contain the exon regions at both ends, with sequences almost identical to the corresponding regions of exons-2 and -3 of *CatA*. Although there are several positions where nucleotide substitutions occurred among accessions, no changes in amino-acid sequences were inferred from the nucleotide sequences (data not shown). All of the introns of the *Oryza* species examined are similar to *p-SINE1* and contain 8-bp flanking direct repeats, called target-site duplications (TSDs), and the B-box (Fig. 3). However, all of them lack an A-box, as is the case with the intron sequences of *CatA* and its homologs in the *O. longistaminata* accessions, as reported previously (Iwamoto et al. 1998). The A-box and B-box are the promoter sequences for RNA polymerase III (Gali et al. 1981). In the *CatA* homologs of Asian and Oceanian wild *Oryza* species, except in *rufipogon* W0120 *CatA2*, part of the B-box is deleted.

When the 400-bp product of *O. rufipogon* W0120 was cloned for DNA sequencing, it was found that the

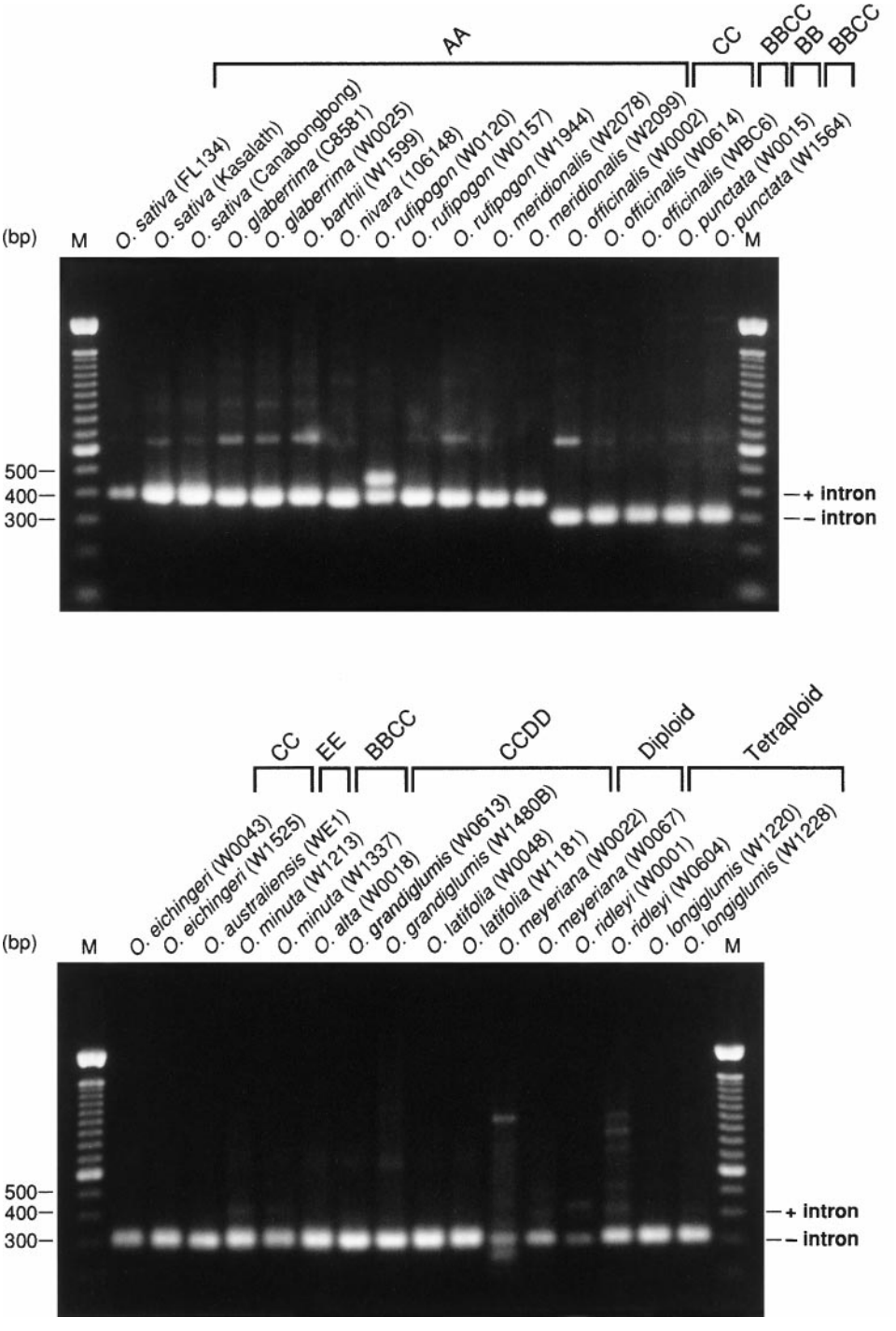


Fig. 2 Agarose gel electrophoresis of the PCR products of *Oryza* species and accessions. The PCR products of the second amplification using AU1 and AR2 as primers were loaded in each lane. *Lane M* 100-bp-ladder markers. Bands with a size of about 400 bp indicate the presence of the intron (+ *intron*), while bands of about 300 bp indicate its absence (− *intron*). Genome types are shown *above* the *brackets*

