GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Chloroplast DNA polymorphism and evolutional relationships between Asian cultivated rice (*Oryza sativa*) and its wild relatives (*O. rufipogon*)

W.J. Li, B. Zhang, G.W. Huang, G.P. Kang, M.Z. Liang and L.B. Chen

College of Life Science, Hunan Normal University, Changsha, Hunan, China

Corresponding authors: M.Z. Liang / L.B. Chen
E-mail: 1962liang@163.com / liwenjia0529@163.com

Genet. Mol. Res. 11 (4): 4418-4431 (2012)
Received March 12, 2012
Accepted June 6, 2012
Published October 9, 2012
DOI http://dx.doi.org/10.4238/2012.October.9.8

ABSTRACT. We analyzed chloroplast DNA (cpDNA) polymorphism and phylogenic relationships between 6 typical *indica* rice, 4 *japonica* rice, 8 *javanica* rice, and 12 Asian common wild rice (*Oryza rufipogon*) strains collected from different latitudes in China by comparing polymorphism at 9 highly variable regions. One hundred and forty-four polymorphic bases were detected. The *O. rufipogon* samples had 117 polymorphic bases, showing rich genetic diversity. One hundred and thirty-one bases at 13 sites were identified with *indica*/*japonica* characteristics; they showed differences between the *indica* and *japonica* subspecies at these sites. The *javanica* strains and *japonica* shared similar bases at these 131 polymorphic sites, suggesting that *javanica* is closely related to *japonica*. On the basis of length analyses of the open reading frame (ORF)100 and (ORF)29-tRNA-Cys(GCA) ($TrnC^{GCA}$) fragments, the *O. rufipogon* strains were classified into *indica/japonica* subgroups, which was consistent with the results of the phylogenic tree assay based on concatenated datasets. These results indicated that differences in *indica* and *japonica* also exist in the cpDNA genome of the *O. rufipogon* strains. However,

these differences demonstrated a certain degree of primitiveness and incompleteness, as an *O. rufipogon* line may show different *indica*/*japonica* attributes at different sites. Consequently, *O. rufipogon* cannot be simply classified into the *indica*/*japonica* types as *O. sativa*. Our data support the hypothesis that Asian cultivated rice, *O. indica* and *O. japonica*, separately evolved from Asian common wild rice (*O. rufipogon*) strains, which have different *indica-japonica* differentiation trends.

Key words: Genetic diversity; Evolutionary relationship; Dual origin; Chloroplast DNA; *Indica-japonica* differentiation

INTRODUCTION

Asian cultivated rice (*Oryza sativa* L.) is a major source of nutrition for more than half of the global population (Vaughan et al., 2003). Asian common wild rice (*O. rufipogon* Griff.) has long been believed to be the direct ancestor of Asian cultivated rice. It serves as a valuable gene pool for the improvement of cultivated rice due to its many extraordinary traits, such as disease and insect resistance, stress tolerance, cytoplasmic male sterility, and others (Khush, 1997). Therefore, a better understanding of the genetic diversity and phylogenic relationships between Asian cultivated rice and Asian common wild rice will be beneficial for the efficient use of the *Oryza* gene pool, and will certainly have a significant impact on rice production (Sun et al., 2001; Vaughan et al., 2003; Garris et al., 2005).

In the past decades, various phylogenetic tools have been used to investigate the evolutionary relationships between closely related *Oryza* species, such as morphological characteristics (Morishima and Oka, 1960), isoenzymes (Cai and Wang, 1992; Wang et al., 1994), rDNA (Fukui et al., 1994), random amplified polymorphic DNA (Sun et al., 1995), restriction fragment length polymorphism (Wang and Tanksley, 1989; Nakano et al., 1992; Sun et al., 1997, 2001), simple sequence repeats (Nishikawa et al., 2005), inter-simple sequence repeats (Joshi et al., 2000), nuclear coding sequences (Chen et al., 1994; Sun et al., 1997), single nucleotide polymorphism (SNP; Caicedo et al., 2007), and combined chloroplast, mitochondrial, and nuclear data analyses (Duan et al., 2007). Nevertheless, due to the differences in methodologies and resulting data, some reports suggested that the *indica-japonica* differentiation occurred not only in cultivated rice but also in its wild relatives, which provided theoretical support for the dual origin of cultivated rice. This hypothesis states that the *indica* and *japonica* subspecies separately evolved from *O. rufipogon*. On the other hand, other studies supported the monophyletic origin of cultivated rice, which indicates that *indica* descended first from *O. rufipogon*, and the *japonica* subspecies in turn evolved from *indica*. Hence, the genetic diversity and evolutionary relationships between *O. sativa* and *O. rufipogon* remain controversial.

The DNA of organelles has been used in phylogenetic studies due to its maternal inheritance, higher genetic stability, and lower mutation frequencies compared with nuclear DNA. In rice, the evolutionary rate of chloroplast DNA (cpDNA) is 3 times higher than that of mitochondrial DNA (mtDNA; Tian et al., 2006). Hence, its maternal inheri-

tance and relatively high mutation rate are useful in elucidating the phylogeny of the species. Hirai et al. (1985) constructed the first physical map of rice cpDNA. Hiratsuka et al. (1989) published the rice chloroplast genome sequence from *Nipponbare*. Five chloroplast genome sequences of rice are available at present. In addition, open reading frame (ORF)100 has been reported as an efficient marker that determines whether a rice chloroplast genome belongs to either the *indica* or the *japonica* type. The ORF100 band of *japonica* rice lags behind that of *indica* rice as a result of a 69-bp deletion in the *indica* ORF100 region (Kanno et al., 1993). Chen et al. (1994) classified 137 cultivated rice into *indica/japonica* based on the ORF100 marker, which was consistent with the classification based on morphological and physiological characteristics. Using this marker, 151 strains of *O. rufipogon* were identified, and the cpDNA of *O. rufipogon* can also be classified into the *indica* and *japonica* types (Sun et al., 1997). Tang et al. (2004) found that *indica* has a 32-bp insertion in the ORF29-tRNA-Cys(GCA) ($TrnC^{GCA}$) spacer, leading to a lag of the *indica* bands related to those of *japonica*. Therefore, ORF29-$TrnC^{GCA}$ can be used as another marker to distinguish the *indica* chloroplast genome from that of *japonica*. Shaw et al. (2005) found that the intron of the s16 (*rps16*) ribosomal protein and the spacer of $TrnT^{UGU}$-$TrnL^{UAA}$ [tRNA-Thr(UGU)-tRNA-Leu(UAA)] are highly variable regions in plant chloroplast genomes. These studies primarily focused on the sequences of the non-coding regions of cpDNA. However, some variable encoding regions of cpDNA that contain important genes involved in photosynthesis also provide effective sources for analyzing the phylogenetic relationships of closely related species.

