Plant Syst. Evol. 249: 55–66 (2004)
DOI 10.1007/s00606-004-0173-8

Plant Systematics and Evolution
Printed in Austria

Origin and phylogeny of *Oryza* species with the CD genome based on multiple-gene sequence data

Y. Bao[1,2] and **S. Ge**[1]

[1]Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China
[2]Department of Biology, Qufu Normal University, Qufu, China

Received December 9, 2003; accepted February 26, 2004
Published online: October 11, 2004

Abstract. The CD genome species in the genus *Oryza* are endemic to Latin America, including *O. alta*, *O. grandiglumis* and *O. latifolia*. Origins and phylogenetic relationship of these species have long been in dispute and are still ambiguous due to their homogeneous genome type, similar morphological characteristics and overlapping distribution. In the present study, we sequenced two chloroplast fragments (*mat*K and *trn*L-*trn*F) and portions of three nuclear genes (*Adh*1, *Adh*2 and *GPA*1) from sixteen accessions representing seven species with the C, CD, and E genomes, as well as one G genome species as the outgroup. Phylogenetic analyses using parsimony and distance methods strongly supported that the CD genome originated from a single hybridization event, and that the C genome species (*O. officinalis* or *O. rhizomatis* instead of *O. eichingeri*) served as the maternal parent while the E genome species (*O. australiensis*) was the paternal donor during the formation of CD genome. In addition, the consistent phylogenetic relationships among the CCDD species indicated that significant divergence existed between *O. latifolia* and the other two (*O. alta* and *O. grandiglumis*), which corroborated the suggestion of treating the latter two as a single species or as taxa within species.

Key words: Origin, phylogeny, *Oryza*, CD genome, *mat*K, *trn*L-*trn*F, *Adh*, *GPA*1.

Polyploidy is an important mechanism of speciation in flowering plants, and approximately 70% of plant species have experienced one or more episodes of polyploidization in history (Stebbins 1950, Wendel 2000). Because of the potential to adapt to a wider range of habitats and survive better in unstable climates than their diploid progenitors, polyploid evolution has been a subject of intensive study for more than half a century (Wendel 2000). Most crop plants are of polyploid origins and the best studied polyploids include many of the world's leading crops such as cotton, wheat, soybean, peanut, banana, etc. (for reviews, see Gaut et al. 2000, Wendel 2000). The genus *Oryza* L., to which cultivated rice belongs, comprises approximately 24 species distributed throughout the world (Vaughan 1989, 1994). In this genus, more than one-third of the species are allotetraploids with different genome combinations, including the BC, CD, HJ and HK constitutions (Vaughan 1994, Ge et al. 2001b). Previous phylogenetic studies based on molecular data suggested that allotetraploids with different genome constitutions originated at different times and different places in history (Dally and Second 1990;

Wang et al. 1992; Ge et al. 1999, 2001b). For example, the BC genome species exhibit several independent origins, with their maternal parents being either the B or the C genome, while the CD genome species most likely originated from a single hybridization event (Ge et al. 1999).

Among the allotetraploids in the genus *Oryza*, the CD genome species (2n = 4x = 48) are endemic to Latin America, consisting of three species, *Oryza alta* Swallen, *O. latifolia* Desv. and *O. grandiglumis* (Doell) Prod. *Oryza latifolia* is widely distributed, occurring in Central and South America as well as the Caribbean islands, while *O. alta* and *O. grandiglumis* are found only in South America, primarily in the Amazon basin (Vaughan, 1989, 1994). Due to their homogeneous genome type, similar morphological characteristics, as well as overlapping distribution, the delimitation and phylogenetic relationships of the three species have long been controversial (Tateoka 1962, Nayar 1973, Jena and Kochert 1991, Aggarwal et al. 1996, Fukui et al. 1997, Buso et al. 2001). In addition, because diploid species with the C and D genomes have not been reported on the American continent, the debates over possible origins of these American tetraploids with the CD genome have been continued for decades (Nayar 1973, Wang et al. 1992, Fukui et al. 1997, Li et al. 2001, Federici et al. 2002). This issue is particularly challenging given the fact that no diploid with the D genome has been identified, despite worldwide efforts (Vaughan 1989, Fukui et al. 1997).

Recently, low-copy nuclear genes in combination with PCR-cloning have proven very effective for addressing allopolyploidization at the species level, because the homoeologous DNA regions or loci in allopolyploids can be easily identified and characterized, and their sequences may then be included in the phylogenetic analysis (Sang and Zhang 1999, Doyle et al. 2000, Baumel et al. 2002). In particular, combined analyses of biparentally inherited nuclear genes and maternally inherited regions such as chloroplast fragment in most flowering plants enable us to unravel polyploid speciation and to identify maternal parents (Small et al. 1998, Ge et al. 1999, Popp and Oxelman 2001). The alcohol dehydrogenase (*Adh*) gene is the most widely used low-copy nuclear gene, whereas chloroplast genes such as *mat*K gene, *trn*L intron and *trn*L-*trn*F spacer are widely used in phylogenetic studies (Soltis and Soltis 1998, Sang 2002). *GPA*1 encodes a G protein α subunit, and functions in various systems of signal transduction in diverse tissues or cells in flowering plants (Ma 1994, Fujisawa et al. 1999). Because of its features such as single copy in higher plants, and its well-characterized gene structure and chromosome location (Fujisawa et al. 1999), *GPA*1 is a potentially useful system for phylogenetic reconstruction. In the present study, we utilized two chloroplast fragments (*mat*K and *trn*L-*trn*F) and three nuclear gene fragments (*Adh*1, *Adh*2 and *GPA*1) to infer the phylogenetic relationships among the three allotetraploids. The specific goal of the study is to reveal the origin and the parental lineages of the CD genome species. Such information may contribute to a better delimitation of species and understanding of speciation and evolution in the genus *Oryza*.