400-bp band consisted of two kinds of PCR products. For distinguishing the two products of *O. rufipogon* W0120, the product containing the shorter intron (74 bp) was named *CatA1* and that containing the longer one (78 bp) *CatA2* (Fig. 3). The lengths of the introns vary slightly: 70 bp (*rufipogon* W1944 *CatA*, *nivara* 106148 *CatA*, *meridionalis* W2078 *CatA*, *meridionalis* W2099 *CatA*), 74 bp (*rufipogon* W0157 *CatA*, *rufipogon* W0120 *CatA1*), 78 bp (*rufipogon* W0120 *CatA2*, *barthii* W1599 *CatA*, *glaberrima* C8581 *CatA*, *glaberrima* W0025 *CatA*), and 86 bp (*sativa* FL134 *CatA*, *sativa* Kasalath *CatA*, *sativa* Canabongbong *CatA*). The 450-bp PCR product of *O. rufipogon* W0120 (Fig. 2) could not be cloned into plasmid vectors. However, when the 450-bp product excised from an agarose gel was sequenced, two sequences corresponding to those of *rufipogon* W0120 *CatA1* and *CatA2* were detected. Therefore, we conclude that the 450-bp product was derived from mis-annealing between single DNA strands of *CatA1* and *CatA2*.

Comparison of the intron sequences revealed nucleotide deletions and insertions (regions 1–10 in Fig. 3).