In the present study, we performed a more comprehensive comparison of the cpDNAs of *O. sativa* and *O. rufipogon* using the sequences of 3 non-coding regions and 6 coding regions to explore the genetic diversity of cpDNA and the evolutionary relationship between *O. sativa* and *O. rufipogon* as well as to reveal the *indica/japonica* attributes of the *javanica* cpDNA. The results will deepen our understanding of the evolutionary relationships between *O. sativa* and *O. rufipogon* and ultimately accelerate the efficient use of the extraordinary genes of *O. rufipogon* for the genetic improvement of cultivated rice.

MATERIAL AND METHODS

Comparison of chloroplast genomes

Four chloroplast genomic sequences, *O. sativa japonica* (*Nipponbare*, AY522330.1, Tang et al., 2004 and *PA64S*, AY522331.1, Tang et al., 2004), *O. sativa indica* (isolate *9311*; AY522329.1, Tang et al., 2004), and *O. nivara* (AP006728.1, Shahid et al., 2004), from the National Center for Biotechnology Information (NCBI) nucleotide database were compared *in silico*. These sequences were aligned with ClustalW, and we manually proofed the alignment for extracting encoding regions that contain many variant sites.

Plant materials

A total of 30 accessions of rice were used in this study, which contained 18 *O. sativa* (6 *indica*, 4 *japonica*, and 8 *javanica*) and 12 *O. rufipogon* from different latitude distributions in China (Table 1).

Table 1. Plant materials used in this study.

No.	Materials	Origin/latitude	Provider
1	*Guangluai-4*[a]	Guangdong, China	Hunan Normal University
2	*IR36*[a]	IRRI	Hunan Normal University
3	*Nanjing-3*[a]	Jiangsu, China	Hunan Normal University
4	*9311*[a]	IRRI	Hunan Normal University
5	*Kasalath*[a]	India	Hunan Normal University
6	*Basmati*[a]	India	Hunan Normal University
7	*Nipponbare*[b]	Japan	Hunan Normal University
8	*Bairizao*[b]	Sichuan, China	Hunan Normal University
9	*Akihikari*[b]	Japan	Hunan Normal University
10	*Aizinuo*[b]	Yunan, China	Hunan Normal University
11	*Lemont*	USA	Hunan Normal University
12	*P002*	USA	Hunan Normal University
13	*IM359*	Unknown	Hunan Normal University
14	*IM273*	Unknown	Hunan Normal University
15	*CPSL017*	Unknown	Hunan Normal University
16	*Rosemont*	USA	Hunan Normal University
17	*D16*	Unknown	Hunan Normal University
18	*Brazilian UR*	Brazil	Hunan Normal University
19	*Hainan CWR*	Hainan (18.73N), China	Hunan Normal University
20	*Hepu CWR*	Guangxi (21.38N), China	Nanning Wild Rice Garden
21	*Gongguan CWR*	Guangxi (21.39N), China	Nanning Wild Rice Garden
22	*Gaozhou CWR*	Guangdong (21.42N), China	South China Agriculture University
23	*Fusui CWR*	Guangxi (22.25N), China	Nanning Wild Rice Garden
24	*Zengcheng CWR*	Guangdong (23.28N), China	Nanning Wild Rice Garden
25	*Tianyang CWR*	Guangxi (23.37N), China	Nanning Wild Rice Garden
26	*Liujiang CWR*	Guangxi (24.08N), China	South China Agriculture University
27	*Guilin CWR*	Guangxi (24.15N), China	South China Agriculture University
28	*Jiangyong CWR*	Hunan (25.41N), China	Hunan Rice Institute
29	*Chaling CWR*	Hunan (26.50N), China	Hunan Normal University
30	*Dongxiang CWR*	Jiangxi (28.14N), China	Hunan Normal University

[a]Typical *indica* rice; [b]typical *japonica* rice.

Chloroplast DNA extraction

Chloroplast DNA was extracted from the fresh leaves of rice using the method mentioned by Rogers and Bendich (1988).

PCR amplification and agarose-gel electrophoresis

Chloroplast genome primers cp1-cp11 were designed with Primer Premier 5.0 based on the chloroplast genome sequence of *9311* (Table 2). The PCR system included 2.5 µL 10X PCR buffer, 0.5 µL forward and reverse primers, 0.3 µL 10 mM dNTPs, 1 U Taq polymerase, 40 ng template DNA, and complementary ultrapure water to 25 µL. PCR amplification was performed as follows: pre-denaturation at 95°C for 3 min, then 95°C for 30 s, 50°-55°C for 30 s, and 72°C for 50-90 s; the cycle was repeated 30 times, and then 72°C for 10 min. The PCR products were fractionated on a 1-3% agarose gel, which contained 0.5 µg/mL ethidium bromide, and then the gel was observed with an ultraviolet gel imaging system.