Materials and methods

Plant materials. Sixteen accessions representing seven species with the C, CD and E genomes were sampled, including four accessions for *O. eichingeri* A. Peter (CC), one accession for *O. australiensis* Domin. (EE) and two accessions for each of two CC species (*O. officinalis* Wall ex Watt and *O. rhizomatis* Vaughan) and three CCDD species (Table 1). In addition, one accession of *O. granulata* Nees et Arn. ex Watt. (GG) was used as an outgroup because evidence showed that this species is the earliest diverging lineage in the genus *Oryza* (Wang et al. 1992, Ge et al. 1999). Seeds of these accessions were kindly provided by the International Rice Genebank at the International Rice Research Institute (Manila, Philippines). All seeds were firstly treated at 50 °C for five days to break dormancy. The dehulled seeds were suspended with a fungicidal liquid, washed thoroughly, and

Table 1. Plant materials used in the present study

Taxon	Genome	Acc. No.	Source	GenBank Accession No.				
				*mat*K	*trn*L-*trn*F	*Adh*1	*Adh*2	*GPA*1
O. officinalis	CC	105085	Philippines	AF148658	AF520764	AF148579	AF148613	AY188586
O. officinalis-1	CC	106159	Papua New Guinea			AY169473	AY169483	
O. eichingeri-s	CC	81803	Sri Lanka	AY176644	AF520766	AY169475	AY169487	AY188588
O. eichingeri-s1	CC	105407	Sri Lanka			AY169476	AY169486	
O. eichingeri-u	CC	101422	Uganda			AY169477	AY169485	AY188589
O. eichingeri-u1	CC	105159	Uganda			AY169478	AY169488	
O. rhizomatis	CC	105448	Sri Lanka	AF148660	AF520765	AF148580	AF148614	AY188587
O. rhizomatis-1	CC	103410	Sri Lanka			AY169474	AY169484	
O. alta	CCDD	105143	Guyana	AF148664	AF520768	AF148583	AF148617	AY188590
						AF148586	AF148620	AY188591
O. alta-1	CCDD	100161	Brazil			AY169479	AY169489	
						AY169480	AY169490	
O. grandiglumis	CCDD	105669	Brazil	AF148666	AF520767	AF148584	AF148619	AY188592
						AF148588	AF148622	AY188593
O. grandiglumis-1	CCDD	105664	Brazil					
O. latifolia	CCDD	105141	Costa Rica	AF148665	AF520769	AF148585	AF148618	AY188594
						AF148587	AF148621	AY188595
O. latifolia-1	CCDD	100914	Mexico			AY169481	AY169491	
						AY169482	AY169492	
O. australiensis	EE	105263	Australia	AF148667	AF520770	AF148589	AF148623	AY188596
O. granulata	GG	C0024	Hainan China	AF148674	AF520771	AF148597	AF148631	AY188597

immersed in warm water (30-35 °C) before germination. Two week-old seedlings were transplanted into pots in the greenhouse. Total DNA was isolated from fresh leaves of individual plants by the CTAB method (Doyle and Doyle 1987).

PCR amplification, cloning and sequencing. To amplify the chloroplast *mat*K and *trn*L intron and *trn*L-*trn*F spacer, and nuclear *Adh*1, *Adh*2 and *GPA*1, polymerase chain reactions (PCR) were performed in a 25-ul reaction with 10-20 ng template DNA, 10 mM Tris-HCl (pH8.3), 2.0 mM $MgCl_2$, 200 μM each dNTP, 5 μM each primer, and 0.75 units of *Taq* polymerase (Takaya). All amplifications were performed in a PTC-200 (PE) thermocycler.

Primers for amplifying *mat*K and *trn*L intron and *trn*L-*trn*F spacer followed those in Ge et al. (1999) and Taberlet et al. (1991), respectively. Amplification program includes 30 cycles of 1 min denaturation at 94 °C, 1 min annealing at 55 °C, and 1.5 min extension at 72 °C. Two *Adh* genes were amplified with primers AdhF1 and Adh1bR for *Adh*1, and AdhF1 and Adh2RR for *Adh*2, respectively (Ge et al. 1999). Amplification was accomplished using a program of 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 1.5 min, followed by a final 10-min extension at 72 °C. Nuclear gene *GPA*1 was amplified with primers GAP1FF and GAP1-14R. These two primers, which are located on exon 9 and exon 14 of the gene, respectively, were designed based on sequences that are conserved across three divergent genera of the grass family, *Oryza* (GenBank accession number L35844), *Hordeum* (AF267485) and *Zea* (AF055471). The PCR procedure includes 30 cycles of 1 min denaturation at 94 °C, 1 min annealing at 52 °C, and 1.5 min extension at 72 °C and a final 10-min extension at 72 °C . Two internal sequencing primers were also used for the *mat*K and *Adh* fragments. All of the primers used for amplification and sequencing in this study are listed in Table 2.

All PCR products were electrophoresed on 1.5% agarose gels. The amplified DNA fragments corresponding to the expected size were cut from the gel and purified with DNA Purification Kit (Pharmacia) according to the manufacturer's manual. Purified products of chloroplast *mat*K and *trn*L-*trn*F were sequenced directly. Nuclear *Adh* of diploid species was sequenced directly, while corresponding fragments of tetraploid species were inserted into pGEM-T-easy vectors (Promega) for cloning. Ge et al. (2001a) developed a rapid and reliable method to identify all of the 10 genomes in *Oryza*. Based on this method, two homeologous loci contributed by diploid parents could be distinguished efficiently from allotetraploid species by restriction enzyme *Eco*RI and *Afl*II, and then were sequenced separately.