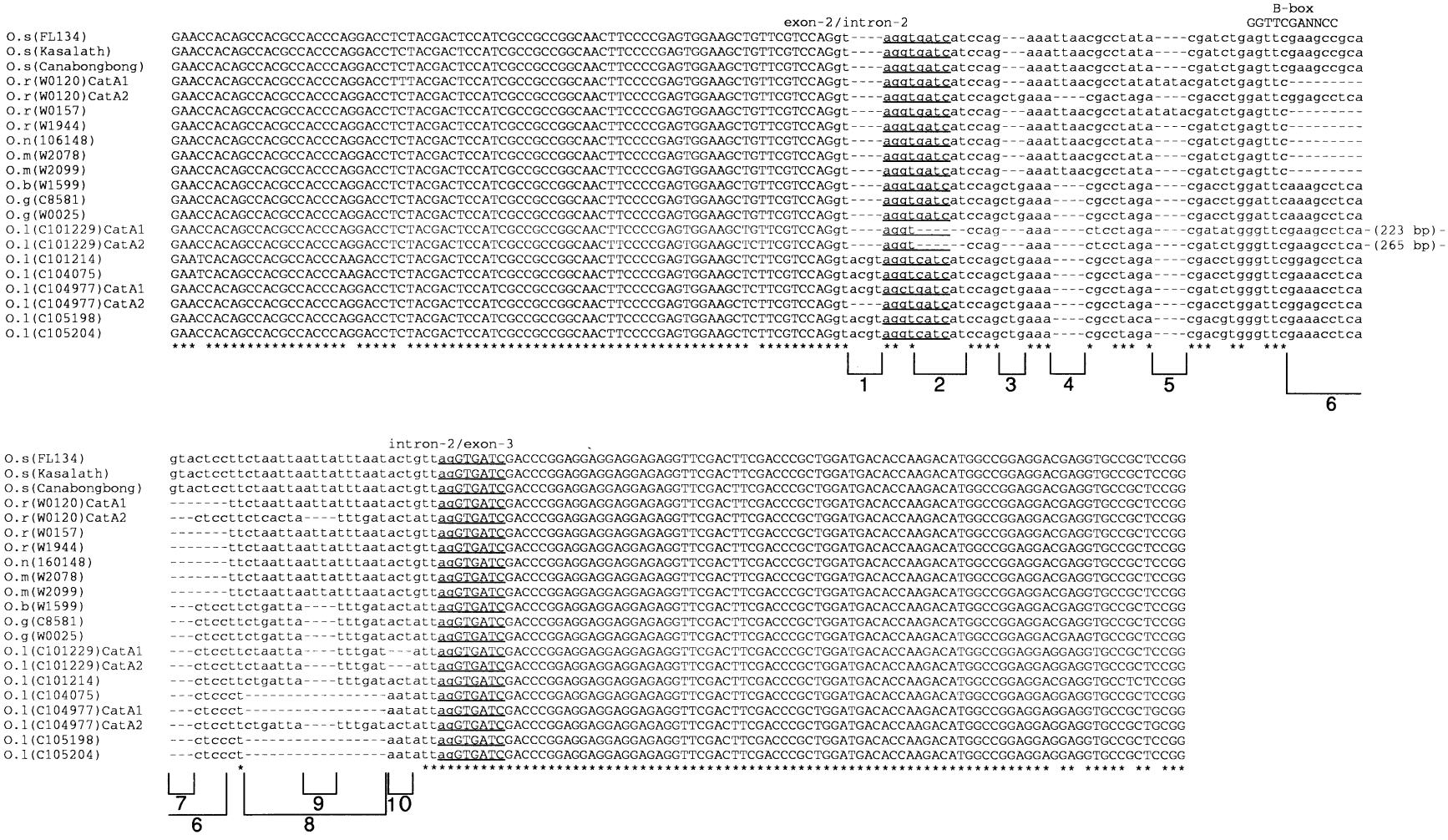


Fig. 3 Nucleotide sequence alignment of the PCR products of the *Oryza* species and accessions with an AA genome. Exon and intron sequences are shown by *uppercase* and *lowercase letters*, respectively. TSDs (see text) are indicated by *underlines*. Gaps, represented by *dashes*, were introduced to maximize the alignment. Conserved nucleotides among sequences are indicated by *asterisks* at the bottom. The B-box consensus sequence (see text) is shown *above* the alignment. The regions 1–10 *below* the alignment show insertions/deletions of nucleotides in the introns. The nucleotide sequences of the *O. longistaminata* accessions (C101229, C101214, C104075, C104977, C105198 and C105204) were taken from Iwamoto et al. (1998)

Some of these are restricted to one or some *Oryza* species, with the exception of *rufipogon* W0120 *CatA2*. Thus (1) a 13-bp sequence (consensus sequence: GAAGCCTCACTCC) is not seen in Asian and Oceanian wild *Oryza* species such as *O. rufipogon*, *O. nivara* and *Oryza meridionalis* (region 6 in Fig. 3), (2) the CTG sequence is seen only in African *Oryza* species such as *O. longistaminata*, *O. barthii* and *O. glaberrima* (region 3), (3) both the TTAA (region 4) and the ATTA (region 9) are observed not only in Asian and Oceanian wild *Oryza* species but also in three varieties of Asian cultivated rice, *O. sativa*, (4) *O. sativa* varieties contain the sequence GTA (region 7), and (5) the ACGT sequence is present in five of eight *CatA* homologs in six *O. longistaminata* accessions (region 1) (Fig. 3). These oligonucleotide sequences could be used for identifying *Oryza* species. In the Asian wild *Oryza* species, *O. rufipogon* W0120, one of two homologs (*rufipogon* W0120 *CatA2*) is more similar to the *CatA* homologs of the African *Oryza* species than to those of the Asian species, while the other homolog (*rufipogon* W0120 *CatA1*) has an intron similar to those of other Asian and Oceanian wild *Oryza* species. There are two possible explanations for this: (1) *O. rufipogon* W0120 originally contained both the Asian/Oceanian-type and the African-type *CatA* homologs, or (2) *rufipogon* W0120 *CatA2* was incorporated into *O. rufipogon* W0120 by crossing with African *Oryza* species. We do not have any data to support either of these possibilities.