DNA sequencing and data analysis

The target fragments were isolated from the agarose gel, purified with the reagent

kit [Tiangen Minipurification kit, Tiangen Biotech (Beijing) Co., Ltd.], and then directly sequenced. Sequence analyses on these polymorphic fragments were performed by the molecular evolutionary genetics analysis (MEGA) version 4.0 software (Tamura et al., 2007). GC content, consistency index, retention index, and Tajima's D statistic were calculated by MEGA 4.0. Phylogenetic analyses were conducted using maximum parsimony (MP) and neighbor joining (NJ) as implemented in MEGA 4. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). The MP analysis was carried out with all SNPs, inserts, and deletions (indels) data. The MP tree was obtained using the close-neighbor-interchange algorithm (Nei and Kumar, 2000). In the NJ analysis (Saitou and Nei, 1987), evolutionary distances were calculated using the maximum composite likelihood model as the number of base substitutions per site. Bootstrap resampling with 1000 replicates was performed to test the reliability of the inferred phylogenetic trees (Felsenstein, 1985).

Table 2. Primer sequence and target fragment of chloroplast DNA.

Primer	Forward sequence (5'-3')	Reverse sequence (5'-3')	Target fragment
cp1	GTGGACCTGACTCCTTGAA	AGCCGAGGTCGTGGTAA	ORF100
cp2	GCAGCCCAAGCGAGACT	AAGGCTCGGCGATACTG	ORF29-*TrnC*^{GCA}
cp3	TTTTCTCCTCATACGGCT	TAGTCTGTTCTATTCGTCCC	*rps16* gene intron
cp4	AGTGGGCTTACATAACAGAAA	ACCAAGGCTCAATACAATCA	*TrnT*^{UGU}-*TrnL*^{UAA}
cp5	AGTCTTTGCCAATGCGATAA	GTTGCTGACCCATACCAC	*psbB*
cp6	CTATGTGGTATGGGTCAGC	AAAGAAGGGATGGGAGAT	*psbB*
cp7	TAGCAACACTTTTCCCACA	GCTCATTCCAATCCTCAA	*rpoc2*
cp8	TGAGGATTGGAATGAGCG	TTTTGGGAGATGGTGGA	*rpoc2*
cp9	TCAAGCATCTCGGTTCG	CGCCTCATTAGCCCATAC	*ccsA*
cp10	TTCCCAGTAAGAGGGTCAA	CTTCTCACTTGTTATGGCTTG	*ndhH*
cp11	AGAGGGAAGTTGTGAGCATT	CGTCTGGGTATGCGTCCT	*psbA*

RESULTS

Identifying the encoding regions with condensed variant sites in *Oryza* cpDNA

We manually examined the sequence alignments by comparing the 4 chloroplast genomic sequences in the NCBI nucleotide database, and found 34 SNPs and 3 indels in the encoding regions. Of these, 19 SNPs could eventually lead to changes in the amino acids (Table 3). Five highly variable encoding regions were then selected for the genetic diversity and phylogenetic relationship analyses of rice. The 5 selected regions are fragments of photosystem II 47-kDa protein (*psbB*), RNA polymerase beta chain (*rpoc2*), cytochrome c biogenesis protein (*ccsA*), NADH dehydrogenase 48-kDa subunit (*ndhH*), and photosystem II protein D1 (*psbA*) genes.

CpDNA genetic diversity analysis of 9 highly variable regions

The ORF100, ORF29-TrnC^{GCA}, rps16 intron, and TrnT^{UGU}-TrnL^{UAA} fragments are currently the most widely used in cpDNA genetic diversity analyses. In these 4 cpDNA regions and the 5 previously mentioned selected variable coding regions, 144 polymorphic bases (bp) were detected among 30 rice accessions (Tables 4 and 5), and 117 polymorphic bases were found among the 12 strains of *O. rufipogon*, suggesting high genetic diversity in the cpDNA of various ecotypes from different distributions.

Table 3. SNPs and Indels found in encoding regions of rice cpDNA by comparing the four whole chloroplast genomic sequences.

Position[a]	Gene	Rice species				Change of amino acid
		9311	*O. nivara*	*Nipponbare-G*	*PA64S*	
412	psbA	CTC	CTC	TTC	TTC	Leu/Phe
4,547	rps16	TGT	TGT	GGT	GGT	Cys/Gly
7,141	psbK	AAC	AAC	AAT	AAT	
8,554	ORF100	D-69	D-69	I-69	I-69	
14,166	ORF91	AGC	AGC	AAC	AAC	Ser/Asn
16,613	psbM	A	-	-	-	
20,547	rpoB	CCA	CCA	CCT	CCT	
27,979-27,980	rpoc2	TTG	TTT	TGG	TGG	Leu/Phe/Trp
28,970	rpoc2	CGA	CGT	CGA	CGA	
29,073	rpoc2	GAC	GAC	AAC	AAC	Asp/Asn
29,945	rps2	CAA	CTA	CAA	CAA	Gln/Leu
30,925	atpI	AAA	AAT	AAA	AAA	Lys/Asn
35,342	atpA	GCA	GCA	GCG	GCG	
37,963	psaB	GTC	GTT	GTC	GTC	
52,090	atpB	TGC	CGC	CGC	CGC	Cys/Arg
53,188	atpB	ACG	CCG	ACG	ACG	Thr/Pro
54,869	rbcL	GGA	GGA	GGG	GGG	
56,545	ORF106	TCC	TCT	TCT	TCT	
56,818	ORF106	GGT	GGT	GGC	GGC	
63,618	petE	CTC	TTC	CTC	CTC	Leu/Phe
65,172	rpl33	GTT	GTC	GTT	GTT	
65,235	rpl33	GCA	GCG	GCA	GCA	
66,354	rpl20	TGG	TGG	TGA	TGA	Trp/Stop condon
69,301	psbB	GTG	GCG	GCG	GCG	Val/Ala
69,833	psbB	ACT	ACC	ACT	ACT	
76,656	rps8	D-21	I-21	D-21	D-21	
77,532	rpl14	ACT	GCT	ACT	ACT	Thr/Ala
77,750	rpl16	TAC	TAC	TGC	TGC	Tyr/Cys
79,426	rps3	TTC	TTC	CTC	CTC	Phe/Leu
102,887	ndhf	TGC	TGC	CGC	CGC	Cys/Arg
104,485	rpl32	ACAA	ACAA	-	-	
105,740	ccsA	TTG	TTG	TTA	TTA	
105,741-105,743	ccsA	CTT	CTT	AGC	AGC	Leu/Ser
112,788	ndhH	ACC	CCC	ACC	ACC	Thr/Pro
113,049	ndhH	TGA	TGA	GGA	GGA	Stop condon/Gly

[a]Position denotes the nucleotide position in the chloroplast genome sequence of *9311*.