Data analysis. Sequences were aligned with CLUSTAL W (Thompson et al. 1994) and refined manually. Phylogenetic analyses were conducted using the parsimony and distance methods as implemented in PAUP* version 4.0 (Swofford 1998). Maximum parsimony (MP) analyses were performed by heuristic search with MULPARS, tree bisection-reconnection (TBR) branch swapping, and RANDOM stepwise addition with 1000 replicates. The sequence data were also analyzed with a neighbor-joining (NJ) method using the Jukes-Cantor and Kimura two parameter distance estimates (Kimura 1980, Saitou and Nei 1987). Topological robustness was assessed by bootstrap analysis with 1000 replicates using simple taxon addition (Felsenstein 1985). Topological congruence of two cpDNA (*mat*K gene and *trn*L intron and *trn*F-*trn*L spacer) trees, two *Adh* (*Adh*1 and *Adh*2) trees, and *Adh* and *GPA*1 trees was accessed with partition-homogeneity test as implemented in PAUP (Johnson and Soltis 1998). Phylogenetic analyses combining two chloroplast regions and combining three nuclear genes were then performed, when no significant incongruence was found between them.

Results

Sequence characteristics. Approximately 4.2 kilobase pairs (kp) of nuclear DNA fragments from three different regions, and 2.4 kb of cpDNA from two regions were sequenced for at least one accession from eight species including three CC species, one EE species and three CCDD species, as well as one GG species as the outgroup. Each of the sampled sequences is characterized in Table 3. Phylogenetically informative characters were observed in all of the regions with the percentage varying from 0.4 % (*mat*K) to 9.5 % (*Adh*1).

Table 2. Primers used for PCR and sequencing

Primer	Sequence	Reference
*mat*K		Ge et al. 1999
matKF1	5′-TAATTAAGAGGATTCACCAG-3′	
matKF2	5′-ATTGCCTTTCCTTGATATCG-3′	
matKR2	5′-ACTACTCGAATTGGAATAG-3′	
maKR1	5′-ATGCAACACCCTGTTCTGAC-3′	
*trn*L-*trn*F		Taberlet et al. 1991
c	5′-CGAAATCGGTAGACGCTACG-3′	
f	5′-ATTTGAACTGGTGACACGAG-3′	
*Adh*1		Ge et al. 1999
AdhF1	5′-CACACCGACGTCTACTTCTG-3′	
AdhF2	5′-AGAGTGTTGGAGAGGGTGTGAC-3′	
Adh1R2	5′-ACTCACAGCAAGGCCTACAGC-3′	
Adh1bR	5′-TCAGCAAGTACCTAAATTATC-3′	
*Adh*2		Ge et al. 1999
AdhF1	5′-CACACCGACGTCTACTTCTG-3′	
AdhF2	5′-AGAGTGTTGGAGAGGGTGTGAC-3′	
Adh2R2	5′-ACAGCAAGGCCAACAGCTCC-3′	
Adh2RR	5′-CCACCGTTGGTCATCTCAAT-3′	
*GPA*1		This paper
GPA1FF	5′-GCAAGAGTACGGACAAATGGTG-3′	
GPA114R	5′-GCTTGCTGCTCTGGAAGTAG-3′	

Two distinct types of sequences were identified at each of two *Adh* and one *GPA*1 genes for all of CCDD species. One type is similar to the sequences of C genome diploids, and the other is similar to that of the E genome diploid. *Adh* genes sequenced include five introns and five exons. The aligned sequences of *Adh*1 were 1867 bp long, of which 195 nucleotide sites were variable and 178 were phylogenetically informative. The *Adh*2 data set contained 1666 nucleotide sites, of which 160 nucleotide sites were variable and 144 were phylogenetically informative. Approximately twice as much polymorphism was found in introns as in exons (data not shown). The percent of GC content is higher in exons (50-61.5%) than in

Table 3. Characteristics of chloroplast and nuclear DNA fragments sequenced

Sequence	Aligned length	Variable sites	Parsimony informative sites	% informative sites	% mean GC contents
CpDNA					
*mat*K	1552	21	6	0.4	33.5
*trn*L-*trn*F	892	37	7	0.8	34.4
Nuclear DNA					
*Adh*1	1867	195	178	9.5	41.3
*Adh*2	1666	160	144	8.6	43.9
*GPA*1	826	68	42	5.1	38.5

introns (28.2–49.5%) for both *Adh* genes. *GPA*1 genes sequenced include five introns and four exons. The aligned sequences of *GPA*1 were 826 bp long, of which 68 nucleotide sites were variable and 42 (5.1%) were phylogenetically informative (Table 3).

As expected (Table 3), cpDNA is less variable than nuclear low-copy genes. Of three single copy nuclear genes, two *Adh* genes provided significantly higher parsimony information than *GPA*1.

Phylogenetic analysis of the chloroplast *mat*K and *trn*L intron and *trn*L-*trn*F spacer. A single most parsimonious tree was obtained for both *mat*K (CI = 0.98, RI = 0.89) and *trn*L-*trn*F sequences (CI = 0.96, RI = 0.90) (Fig. 1A). Three CC species formed a monophyletic group sister to the clade consisting of three CCDD species. The EE species was at the basal position on both trees. By partition-homogeneity test, two data sets of *mat*K and *trn*T-*trn*F were statistically congruent ($p = 1.00$). Therefore, phylogenetic analysis based on the combined data set was conducted, which generated one most parsimonious tree with CI of 0.97 and RI of 0.90 (Fig. 1B). In the combined tree, the E genome species (*O. australiensis*) was at the basal position and the remaining species formed a highly supported clade (96% bootstrap). Within the clade, three CC and three CCDD species formed two monophyletic groups with the bootstrap values of 70% and 97%, respectively. For three CCDD species, *O. alta* and *O. grandiglumis* were more closely related.