Although *O. sativa* cv. FL134, Kasalath, and Canabongbong are japonica-, indica-, and javanica-type Asian cultivated rice varieties, respectively, the sequences of the introns, as well as those of the flanking exon regions, are identical among these three varieties. The PCR product of African cultivated rice *O. glaberrima* C8581 has the same sequence as that of the African wild *Oryza* species *O. barthii* W1599, and the difference between the sequences of *O. glaberrima* W0025 and *O. barthii* W1599 involves only one nucleotide substitution in the exon region. These results support the possibility which has been suggested previously (Chang 1976; Watanabe 1997), that *O. glaberrima* arose from the ancestor of *O. barthii*.

Insertion of the intron during evolution

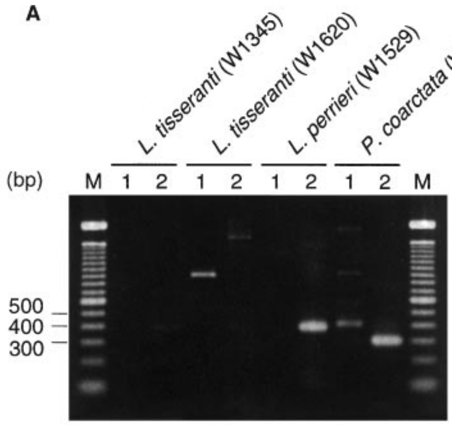
The results shown in Figs. 2 and 3 raised the question of whether the intron had been inserted only in the ancestor of the *Oryza* species with an AA genome, or whether the intron had already been acquired in the common ancestor of the genus *Oryza*, followed by excision of the intron except for those species with an AA genome. We asked, therefore, whether the intron corresponding to intron-2 of *CatA* is present in genera related to *Oryza*. Two accessions of *Leersia tisseranti* (W1345 and W1620), and one each of *Leersia perrieri* (W1529) and *Porteresia coarctata* (W0551) were used for PCR analysis. These species belong to the tribe *Oryzeae* and were previously classified within the genus *Oryza* (Tateoka 1963). There are 17 species, including *L. tisseranti* and *L. perrieri*, in the genus *Leersia* (Vaughan 1994). On the other hand, the genus *Porteresia* consists of only a single species. PCR analyses of these four accessions using the rice primers (Fig. 1) showed that *L. perrieri* W1529 and *P. coarctata* W0551 had one major band (Fig. 4 A, lanes 2), the size of which was equal to that of the *CatA* homolog with (*L. perrieri* W1529) or without (*P. coarctata* W0551) the intron. With the two accessions of *L. tisseranti*, no major band was detected after the second PCR amplification (lanes 2 in Fig. 4 A). These results suggest a divergence of the nucleotide sequence of the *CatA* homolog in *L. tisseranti*, indicating that further study is required for evaluating the presence or absence of the intron in this species.

Sequencing of the PCR products revealed that *L. perrieri* W1529 has an intron sequence similar to *p-SINE1*, and that *P. coarctata* W0551 contains no intron in the *CatA* homolog (Fig. 4 B). This suggests that the *p-SINE1*-like retroposon was inserted into the coding region of the *CatA* homolog of the common ancestor of *Oryza* and *Leersia* to form the intron, and that the intron was then lost in the ancestor(s) of *Oryza* species with genomes other than the AA genome (Fig. 5). According to the results of chloroplast-DNA restriction fragment analysis (Zhang and Second 1989), *Porteresia* seems to be more closely related to *Oryza* than to *Leersia*. Therefore, it is likely that the ancestor of *Porteresia* had the *p-SINE1*-like intron and then lost it after evolutionary divergence from *Oryza* (Fig. 5).