A total of 141 polymorphic bases were found in the cpDNA fragments of *O. sativa*, of which 131 *indica-japonica*-characteristic bases spread across 13 sites were identified (marked "4" in Tables 4 and 5). Six *indica* and 4 *japonica* strains were clearly identified at these *indica-japonica*-specific sites, and 8 *javanica* strains had sequences similar to those of *japonica* at these sites. Based on the 131 *indica-japonica*-characteristic bases, the cpDNA classification results of these *O. sativa* strains coincide with the results of the length polymorphism of the ORF100 and ORF29 -TrnC^{GCA} fragments (Figure 1). Hence, the ORF100 and ORF29-TrnC^{GCA} fragments are reliable molecular markers that can be used to classify the *O. sativa* cpDNA into the *indica/japonica* types. The *O. rufipogon* strains can also be classified into the *indica/japonica* types according to these 2 fragments, which agree with the results of Sun et al. (1997). Nevertheless, no *O. rufipogon* strain had sequences similar to those of *indica* or *japonica* in all the 131 *indica-japonica*-characteristic bases. Some *O. rufipogon* strains had more *indica*-specific bases, whereas the others contained more *japonica*-specific bases. For example, *O. rufipogon* from Zengcheng had 117 *indica*-specific bases and 14 *japonica*-specific bases, whereas Chaling had 9 *indica*-specific bases and 122 *japonica*-specific bases. Therefore, although a certain *indica-japonica* differentiation trend exists in *O. rufipogon*, they could not be classified into the *indica*/*japonica* types as simply as those in *O. sativa*.

Table 4. Sequence divergence of the 4 cpDNA regions among 30 materials.

Material	Gene and its base site No.													
	ORF100	ORF29-TrnC^{GCA}		*rps16* intron		TrnT^{UGU}-TrnL^{UAA}								
	248-316[a]	130-131[a]	163-194[a]	267-273[a]	309-312	322-326[a]	413[a]	463	479	520	572	599	635	774-779[a]
9311	D-69	AA	I-32	CTTTATC	-	-	-	G	A	A	G	T	A	-
IR36	D-69	AA	I-32	CTTTATC	-	-	-	G	A	A	G	T	A	-
Nanjing-3	D-69	AA	I-32	CTTTATC	-	-	-	G	C	G	A	T	A	-
Guangluai-4	D-69	AA	I-32	CTTTATC	CTTT	-	-	G	A	A	G	T	A	-
Kasalath	D-69	AA	I-32	CTTTATC	CTTT	-	-	G	A	A	G	T	A	-
Basmati	D-69	AA	I-32	CTTTATC	CTTT	-	-	G	A	A	G	T	A	-
Nipponbare	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Bairizao	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
*Akihika*ri	I-69	GG	D-32	-	-	TATAT	T	A	A	A	G	C	T	AGAAAA
Aizinuo	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Lemont	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
P002	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
IM359	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
IM273	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
CPSL017	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Rosemont	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
D16	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Brazilian UR	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Hainan CWR	D-69	AA	I-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Hepu CWR	D-69	AA	I-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Gongguan CWR	D-69	AA	I-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Gaozhou CWR	I-69	GG	D-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Fusui CWR	I-69	GG	D-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Zengcheng CWR	D-69	AA	I-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Tianyang CWR	D-69	AA	I-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Liujiang CWR	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Guilin CWR	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Jiangyong CWR	D-69	AA	I-32	-	-	-	-	G	A	A	G	T	A	AGAAAA
Chaling CWR	I-69	GG	D-32	-	-	TATAT	T	G	A	A	G	T	A	AGAAAA
Dongxiang CWR	D-69	AA	I-32	-	-	-	-	G	A	A	G	T	A	AGAAAA

[a]*indica-japonica*-characteristic sites. All the GenBank accession Nos. of these sequences were listed in Table S1.

Sequence characteristics and phylogenetic analysis based on concatenated datasets

A total of 9080 bases of the cpDNA regions were sequenced separately from 30 *Oryza* accessions. The variable sites, varying from 1 (0.15%) to 6 (0.73%), were screened in all of the sampled sequences, except in the ORF100 region. The ORF100, ORF29-TrnC^{GCA}, rps16 intron, and TrnT^{UGU}-TrnL^{UAA} regions contained different numbers of informative indels, varying from 12 to 69 bp (Table 6). The total number of variable sites was 22 (0.24%), and that of the parsimoniously informative characters was 14 (0.15%). We tested the selection and neutrality with Tajima's D, using the sequences of *O. sativa* and *O. rufipogon* in each region. The results showed no significant value in any of the regions, which indicated that the areas are evolutionarily neutral and without selective pressure.

Two phylogenetic trees were reconstructed via the MP and NJ methods using the combined sequences of the 9 cpDNA regions to reveal the evolutionary relationship between *O. sativa* and *O. rufipogon*. The MP phylogenetic tree was assembled from all of the SNPs

(22 sites) and the length variations (8 sites) to fully extract information regarding the DNA sequences (Figure 2). On the other hand, the NJ tree was constructed based only on the SNP data (Figure 3). The MP and NJ trees showed nearly similar topologies, with little difference in the bootstrap values, indicating that the studied rice species were classified into 2 main clusters: 1 branch consisted of the *indica* and the 8 *O. rufipogon* strains, and the other consisted of the *japonica*, *javanica*, and the rest of the 4 *O. rufipogon* strains. These relationships were supported by bootstrap values of greater than 50% (Figures 2 and 3). The results also support the existence of *indica-japonica* differentiation trends in *O. rufipogon*.

Table 5. Sequence divergence of the 5 new selected cpDNA encoding regions among 30 materials.