Phylogenetic analysis of the *Adh* genes. Two most parsimonious trees for *Adh*1 and one most parsimonious tree for *Adh*2 were obtained in the study. Strict consensus trees of *Adh*1 (CI = 0.91 and RI = 0.95) and *Adh*2 (CI = 0.92 and RI = 0.95) were shown in

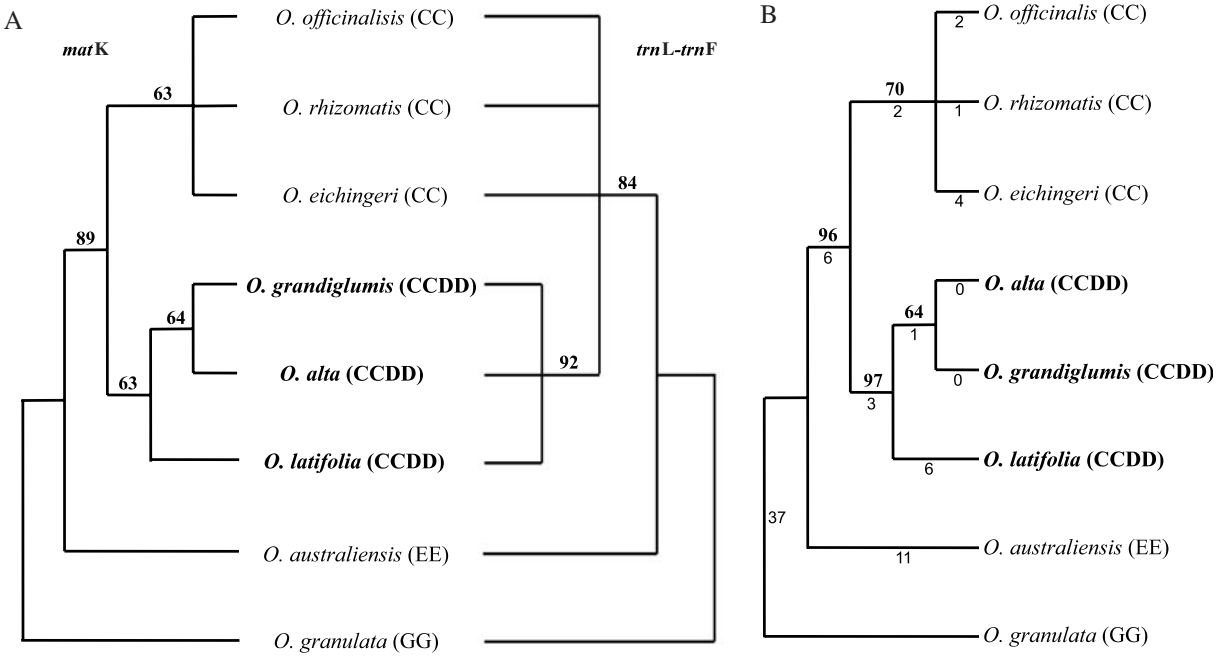

Fig. 1. Phylogenetic trees inferred from cpDNA data sets. **A** Left, the single most parsimonious tree based on *mat*K gene sequences (Tree length = 45, CI = 0.98, RI = 0.89); Right, the single most parsimonious tree based on *trn*L intron and *trn*L-*trn*F spacer sequences (Tree length = 28, CI = 0.96, RI = 0.90). **B** The single most parsimonious tree based on the combined *mat*K and *trn*L-*trn*F data sets. Numbers above branches indicate bootstrap values above 50 % and those below branches indicate nucleotide substitution. Boldface indicates allotetraploid species

Fig. 2A. Almost identical topologies with two monophyletic clades were demonstrated on two *Adh* trees. One clade included three CC species and the C-like copies of three CCDD species with high bootstrap supports (100% for both *Adh* genes), while the other clade comprised the EE species and the D-like copies of the CCDD species with 99% (*Adh*1) and 100% (*Adh*2) bootstrap supports, respectively. Partition-homogeneity test indicated that two data sets were not incongruent ($p = 1.00$), and then they were combined for further phylogenetic analysis. One most parsimonious tree (CI = 0.91 and RI = 0.95) was obtained based on the combined data sets. Similarly, the combined tree was split into two main clades with 100 % bootstrap values (Fig. 2B). In the first clade, diploid *O. officinalis* and *O. rhizomatis* were clustered together (92% bootstrap), and then formed a clade with the C-like copies of three CCDD species (91% bootstrap). Four accessions of *O. eichingeri* formed a moderately-supported group (63%). In the second clade, the D-like copies of three CCDD species formed a highly supported monophyletic clade with the EE species *O. australiensis* (100% bootstrap). It is interesting to note that in contrast to the CC species where accessions from same species were clustered together, either C or D-like copies from *O. alta* and *O. grandiglumis* were mixed together (Fig. 2B). Trees generated by Neighbor–joining (NJ) method had topologies identical to the most parsimonious trees based on either separate or combined data sets (not shown).