Evolutionary divergence of *Oryza* species with AA genomes

There have been several reports on the evolutionary relationships among *Oryza* species. These were based on analyses of the electrophoretic patterns of the large subunit of Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) (Ichikawa et al. 1986), restriction fragment length polymorphism (RFLP) of either genomic DNA (Jena and Kochert 1991; Wang et al. 1992) or chloroplast DNA (Ichikawa et al. 1986; Dally and Second 1990), and random amplified polymorphic DNA (RAPD) (Ishii et al. 1996). Dendrograms have been constructed to show phylogenetic relationships among the *Oryza* species with an AA genome (Dally and Second 1990; Wang et al. 1992; Ishii et al. 1996). However, no phylogenetic tree has previously been constructed using nucleotide sequences of *Oryza* species with the AA genome. Nucleotide sequence analysis is more sensitive to genetic change than RFLP or RAPD analysis. We therefore analyzed the nucleotide sequences of the *p-SINE1*-like intron found in

A



B

```
                                                                                                           exon-2/intron-2
O.sativa                     GAACCACAGCCACGCCACCCAGGACCTCTACGACTCCATCGCCGCCGGCAACTTCCCCGAGTGGAAGCTGTTCGTCCAGgt----aggtgatc
O.longistaminata(C101214)    GAATCACAGCCACGCCACCCAAGACCTCTACGACTCCATCGCCGCCGGCAACTTCCCCGAGTGGAAGCTCTTCGTCCAGgtacgtaggtcatc
L.perrieri(W1529)            GAATCGCAGCCACGCCACCCAAGACCTCTACGACTCCATCGCCGCCGGCAACTTCCCCGAGTGGAAGCTCTTCGTCCAGgtacgtaggtcatc
P.coarctata(W0551)           GAACCACAGCCACGCGACGCAGGACCTCTACGACTCCATCGCCGCCGGCAACTTCCCGGAGTGGAAGCTGTTCGTCCAG-------------
                             *** * ********* ** ** ************************************ *********** *********

                                                        B-box
                                                     GGTTCGANNCC                              intron-2/exon-3
O.sativa                     atccag---aaattaacgcctatacgatctgagttcgaagccgcagtactccttctaattaattatttaatactgttagGTGATCGACCCGGA
O.longistaminata(C101214)    atccagctgaaa----cgcctagacgacctgggttcggagcctca---ctccttctgatta----tttgatactattagGTGATCGACCCGGA
L.perrieri(W1529)            atccagctgaaa----cgcctagacgacctgggttcggagcctca---ctccttctgatta----tttgatactattagGTGATCGACCCGGA
P.coarctata(W0551)           ------------------------------------------------------------------------------GTGATCAACCCGGA
                                                                                                           ****** *******

O.sativa                     GGAGGAGGAGAGGTTCGACTTCGACCCGCTGGATGACACCAAGACATGGCCGGAGGACGAGGTGCCGCTCCGG
O.longistaminata(C101214)    GGAGGAGGAGAGGTTCGACTTCGACCCGCTGGATGACACCAAGACATGGCCGGAGGACGAGGTGCCTCTCCGG
L.perrieri(W1529)            GGAGGAGGAGAGGTTCGACTTCGACCCGCTGGATGACACCAAGACATGGCCGGAGGACGAGGTGCCTCTCCGG
P.coarctata(W0551)           GGAGGAGGAGCGGTTCGACTTCGACCCGCTGGACGACACCAAGACATGGCCGGAGGACGAGGTGCCGCTCCGC
                             ********** ********************** ******************************** *****
```

Fig. 4 **A** Agarose-gel electrophoresis of the PCR products of four accessions of species in genera related to *Oryza*. The PCR products of the first amplification with *L. tisseranti* W1345 and W1620, *L. perrieri* W1529 and *P. coarctata* W0551 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. **B** Nucleotide sequence alignment of the PCR products of *L. perrieri* W1529 and *P. coarctata* W0551 with the corresponding region of *O. sativa CatA* (Higo and Higo 1996) and the *CatA* homolog in *O. longistaminata* C101214 (Iwamoto et al. 1998). Exon and intron sequences are shown by *uppercase* and *lowercase letters*, respectively. TSDs (see text) are indicated by *underlines*. Gaps, represented by *dashes*, were introduced to maximize the alignment. Conserved nucleotides among the sequences are indicated by *asterisks* at the bottom. The B-box consensus sequence (see text) is shown *above* the alignment