Material	Gene and its base site No.							
	psbB		*rpoc2*		*ccsA*	*ndhH*		*psbA*
	550[a]	1082	3293[a], 3294	4387[a]	540-543[a]	109	370[a]	331[a]
9311	T	T	TG	G	GCTT	A	T	C
IR36	T	T	TG	G	GCTT	A	T	C
Nanjing-3	T	T	TG	G	GCTT	A	T	C
Guangluai-4	T	T	TG	G	GCTT	A	T	C
Kasalath	T	T	TG	G	GCTT	A	T	C
Basmati	T	T	TG	G	GCTT	A	T	C
Nipponbare	C	T	GG	A	AAGC	A	G	T
Bairizao	C	T	GG	A	AAGC	A	G	T
Akihikari	C	T	GG	A	AAGC	A	G	T
Aizinuo	C	T	GG	A	AAGC	A	G	T
Lemont	C	T	GG	A	AAGC	A	G	T
P002	C	T	GG	A	AAGC	A	G	T
IM359	C	T	GG	A	AAGC	A	G	T
IM273	C	T	GG	A	AAGC	A	G	T
CPSL017	C	T	GG	A	AAGC	A	G	T
Rosemont	C	T	GG	A	AAGC	A	G	T
D16	C	T	GG	A	AAGC	A	G	T
Brazilian UR	C	T	GG	A	AAGC	A	G	T
Hainan CWR	C	T	TG	G	GCTT	A	T	C
Hepu CWR	T	T	TG	G	GCTT	A	T	C
Gongguan CWR	T	T	GG	G	GCTT	A	T	C
Gaozhou CWR	C	T	GG	A	GCTT	A	G	T
Fusui CWR	C	T	GG	A	GCTT	A	G	T
Zengcheng CWR	C	T	TT	G	GCTT	A	T	C
Tianyang CWR	C	C	TT	G	GCTT	C	T	C
Liujiang CWR	C	T	GG	A	GCTT	A	G	T
Guilin CWR	C	T	GG	G	GCTT	A	G	T
Jiangyong CWR	T	T	TG	A	GCTT	A	G	T
Chaling CWR	T	T	TG	G	GCTT	A	T	C
Dongxiang CWR	T	T	TG	G	GCTT	A	T	C

[a]*indica-japonica*-characteristic sites. All the GenBank accession Nos. of these sequences were listed in Table S1.

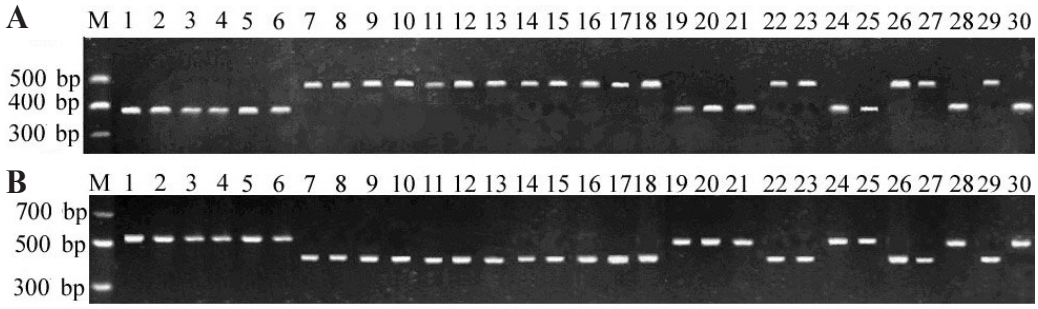

Figure 1. Agarose gel eletrophoresis results of ORF100 and ORF29-TrnC^{GCA} fragments. **A.** ORF100 fragments; **B.** ORF29-TrnC^{GCA} fragments. *Lane M* = molecular marker; *lanes 1-30* = different materials according to Table 1.

Table 6. Sequence characteristics of 9 regions and combined datasets.

Locus	Aligned length (bp)	GC (%)	No. of variable sites (%)	No. of informative sites (%)	Consistency index	Retention index	No. of informative indels	Tajima's D
ORF100	456	32.5	0	0	-	-	1 (69 bp)	n/c
ORF29-TrnC^{GCA}	514	37.8	2 (0.39)	2 (0.39)	1.000	1.000	1 (32 bp)	1.083071
rps16 intron	684	32.1	1 (0.15)	0	-	-	3 (17 bp)	-1.155588
TrnT^{UGU}-TrnL^{UAA}	824	29.3	6 (0.73)	0	-	-	3 (12 bp)	-1.091074
psbB	1688	43.9	2 (0.12)	2 (0.12)	1.000	1.000	0	0.737635
rpoc2	2474	38.1	4 (0.16)	3 (0.12)	0.800	0.960	0	0.959302
ccsA	859	33.9	4 (0.47)	4 (0.47)	1.000	1.000	0	1.092620
ndhH	964	37.8	2 (0.21)	2 (0.21)	1.000	1.000	0	0.838228
psbA	617	42.9	1 (0.16)	1 (0.16)	1.000	1.000	0	1.590843
Total	9080	37.5	22 (0.24)	14 (0.15)	1.000	1.000	8 (130 bp)	0.413816

n/c = not counted.