Phylogenetic analysis of nuclear *GPA1* gene and combined nuclear *Adh* and *GPA1* genes. Strict consensus tree of two most parsimonious

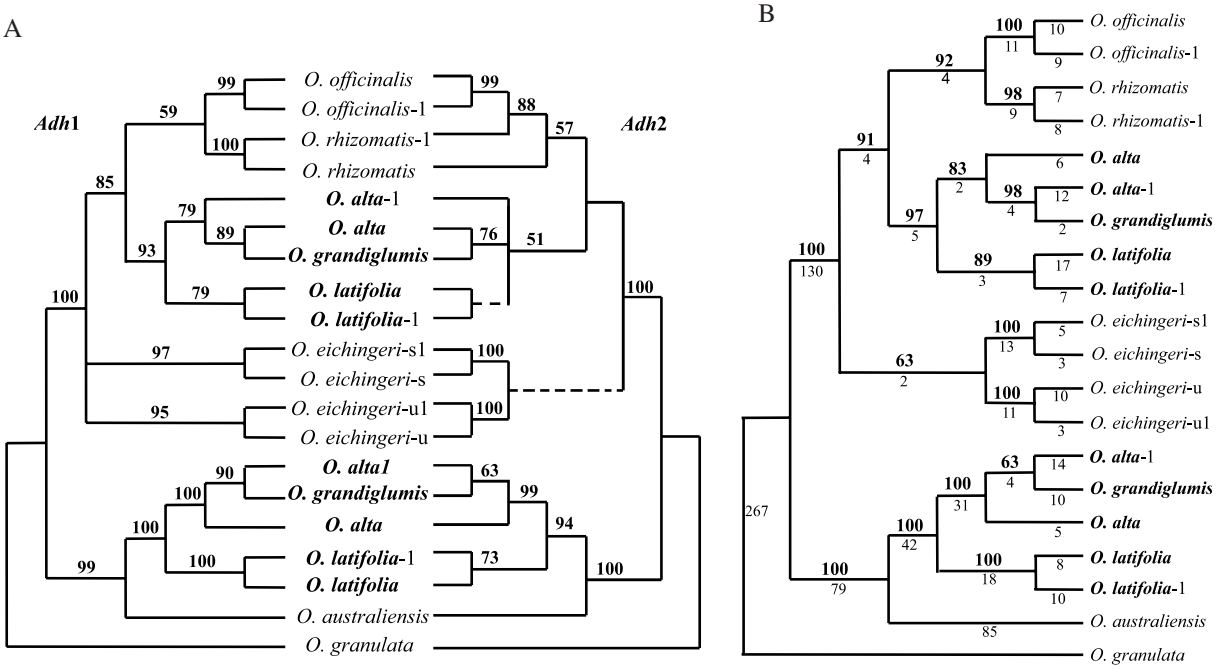

Fig. 2. Phylogenetic trees inferred from two *Adh* data sets. **A** Left, the strict consensus of two most parsimonious trees based on *Adh*1 sequences (Tree length = 465, CI = 0.91, RI = 0.95). Right, the single most parsimonious tree based on *Adh*2 sequences (Tree length = 406, CI = 0.92, RI = 0.95). **B** The single most parsimonious tree based on the combined *Adh*1 and *Adh*2 data sets (tree length = 870, CI = 0.91, RI = 0.95). Numbers above branches indicate bootstrap values above 50 % and those below branches indicate nucleotide substitution. Dashed lines indicate the branch with bootstrap lower than 50%. Boldface indicates allotetraploid species

trees for *GPA1* gene (CI = 0.92 and RI = 0.93) was shown in Fig. 3. Three CC species and the C-like copies of three CCDD species formed a monophyletic group with 100% bootstrap support, although the relationships of three CC species were not determined because of low resolution. The other clade consisted of the EE species and the D-like copies of the CCDD species with 99% bootstrap support. Partition-homogeneity test indicated that two *Adh* data sets and one *GPA1* data set were not incongruent (p = 0.68) and then were combined for further phylogenetic analysis. One most parsimonious tree (CI = 0.93 and RI = 0.94) was obtained based on the combined data sets of three nuclear genes, showing the same topological structure as the combined Adh tree (Fig. 4A). Two monophyletic clades were supported by 100% bootstrap with CC species and the C-like copies of CCDD species as one clade, and the EE species and the D-like copies of CCDD species as the other clade. In addition, the C-like copies of the CCDD species clustered first with *O. officinalis*-*O. rhizomatis* clade and then with *O. eichingeri* in the C genome clade. As for the CCDD species, *O. alta* and *O. grandiglumis* were closely related to each other on either C genome clade (95% bootstrap) or D (E) genome clade (100% bootstrap). Trees generated by Neighbor–joining (NJ) method had topologies identical to the most parsimonious trees (not shown). NJ tree based on the combined data set of three genes also resulted in two highly supported monophyletic clades corresponding to the C and D (E) genome clusters, respectively (Fig. 4B).

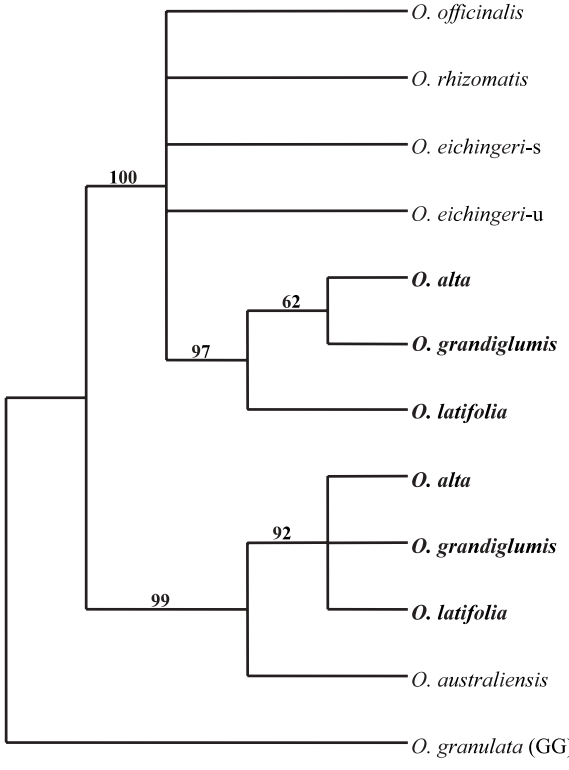

Fig. 3. Strict consensus of two most parsimonious trees inferred from *GPA*1 sequences (Tree length = 124, CI = 0.92, RI = 0.92). Numbers above branches indicate bootstrap values above 50%, and boldface indicates allotetraploid species

