Theor Appl Genet (2005) 110: 696–705
DOI 10.1007/s00122-004-1895-2

ORIGINAL PAPER

Tomotaro Nishikawa · Duncan A. Vaughan
Koh-ichi Kadowaki

Phylogenetic analysis of *Oryza* species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes

Received: 25 July 2004 / Accepted: 26 November 2004 /

Abstract Simple sequence repeats (SSR) and their flanking regions in the mitochondrial and chloroplast genomes were sequenced in order to reveal DNA sequence variation. This information was used to gain new insights into phylogenetic relationships among species in the genus *Oryza*. Seven mitochondrial and five chloroplast SSR loci equal to or longer than ten mononucleotide repeats were chosen from known rice mitochondrial and chloroplast genome sequences. A total of 50 accessions of *Oryza* that represented six different diploid genomes and three different allopolyploid genomes of *Oryza* species were analyzed. Many base substitutions and deletions/insertions were identified in the SSR loci as well as their flanking regions. Of mononucleotide SSR, G (or C) repeats were more variable than A (or T) repeats. Results obtained by chloroplast and mitochondrial SSR analyses showed similar phylogenetic relationships among species, although chloroplast SSR were more informative because of their higher sequence diversity. The CC genome is suggested to be the maternal parent for the two BBCC genome species (*O. punctata* and *O. minuta*) and the CCDD species *O. latifolia,* based on the high level of sequence conservation between the diploid CC genome species and these allotetraploid species. This is the first report of phylogenetic analysis among plant species, based on mitochondrial and chloroplast SSR and their flanking sequences.

Communicated by J.S. Heslop-Harrison

T. Nishikawa · D. A. Vaughan · K. Kadowaki (✉)
Genetic Diversity Department,
National Institute of Agrobiological Sciences,
2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan
E-mail: kadowaki@affrc.go.jp
Tel.: +81-29-8387449
Fax: +81-29-8387408

Introduction

Simple sequence repeats [(SSR) also called microsatellites] are regions where mono-, di-, tri-, and tetranucleotide base pairs are multiplied repeatedly. They are well distributed throughout the genomes of most eukaryotic species and are known to be highly variable. Therefore, information from SSR analysis has been widely used to detect polymorphism in the nuclear genome among species. Polymorphism in SSR is generally believed to be the result of replication error (Jarne and Lagoda 1996; Moxon and Wills 1999), and it occurs at a higher rate than the mutation in non-repetitive DNA (Wierdl et al. 1997).

There are three genomes in plant cells: nuclear, chloroplast, and mitochondrial. Microsatellite markers are a powerful tool that have been used to investigate genomic structure, and fine mapping by SSR is now available for the nuclear genomes of some plants (e.g., Akkaya et al. 1992; Chin et al. 1996; Taramino and Tingey 1996; Temnykh et al. 2000). Complete SSR information has been reported for the chloroplast genome of rice (Ishii and McCouch 2000). However, very limited information is available on mitochondrial SSR. Comparative studies of the extent of variability between chloroplast and mitochondrial SSR have not been reported. In addition, the level of variability among SSR sequences and their flanking regions in the chloroplast and mitochondrial genomes have not been examined.

It is reported that nucleotide substitution is lowest in mitochondrial DNA and highest in the nuclear DNA (Wolfe et al. 1987). In this respect, nuclear and cytoplasmic SSR, respectively, are useful molecular markers to determine phylogenetic relationships among closely or less closely related organisms. Maternal inheritance of extra-nuclear genomes has been effectively used to elucidate genetic relationships among species, particularly in relation to the origin of polyploid plants (Nishikawa et al. 2002).

Earlier chloroplast SSR studies have been based on polymorphism detected by silver staining of amplified PCR products that include SSR plus flanking sequences. This method provides relatively good resolution; however, it is difficult to determine (1) whether the size differences are caused by the length of the SSR itself or that of its flanking sequence, and (2) whether there is substitution within the SSR sequence. Chloroplast SSR homoplasy has been reported in the genus *Clusia* (Hale et al. 2004). Therefore, size homoplasy (Doyle et al. 1998), where the same molecular weight contains different internal mutations, should be monitored by sequencing rather than estimation of size on a gel.

Chloroplast SSR have been identified in fully sequenced chloroplast genomes of different plants. However, to the best of our knowledge, mitochondrial SSR have not been studied from completely sequenced mitochondrial genomes. To date, there is only one published report of mitochondrial SSR where a part of the genomic information was used (Soranzo et al. 1999). The complete nucleotide sequence of the mitochondrial genome of higher plants is now available for Arabidopsis (Unseld et al. 1997), sugar beet (Kubo et al. 2000), and rice (Notsu et al. 2002).