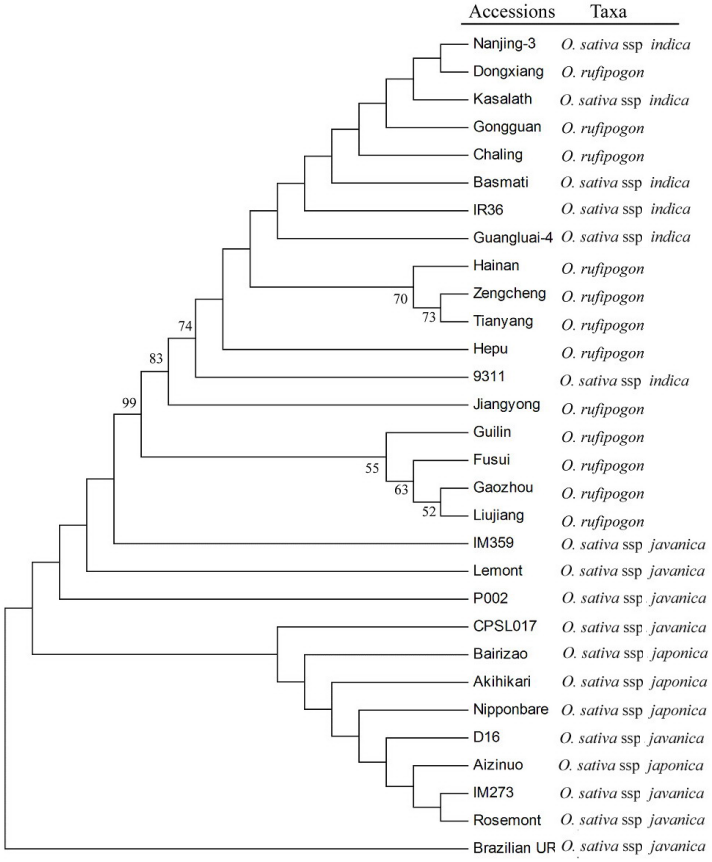

Figure 2. Evolutionary history inferred using the maximum parsimony method. A strict consensus tree of the 288 most parsimonious trees (length = 25) is shown. The consistency index is 0.705882, the retention index is 0.960938, and the composite index is 0.768750 for all sites and parsimony-informative sites. The MP tree was obtained using the close-neighbor-interchange algorithm. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

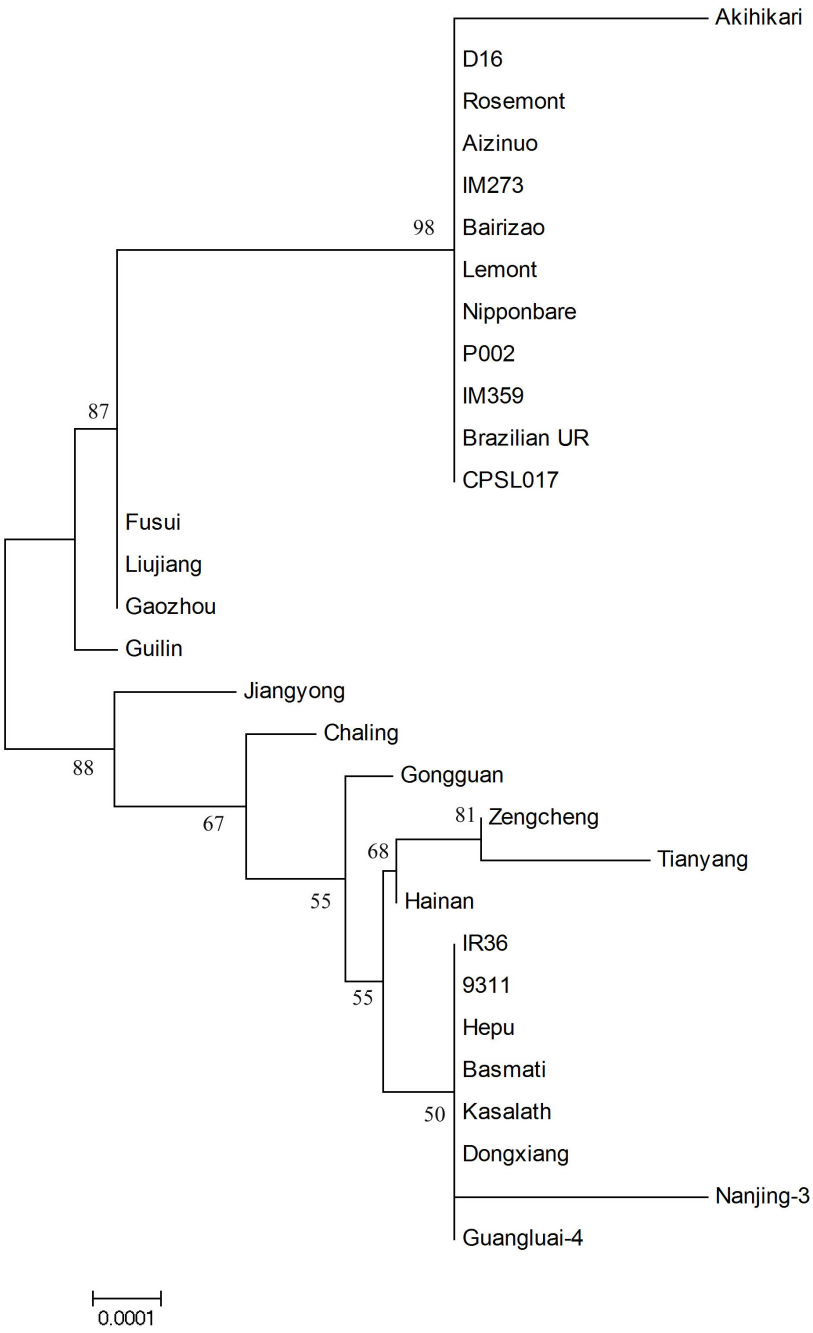

Figure 3. Evolutionary history inferred using the neighbor-joining method. The optimal tree with the sum of branch length = 0.00286400 is shown. The evolutionary distances were computed using the maximum composite likelihood model. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