Discussion

Traditional genome analysis through assessment of chromosome pairing in an interspecific hybrid has contributed greatly to plant taxonomy and classification, particularly for economically important groups such as *Triticum*, *Hordeum*, *Oryza*, *Avena*, *Brassica*, and *Gossypium*. By studying the meiotic pairing of hybrids between *Oryza* species, Morinaga (1939, 1943) identified five different genomes with the A, B, C, D, and E genomes in diploids and BC and CD genomes in tetraploids. Although many previous studies indicated that the CD genomes originated from a single hybridization event (Jena and Kochert 1991, Wang et al. 1992, Aggarwal et al. 1996, Ge et al. 1999, Buso et al. 2001), the diploid donors of CD genome species have remained unclear. Besides supporting the single origin of the CD genome species, the present study was in good agreement with the previous recognition that the C genome served as the maternal parent of these tetraploids (Ge et al. 1999, Buso et al. 2001). Wang et al. (1992) found that *O. eichingeri* of Africa is the closest living relatives of the CD genome tetraploids. Our data, however, indicated that *O. officinalis* and *O. rhizomatis* other than *O. eichingeri* were

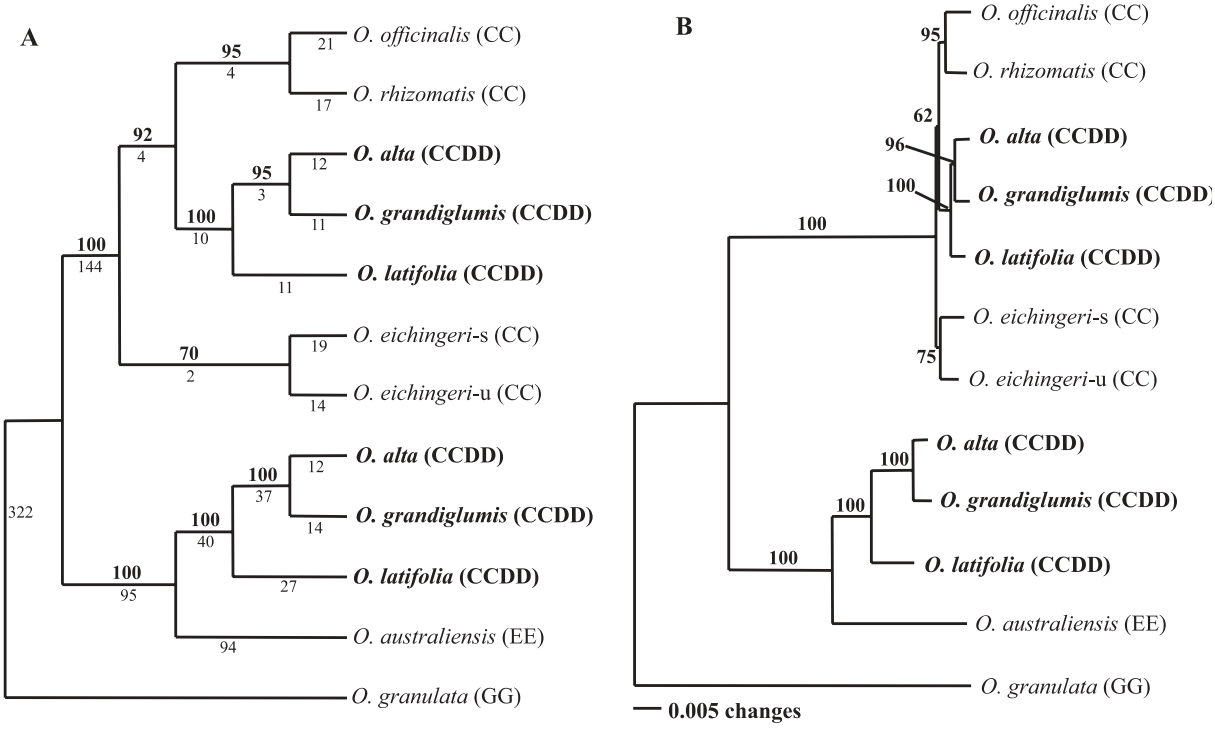

Fig. 4. Phylogenetic trees inferred from the combined *Adh*1, *Adh*2 and *GPA*1 data sets. **A** The single most parsimonious tree (Tree length = 916, CI = 0.93, RI = 0.94). **B** Neighbor joining tree. Numbers above branches indicate bootstrap values above 50 % and those below branches indicate nucleotide substitutions; Boldface indicates the allotetraploid species

more closely related to the CD genome species (Fig. 2 and Fig. 4).

The other donor of the CD genome is the most controversial issue and have long been the subject of dispute because no diploid species carrying the D genome has been identified thus far (Jena and Kochert 1991, Fukui et al. 1997, Li et al. 2001). To date, several hypotheses have been proposed. The most common explanation is that the D genome is extinct and remains only in the CCDD species (Jena and Kochert 1991). An alternative explanation for the origin of D genome, as proposed by Gopalakrishnan and Sampath (1967), is that the genomes C and D are closely related and the D may be merely a variant of the C genome (Nayar 1973). Using genomic *in situ* hybridization (GISH), Fukui et al. (1997) found that the overall nucleotide sequence homology between the B and C genomes was less than that between the C and D genomes, thus supporting a C-genome origin of the D genome. Based on nuclear RFLP data, Wang et al. (1992) found that the CD genome species had smaller genetic distances with C and E genome species than with other diploid species, and suggested that the E genome was related to the D genome and might have played a role in the formation of CCDD species. Ge et al. (1999) investigated the phylogenetic relationship among 23 *Oryza* species using sequence data from two nuclear *Adh* genes and a chloroplast *mat*K gene. Their *Adh* gene trees presented clearly that two distinct types of copies were found for each of three CD genome species and one of the copies formed a monophyletic group with the EE species with 100% bootstrap support (Ge et al. 1999). Our present study, based on three nuclear gene sequences with more accessions representing the C, CD and E genome species, strongly supported the close affinity between

the D and E genomes, with the E genome as the paternal donor during the formation of CCDD species (Figs. 1–4). Therefore, the hypothesis that the D genome is merely a variant of the C genome is less likely according to the evidence available.