homologs of the catalase gene *CatA*, in order to examine the evolutionary relationships among the AA-genome *Oryza* species. *p-SINE1* is one of the short interspersed repetitive elements (SINEs) which are distributed in the genomes of higher eukaryotes, and SINEs are good markers for examining evolutionary relationships (Okada 1991). Because the *p-SINE1*-like sequence is located between two conserved exons, exons-2 and -3, of *CatA*, we were able to amplify the intron sequences in the homologs in other *Oryza* species, and also in some species in genera related to *Oryza*, using primers with the flanking exons sequences, and then use them for estimating the evolutionary relationships among these species.

Interestingly, the *CatA* homolog (*perrieri* W1529 *CatA*) in a genus releted to *Oryza*, *L. perrieri* W1529, has an intron with the same sequence as that in *longistaminata* C101214 *CatA* (Fig. 4 B). Because the distribution of *O. longistaminata* C101214 (germ plasm collected in the Ivory Coast) is geographically distant from that of *L. perrieri* W1529 (Madagascar), it is not likely that *O. longistaminata* C101214 and *L. perrieri* W1529 crossed so as to have the same *CatA* homolog. Rather, *O. longistaminata* C101214 and *L. perrieri* W1529 seem to share the trait of the *CatA* homolog that existed in the genome of the common ancestor of *Oryza* and *Leersia*.

A phylogenetic tree of the AA-genome *Oryza* species was constructed based on the nucleotide sequences of the PCR products of the *CatA* homologs (Fig. 6).

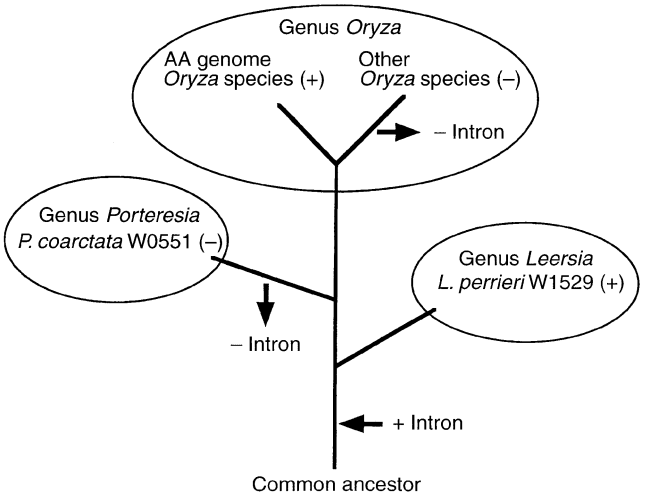


Fig. 5 Schematic representation of predicted gain and loss of the intron in the *CatA* homolog during evolutionary divergence of *Oryza* species, *L. perrieri* W1529 and *P. coarctata* W0551. Events of the presumed gain or loss of the intron are indicated by *arrows*. It is assumed that the intron was absent in the common ancestor. *CatA* homologs in AA-genome *Oryza* species and *L. perrieri* W1529 have the intron (+), whereas those in other *Oryza* species and *P. coarctata* W0551 do not (–)

Estimates of when the ten nucleotide sequences (regions 1–10 in Fig. 3) were deleted or inserted in the intron sequence during the divergence of the *Oryza* species are shown on the tree. Because *longistaminata* C101214 *CatA* seems to be the most ancestral of the *CatA* homologs among the *Oryza* species examined (see above), *longistaminata* C101214 *CatA* is placed near the root of the tree. In this tree the *CatA* homologs are roughly divided into two groups: those in the Asian/Oceanian *Oryza* species, with the others in the African species. Furthermore, the Asian/Oceanian *Oryza* group could itself be divided into two subgroups: the *CatA* homologs of three varieties of *O. sativa* form one subgroup, and those of *O. nivara*, *O. rufipogon* and *O. meridionalis*, all of which are wild *Oryza* species distributed in Asia and Australia, form the other subgroup. This suggests that, during the course of the evolution, the ancestor of the Asian/Oceanian *Oryza* species diverged from the ancestor of the African *Oryza* species