DISCUSSION

CpDNA genetic diversity of *O. sativa*

Indica and *japonica* are 2 important subspecies of *O. sativa*. Their agronomic characteristics and ecological adaptability show remarkable differences, and this characteristic association of traits has been explained by hybrid sterility or by reproductive barriers (Sato and Morishima, 1988; Harushima et al., 1998). The nuclear genetic diversities between *indica* and *japonica*, particularly the nuclear genetic differences related to important agronomic traits, have been widely investigated in recent decades (Second, 1982; Wang and Tanksley, 1989; Chen et al., 1994; Garris et al., 2005). However, our understanding of the cytoplasmic genetic diversity between them, as well as of the relationship between their genetic diversity, agronomic characteristics, and phylogeny, is still limited (Kanno et al., 1993; Chen et al., 1994; Sun et al., 1997; Shaw et al., 2005). In the present study, 131 *indica-japonica* characteristic bases involved in 13 sites were identified based on the sequence analysis of 6 *indica*, 4 *japonica*, and 8 *javanica*, as listed in Table 1. The results revealed abundant polymorphism between the *indica* and *japonica* subspecies, not only in the non-coding regions, but also in the coding regions. These polymorphic bases in the encoding regions could lead to changes in the amino acids. However, whether these amino acid changes are related to the different phenotypes of these 2 subspecies remains to be determined. The *indica-japonica* classification results of these *O. sativa* strains in all of the *indica-japonica* characteristic nucleotides are highly consistent with those of the ORF100 and ORF29-TrnC^{GCA} fragments. Therefore, the analysis of the amplified fragment lengths of ORF100 and ORF29-TrnC^{GCA} using agarose gel electrophoresis is a simple, fast, and accurate method of identifying the *indica/japonica* types in the *O. sativa* cytoplasm.

Javanica showed very wide compatibility, as shown by the fairly good bearing rates of both its *indica/javanica* and *japonica/javanica* hybrids, which has great significance for achieving strong heterosis. However, the attributes of *javanica* remain controversial (Wang and Sun, 2000). Some scholars advocated that *O. sativa* L. should be classified into the *indica* (*O. sativa* ssp *indica*) and *japonica* (*O. sativa* ssp *japonica*) subspecies (Cheng, 1985), and that *javanica* is an ecotype of *japonica.* However, others argued that it should be divided into 3 subspecies, namely, *indica*, *japonica*, and *javanica* (Chang, 1976; Yuan, 1987). The *javanica* studied in our research has a relatively close relationship with *japonica* and is clustered into the *japonica* branch in both the MP and NJ trees; therefore, our results support the former classification system.

CpDNA genetic diversity and *indica-japonica* differentiation in *O. rufipogon*

The *O. rufipogon* strains used in our study were distributed from Yaxian (18.73N) of the Hainan Province to Dongxiang (28.14N; the highest latitude of *O. rufipogon* distribution) of the Jiangxi Province, both located in China. The plant types, photosensitivities, cold tolerance, and overwintering capabilities of the *O. rufipogon* strains from the different latitude distributions showed clear differences. The *O. rufipogon* strains from Dongxiang (28.14N), Chaling (26.50N), Jiangyong (25.41N), Guilin (24.15N), and Liujiang (24.08N) can tassel, bloom, and seed from September to November under natural conditions in Changsha (28.12N), whereas those from Hainan (18.73N) can only complete their entire life cycle under a 10-h short-day treatment. These results indicated differences in the photosensitivity of

floral induction, although the *O. rufipogon* strains were all typical short-day plants, the photosensitivity strengthened with decreasing distribution latitude. These *O. rufipogon* strains also showed significant differences in their cold tolerance and overwintering capabilities under natural conditions in Changsha. Those from the lowest latitude of Hainan (18.73N) could not live through winter; those from the 21.38N to 25.41N region could not survive overwintering in normal years without freezing weather, and only those from Chaling (26.50N) and Dongxiang (28.14N) safely survived the winter of 2008, which had a rare snow disaster that lasted for 20 days. These results demonstrated the biological diversity of *O. rufipogon*, which resulted from the difference in regional distributions (Li et al., 2001).

Our study also indicated rich genetic polymorphism in the cpDNA of *O. rufipogon* from different distributions. The *O. rufipogon* strains from near geographic distributions often had similar sequences and closer relationships. For instance, the *O. rufipogon* from Gaozhou (21.42N) and Fusui (22.25N) had perfectly similar sequences and were clustered into a clade based on the sequence analyses of these 9 highly cpDNA variable regions (Figures 2 and 3). The *O. rufipogon* strains showed different trends of *indica-japonica* differentiation in the 131 identified *indica-japonica* characteristic bases, which were validated by both the MP and NJ trees. Hence, these trends provide evidence on the level of cytoplasmic genetic information for dual origins; *indica* and *japonica* must have evolved separately from *O. rufipogon* strains with different *indica/japonica* differentiation trends (Chang, 1976; Second, 1982; Londo et al., 2006; Duan et al., 2007; Kumagai et al., 2010). However, no *O. rufipogon* strains that are completely identical to those of *indica* or *japonica* were found at these sites in our research, suggesting that the *indica-japonica* differentiation of *O. rufipogon* is relatively primitive and incomplete. Thus, the agarose gel electrophoresis results of the ORF100 and ORF29-TrnC^{GCA} fragments could only be used in the preliminary analysis of the *indica-japonica* differentiation trends in *O. rufipogon*, and we could not simply determine the characteristics of the *O. rufipogon* cpDNA based on these results. In addition, all of these corresponding *indica*/*japonica*-specific sequences, except the *indica*-specific sequences "CTTTATC" at positions 267-273 of the *rps16* intron and the *japonica*-specific sequences "AAGC" at positions 540-543 of the *ccsA* gene, could be detected in the *O. rufipogon* strains. When position 370 of the *ndhH* gene fragment is taken as an example, the *indica*-specific base C was found in *O. rufipogon* from Gaozhou, Fusui, Liujiang, Guilin, and Jiangyong, whereas the *japonica*-specific sequence G was found in the rest of the *O. rufipogon* strains (Tables 4 and 5). Hence, the *indica/japonica*-specific sequences of the cpDNA in *O. sativa* were mainly inherited from the *O. rufipogon* strains that have different *indica/japonica* differentiation trends. In addition, a small number of *indica/japonica*-specific sequences have independently formed through mutation during the long process of natural selection and artificial selection. This phenomenon is another theory that supports the dual origin of *O. sativa*.

ACKNOWLEDGMENTS

We thank the South China Agriculture University and Nanning Wild Rice Garden for providing the plant material for this study. Research supported by the National High Technology Research and Development Program (863 Program, #2010AA101305), Key Programs for Science and Technology Development of Hunan Province (#2009FJ1004-1) and the National Natural Science Foundation of China (#31171473/C1302).