Recently, Li et al. (2001) conducted a GISH study on *O. officinalis* complex and suggested that the E genome might not be the direct donor of the CD genome because the E genome was closer to the C than to the D genome. It should be noted, however, that the above results of both Fukui et al. (1997) and Li et al. (2001) were exclusively based on total genomic *in situ* hybridization (GISH), which presents technical limitations when used for phylogenetic study on allopolyploids. This is because GISH, with total genomic DNA as a probe, provides information about similarities between repetitive DNA from related species (Belyayev et al. 2000). As evidenced by Belyayev et al. (2000) in allotetraploid *Triticum dicoccoides* (AABB), the ancestral B-genome sequences have spread throughout the AB tetraploid genome to a greater extent than the ancestral A-genome sequences, because of interlocus concerted evolution and (or) colonization. Consequently, GISH experiments evidently reflect general tendencies of intrapolyploid repetitive sequence interaction (Belyayev et al. 2000), and may be less useful for phylogenetic reconstruction in this particular case.

Whether the three CCDD species should be treated as separate entities or just a single species has been disputed for decades (Nayar 1973, Jena and Kochert 1991, Aggarwal et al. 1996, Buso et al. 2001). In his comprehensive review on *Oryza* species, Roschevicz (1931) treated these American tetraploids as two separate species (*O. grandiglumis* and *O. latifolia*). The third species *O. alta*, which had been known as *O. latifolia* var. *grandispiculis* (Chevalier 1932), was described by Swallen (1936) (see the review by Nayar 1973). These treatments have been kept by many later workers (Chatterjee 1948; Tateoka 1962; Vaughan 1989, 1994, Lu 1999). However, accumulating morphological, cytogenetic and distribution data suggested that the three species are better considered as conspecific (Nayar 1973). Tateoka (1962) pointed out that spikelet length could separate *O. latifolia* from the other two tetraploids (*O. alta* and *O. grandiglumis*), but the variation on sterile lemmas between *O. grandiglumis* and *O. alta* was not stable. Much closer relationships between *O. alta* and *O. grandiglumis* have been demonstrated from previous molecular data (Wang et al. 1992; Aggarwal et al. 1996, 1999; Ge et al. 1999; Federici et al. 2002). Based on total genomic hybridization, for example, Aggarwal et al. (1996) showed that *O. latifolia* was the most divergent among three CCDD species, and *O. alta* and *O. grandiglumis* were more similar to each other. Federici et al. (2002) conducted a RFLP analysis on the *Oryza officinalis* complex and found that *O. alta* and *O. grandiglumis* had almost identical hybridization patterns and were clearly separated from *O. latifolia*. In this study, we found that *O. alta* and *O. grandiglumis* were always mixed together on both nuclear *Adh* and *GPA*1 trees, but they were all separated clearly from *O. latifolia*. These data demonstrated that *O. alta* and *O. grandiglumis* are very closely related genetically but *O. latifolia* can be distinguished from the former two. Therefore, we suggest that *O. alta* and *O. grandiglumis* should be better treated as a single species (*O. grandiglumis*), and *O. latifolia* remains a separated species, in agreement with the treatment of Roschevicz (1931). However, a valid taxonomic treatment should be further conducted based on a comprehensive revision with combined morphological, cytological and molecular evidence.

We thank Tao Sang of Michigan State University (East Lansing, USA) and Bao-rong Lu of Fudan University (Shanghai, China) for their encouragement and assistance. We are also grateful to the International Rice Research Institute (Manila, Philippines) for providing plant material for this study. This research was supported by the Chinese Academy of Sciences (kscxz-sw-101A), the National Natural Science Foundation of China

(30025005) and the Program for Key International S & T Cooperation Project of P. R. China (2001CB711103).

References

Aggarwal R. K., Brar D. S., Huang N., Khush G.S. (1996) Differentiation within CCDD genome species in the genus *Oryza* as revealed by total genomic hybridization and RFLP analysis. Rice Genet News. 13: 54–57.

Baumel A., Ainouche M.L., Bayer R. J., Ainouche A.K., Misset, M.T. (2002) Molecular phylogeny of hybridizing species from the genus *Spartina* Schreb. (Poaceae). Mol. Phylogenet. Evol. 22: 303–314.

Belyayev A., Raskina O., Korol A., Nevo E. (2000) Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. Genome 43: 1021–1026.

Buso G. S. C., Rangel P.H., Ferreira M. E. (2001) Analysis of random and specific sequences of nuclear and cytoplasmic DNA in diploid and tetraploid American wild rice species (*Oryza* spp.). Genome 44: 476–494.

Chatterjee D. (1948) A modified key and enumeration of the species of *Oryza* L. Indian J. Aric. Sci. 18: 185–192.

Chevalier A. (1932) Nouvelle contribution a l'etude systematique des *Oryza*. Rev. Bot. Appl. Agric. Trop. 12: 1014–1032.

Dally A. M., 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.

Doyle J. J., Doyle J. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19: 11–15.