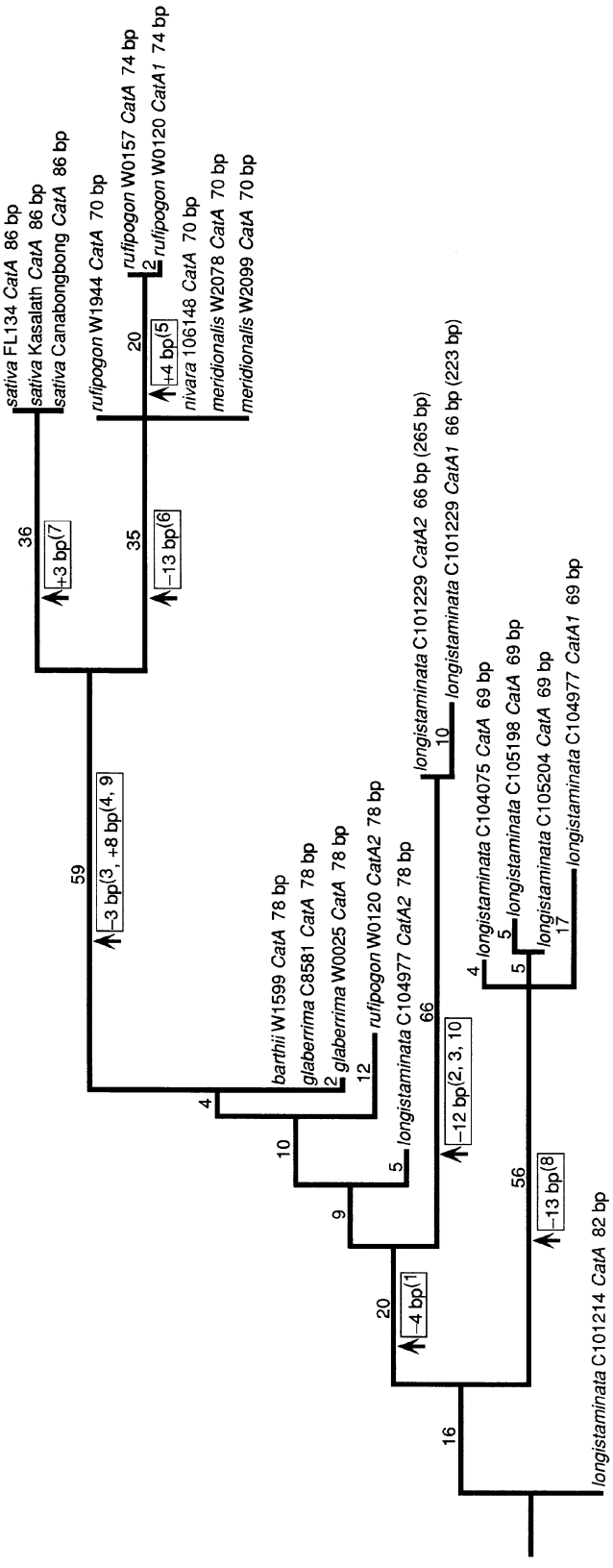


Fig. 6 Rooted phylogenetic tree, constructed by the parsimony method, based on the nucleotide sequences of the PCR products of the *CatA* homologs in the *Oryza* species and accessions with an AA genome. The *CatA* homolog of *O. longistaminata* C101214 (*longistaminata* C101214 *CatA*) is located near the origin of the AA-genome *Oryza* species (see text). The size of each intron is shown in base pairs. Events of presumed addition (+) or deletion (–) of nucleotides during evolution, shown within *open boxes*, are indicated by *arrows*, the location in the intron (regions 1–10 in Fig. 3) being indicated by *superscripts*. Branch lengths are shown *above* the *branches*. The nucleotide sequences of *longistaminata* C101229 *CatA1* and *CatA2* used for constructing the tree are restricted to the region corresponding to the sequences in other *Oryza* species and do not include the additional *p-SINE1*-like region, the AT-rich region and the inverted repeat region specifically seen in *longistaminata* C101229 *CatA1* and *CatA2* (Iwamoto et al. 1998). The full-length sizes of the introns in these homologs are shown in *parentheses*

and subsequently diverged into the ancestors of the Asian cultivated rice and the Asian/Oceanian wild *Oryza* species. Assuming that the phylogenetic tree in Fig. 6 reflects the true evolutionary history of the AA-genome *Oryza* species, we hypothesize that those species with an AA genome originated from the ancestor of *O. longistaminata* in Africa, then spread to other areas and differentiated into a number of species, some changing from perennial to annual.