Supplementary material

REFERENCES

Cai HW and Wang XK (1992). Classification of Asian rices by esterase isozymes. *Southeast China J. agric. Sci.* 5: 19-22.
Caicedo AL, Williamson SH, Hernandez RD, Boyko A, et al. (2007). Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* 3: 1745-1756.
Chang TT (1976). The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica* 25: 435-441.
Chen WB, Sato YI, Nakamura I and Nakai H (1994). Indica-Japonica differentiation in Chinese rice landraces. *Euphytica* 74: 195-201.
Cheng KS (1985). A statistical evaluation of the classification of rice cultivars into hsien and keng subspecies. *RGN* 2: 46-48.
Duan SH, Li SQ, Li YS, Xiong Y, et al. (2007). Distribution and SNPs of the rice CMS-related gene in AA-genome of *Oryza* species. *Yi Chuan* 29: 455-461.
Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.
Fukui K, Ohmido N and Khush GS (1994). Variability in rDNA loci in the genus *Oryza* detected through fluorescence in situ hybridization. *Theor. Appl. Genet.* 87: 893-899.
Garris AJ, Tai TH, Coburn J, Kresovich S, et al. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169: 1631-1638.
Harushima Y, Kurata N, Yano M, Sasaki T, et al. (1998). Detection of interactive loci for genotype segregation in F2 populations between japonica and indica rice crosses. *Breed. Sci.* 4: 82-101.
Hirai A, Isshibashi T, Morikami A, Iwatsuki N, et al. (1985). Rice chloroplast DNA: a physical map and the location of the genes for the large subunit of ribulose 1, 5-bisphosphate carboxylase and the 32 kD photosystem II reaction center protein. *Theor. Appl. Genet.* 70: 117-122.
Hiratsuka J, Shimada H, Whittier R, Ishibashi T, et al. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217: 185-194.
Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK, et al. (2000). Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. *Theor. Appl. Genet.* 100: 1311-1320.
Kanno A, Watanabe N, Nakamura I and Hirai A (1993). Variations in chloroplast DNA from rice (*Oryza sativa*): Differences between deletions mediated by short direct-repeat sequences within a single species. *Theor. Appl. Genet.* 86: 579-584.
Khush GS (1997). Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* 35: 25-34.
Kumagai M, Wang L and Ueda S (2010). Genetic diversity and evolutionary relationships in genus *Oryza* revealed by using highly variable regions of chloroplast DNA. *Gene* 462: 44-51.
Li DY, Liang YM and Yang HQ (2001). Genetic diversity of agricultural crops germplams in Guangxi. *Acta Bot. Yunnanica* 23: 18-21.
Londo JP, Chiang YC, Hung KH, Chiang TY, et al. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci. U. S. A.* 103: 9578-9583.
Morishima H and Oka HI (1960). The pattern of interspecific variation in the genus *Oryza*: Its quantitative representation by statistical methods. *Evolution* 14: 153-165.
Nakano M, Yoshinura A and Iwata N (1992). Phylogenetic study of cultivated rice and its wild relatives by RFLP. *Rice Genet. Newsl.* 9: 132-134.
Nei M and Kumar S (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.
Nishikawa T, Vaughan DA and Kadowaki K (2005). Phylogenetic analysis of *Oryza* species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. *Theor. Appl. Genet.* 110: 696-705.
Rogers OS and Bendich AJ (1988). Extraction of DNA Plant Tissue, Plant Molecular. In: Plant Molecular Biology Manual (Gelvin SB, Schilpe RA and Verna DS, eds.). Kluwer Academic Publishers, Dordecht, 1-10.
Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
Sato YI and Morishima H (1988). Distribution of the genes causing F2 chlorosis in rice cultivars of the Indica and Japonica types. *Theor. Appl. Genet.* 75: 723-727.
Second G (1982). Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jap. J. Genet.* 57: 25-57.
Shahid MM, Nishikawa T, Fukuoka S, Njenga PK, et al. (2004). The complete nucleotide sequence of wild rice (*Oryza nivara*)

chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. *Gene* 340: 133-139.
Shaw J, Lickey EB, Beck JT, Farmer SB, et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92: 142-166.
Sun CQ, Mao L, Wang ZS, Zhu LH, et al. (1995). A Primary Study of Cultivated Rice (*Oryza sativa*) and Common Chinese Wild Rice (*O. rufipogon*) Using Random Amplified Polymorphic DNA (RAPD). *Chin. J. Rice Sci.* 9: 1-6.
Sun CQ, Wang XK, Yoshimura A and Iwata N (1997). Indica-japonica differentiation of chloroplast DNA in *O. rufipogon* Griff. and *O. sativa* L. *J. Agric. Biotech.* 5: 319-324.
Sun CQ, Wang XK, Li ZC, Yoshimura A, et al. (2001). Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. *Theor. Appl. Genet.* 102: 157-162.
Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
Tang J, Xia H, Cao M, Zhang X, et al. (2004). A comparison of rice chloroplast genomes. *Plant Physiol.* 135: 412-420.
Tian X, Zheng J, Hu S and Yu J (2006). The rice mitochondrial genomes and their variations. *Plant Physiol.* 140: 401-410.
Vaughan DA, Morishima H and Kadowaki K (2003). Diversity in the *Oryza* genus. *Curr. Opin. Plant Biol.* 6: 139-146.
Wang XK and Sun CQ (2000). Origin and evolution of biodiversity and classification of *Oryza sativa* L. *J. Plant Genet. Resour.* 1: 48-53.
Wang XK, Cai HW, Sun CQ, Wang ZS, et al. (1994). The preliminary study on the primitive type of *Oryza rufipogon* Griff. in China and its Hsien-Keng differentiation. *Chin. J. Rice Sci.* 8: 205-210.
Wang ZY and Tanksley SD (1989). Restriction fragment length polymorphism in *Oryza sativa* L. *Genome* 32: 1113-1118.
Yuan LP (1987). The stratagement thinking of hybrid rice breeding. *Hybrid Rice* 2: 1-3.