Doyle J. J., Doyle J. L., Brown A. H. D., Pfeil B.E. (2000) Confirmation of shared and divergent genomes in the *Glycine tabacina* polyploid complex (Leguminosae) using histone H3-D sequences. Syst. Bot. 25: 437448.

Federici M. T., Shcherban A. B., Capdevielle F., Francis M., Vaughan D. (2002) Analysis of genetic diversity in the *Oryza officinalis* complex. Biotech. 5: 173–181.

Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

Fukui K., Shishido R., Kinoshita T. (1997) Identification of the rice D-genome chromosomes by genomic in situ hybridisation. Theor. Appl. Genet. 95: 1239–1245.

Fujisawa Y., Kato T., Ohki S., Ishikawa A., Kitano H., Sasaki T., Asahi T., Iwasaki Y. (1999) Suppression of the heterotrimeric G protein causes abnormal morphology, including dwarfism, in rice. Proc. Natl. Acad. Sci. USA 96: 7575–7580.

Gaut B. S., d'Ennequin M. L. T., Peek A. S., Sawkins M.C. (2000) Maize as a model for the evolution of plant nuclear genomes. Proc. Natl. Acad. Sci. USA 97: 7008–7015.

Ge S., Sang T., Lu B.-R., Hong D. Y. (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA 96: 14400–14405.

Ge S., Sang T., Lu B.-R., Hong D. Y. (2001a) Rapid and reliable identification of rice genomes by RFLP analysis of PCR- amplification *Adh* genes. Genome 44: 1136–1142.

Ge S., Sang T., Lu B.-R., Hong D. Y. (2001b) Phylogeny of the genus *Oryza* as revealed by molecular approaches. In: Khush G. S., Brar D.S., Hardy B. (eds.) Rice genetics IV. Proceedings of the Fourth International Rice Genetics Symposium. IRRI, Philippines, pp. 89–105.

Gopalakrishnan R., Sampath S. (1967) Taxonomic status and the origin of American tetraploid species of the series *Latifoliae* Tateka in the genus *Oryza*. India J. Agric. Sci. 37: 465–475.

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

Johnson L. A., Soltis D. E. (1998) Assessing congruence: Empirical examples from molecular data. In: Soltis D. E., Soltis P. S., Doyle J. J. (eds.) Molecular systematics of plants. Kluwer Academic Publishers, Boston, pp. 297–348.

Kimura M. (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.

Li C.-B., Zhang D.-M., Ge S., Lu B.-R., Hong D.-Y. (2001) Differentiation and inter-genomic relationships among C, E and D genomes in the *Oryza officinalis* complex (Poaceae) as revealed by multicolor genomic in situ hybridization. Theor. Appl. Genet. 103: 197–203.

Lu B.-R. (1999) Taxonomy of the genus *Oryza* (Poaceae): historical perspective and current status. Intern. Rice Res. Not. 24: 4–8.

Ma H. (1994) GTP-binding proteins in plants: new members of an old family. Plant Mol. Biol. 26: 1611–1636.

Morinaga T. (1939) Cytogenetics of rice (*Oryza sativa* L.). Bot. Zool. 7: 179–183.

Morinaga T. (1943) Cytogenetical studies on *Oryza sativa* L. VI. The cytogenetics of F_1 hybrids of *O. minuta* Presl. and *O. latifolia*. Jpn. J. Bot. 12: 347–357.

Nayar N. M. (1973) Origin and cytogenetics of rice. Adv. Genet. 17: 153–292.

Popp M., Oxelman B. (2001) Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. Mol. Phylogenet. Evol. 20: 474–481.

Roschevicz R. I. (1931) A contribution to the study of rice. Turdy Prikl. Bot. Genet. Selek. 27: 3–133.

Saitou N., Nei M. (1987) The neighbor-joining method: a new method for reconstruction phylogenetic trees. Mol. Biol. Evol. 4: 406–425.

Sang T., Zhang D. (1999) Reconstructing hybrid speciation using sequences of low copy nuclear genes: Hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. Syst. Bot. 24: 148–163.

Sang T. (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37: 121–147.

Small R. L., Ryburn J. A., Cronn R. C., Seelanan T., Wendel J. F. (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Amer J. Bot. 85: 1301–1315.

Soltis D. E., Soltis P. S. (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis D. E., Soltis P. S., Doyle J.J. (eds.) Molecular systematics of Plants. Kluwer Academic Publishers, Boston, pp. 1–42.

Stebbins G. L. (1950) Variation and Evolution in plants. Columbia University Press, New York.

Swallen J. R. (1936) *Oryza*. In: Biology of the Maya Area. Misc. Pap. IX. Carnegie Inst. Wash. Publ. No. 461, pp. 155–156.

Swofford D. L. (1998) PAUP* 4.0: Phylogenetic Analysis Using Parsimony. Beta version 4.0b. Sinauer Associates, Sunderland, MA.

Taberlet P., Gielly L., Pautou G., Bouvet J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17: 1105–1109.

Tateoka T. (1962) Taxonomic studies of *Oryza* I. *O. latifolia* complex. Bot. Mag. Tokyo 75: 418–427.

Thompson J. D., Higgins D. G., Gibson T. J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic. Acids Res. 22: 4673–4680.

Vaughan D. A. (1989) The genus *Oryza* L. Current status of taxonomy. International Rice Research Institute, Philippines.

Vaughan D. A. (1994) The wild relatives of rice: A genetic resources handbook. Philippines, IRRI, International Rice Research Institute.

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

Wendel J. F. (2000) Genome evolution in polyploids. Plant Mol. Biol. 42: 225–249.

Address of the authors: Ying Bao, Department of Biology, Qufu Normal Univerity, Qufu 273165, China. Song Ge (e-mail: gesong@ns.ibcas.ac.cn), Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.