The *CatA* homologs of *O. meridionalis* W2078 and W2099, *O. nivara* 106148 and *O. rufipogon* W1944 have the same intron sequence. However, these *Oryza* species grow in different areas: *O. meridionalis* W2078 and W2099 are found in Australia, and *O. nivara* 106148 and *O. rufipogon* W1944 in Asia. Therefore, we assume that there was a common ancestor of *O. meridionalis* W2078 and W2099, *O. nivara* 106148 and *O. rufipogon* W1944, and that the ancestor of *O. nivara* 106148 and *O. rufipogon* W1944 was then distributed in Asia, while that of *O. meridionalis* W2078 and W2099 was distributed in Australia.

One of the two *CatA* homologs in *O. longistaminata* C104977 (*longistaminata* C104977 *CatA2*) is closest to the homologs of *O. barthii* and *O. glaberrima* among the eight homologs of *O. longistaminata* accessions in the phylogenetic tree (Fig. 6). However, their life cycles are different; *O. longistaminata* is a perennial while *O. barthii* and *O. glaberrima* are both annuals. This suggests that the African perennial AA-genome *Oryza* species *O. longistaminata* C104977 may be evolutionarily close to the two annual AA-genome *Oryza* species in Africa.

As the *p-SINE1*-like intron is only seen in the AA-genome *Oryza* species, this intron may be useful as an AA genome-specific marker.

Acknowledgements We are grateful to Drs. M. Maekawa and A. Saito for providing the plant materials. We also thank Dr. D. A. Vaughan for critical reading of the manuscript. This work was supported by grants from the Ministry of Agriculture, Forestry and Fisheries, Japan. The experiments comply with the current laws of 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

Aggarwal RK, Brar DS, Khush GS (1997) Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. Mol Gen Genet 254:1–12

Chang TT (1976) The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25:425–441

Dally AM, Second G (1990) Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section Oryza). Cladistic-mutation and genetic-distance analysis. Theor Appl Genet 80:209–222

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, McPeek MA, Thomas TL, McClung CR (1998) Intron loss and gain during evolution of the catalase gene family in angiosperms. Genetics 149:355–365

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

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

Ichikawa H, Hirai A, Katayama T (1986) Genetic analyses of *Oryza* species by molecular markers for chloroplast genomes. Theor Appl Genet 72:353–358

Ishii T, Nakano T, Maeda H, Kamijima O (1996) Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. Genes Genet Syst 71:195–201

Iwamoto M, Maekawa M, Saito A, Higo H, Higo K (1998) Evolutionary relationship of plant catalase genes inferred from exon-intron structures: isozyme divergence after the separation of monocots and dicots. Theor Appl Genet 97:9–19

Jena KK, Kochert G (1991) Restriction fragment length polymorphism analysis of CCDD genome species of the genus *Oryza* L. Plant Mol Biol 16:831–839

Okada N (1991) SINEs. Curr Opin Genet Dev 1:498–504

Tateoka T (1963) Taxomic studies of *Oryza*. III. Key to the species and their enumeration. Bot Mag Tokyo 76:165–173

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 (1989) The genus *Oryza* L.: current status of taxonomy. IRRI Res Paper Ser 138

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

Wang ZY, Second G, Tanksley SD (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor Appl Genet 83:565–581

Watanabe Y (1997) Phylogeny and geographical distribution of genus *Oryza*. In: Matsuo T, Futsuhara Y, Kikuchi F, Yamaguchi H (eds) Science of the rice plant. Vol. 3. Genetics. Food and Agriculture Policy Research Center, Tokyo, Japan, pp 29–39

Zhang SH, Second G (1989) Phylogenetic analysis of the tribe Oryzeae: total chloroplast DNA restriction fragment analysis (a preliminary report). Rice Genet Newslett 6:76–80