Rice is an important cereal worldwide. The genus *Oryza* includes approximately 22 species that have a pan-tropical distribution (Vaughan 1994; Khush 1997). *Oryza* consists of diploid and allotetraploid species. Genome designations have been reported for *Oryza* species based on chromosome pairing at meiosis of F_1 hybrids between different taxa and total genomic DNA hybridization, and these methods revealed that *Oryza* species have AA, BB, CC, EE, FF, or GG genome, and the allotetraploids have BBCC, CCDD, or HHJJ genome. The present study was undertaken to assess the level of genetic variation in the genomes of cytoplasmic organelles based on SSR. SSR were initially surveyed in the rice mitochondrial genome by computer search. Information obtained was used along with the chloroplast SSR already reported for the systematic study of different accessions of genus *Oryza*. The objectives of this investigation were (1) to elucidate phylogenetic relationships among *Oryza* species, based on cytoplasmic organelle genomic markers, and (2) to determine the frequency of nucleotide alteration between SSR and their flanking regions from different organelle genomes.

Materials and methods

Plant materials

A total of 50 accessions representing 21 different *Oryza* species were used in this investigation (Table 1). Two accessions of *Leersia tisserantii* were used as an outgroup for taxonomic analysis.

DNA extraction

Total DNA was extracted from fresh leaves of a single plant by the method of Murray and Thompson (1980).

PCR amplification and subsequent DNA sequence analysis

PCR was performed using KOD-plus-DNA polymerase (Toyobo, Japan), with the following program: 2 min at 94°C followed by 30 cycles for 15 s at 94°C, 30 s at 50°C, 1 min at 68°C. Amplified products were purified by Suprec-02 (TaKaRa Bio, Japan) and the sequence was determined using the protocol of CEQ DTCS Quick Start Kit (Beckman Coulter, USA).

SSR sequence analysis

In order to exclude ambiguity resulting from size homoplasy, we performed DNA sequence analysis for the 12 SSR regions—seven from the mitochondria and five from the chloroplast—to evaluate the exact numbers of SSR along with the flanking sequence (Table 2). The sequences of *O. sativa* cv. Nipponbare were used as a standard for SSR and their flanking region nucleotide sequences for both chloroplast and mitochondrial SSR. Phylogenetic analysis of the genus *Oryza* was conducted based on DNA sequences obtained.

Data analysis

Sequence from *O. sativa* cv. Nipponbare is standard in this paper (mitochondrial genome is AB076665 and AB076666, chloroplast genome is X15901). The sequence matrix was obtained by multiple alignment, using Clustal X, version 1.81 (Thompson et al. 1997). Phylogenetic analysis was performed by the maximum parsimony phylogenetic estimation, using PAUP*, version 4.0b10 (Swofford 1998). The most parsimonious tree was obtained using the heuristic search option involving 100 replications of random addition sequence and tree bisection–reconnection (TBR) branch swapping. All characters were specified as unweighted. Variation in the repeat length was handled as an independent character, and all gaps were handled as insertion/deletion events. A strict consensus tree was computed from all obtained trees. The tree was unrooted, and rooted using the outgroup method. Bootstrap analyses (Felsenstein 1985) were conducted for assessing the reliability of the tree. One thousand replications were calculated using the heuristic search option with TBR branch swapping and random sequence–addition sequence algorithms.

Table 1 List of genus *Oryza* taxa analyzed in this study

Name of species	Accession no.[a]	Origin	Genome
Genus *Oryza*			
Section *Oryza*			
O. *sativa* complex			
O. *sativa* L. cv. Nipponbare	AA		
O. *sativa* L. cv. Kasalath	AA		
O. *nivara* Sharma et Shastry	SL10	Sri Lanka	AA
O. *rufipogon* Griff.	W0120	India	AA
O. *rufipogon* Griff.	W0157	India	AA
O. *rufipogon* Griff.	W2078	Australia	AA
O. *rufipogon* Griff.	W2099	Australia	AA
O. *glaberrima* Steud.	Tog5923	Nigeria	AA
O. *barthii* A. Chev.	W0652	Sierra Leone	AA
O. *longistaminata* Chev. et. Roehr.	W1421	Mali	AA
O. *longistaminata* Chev. et. Roehr.	W1499	Unknown	AA
O. *longistaminata* Chev. et. Roehr.	W1650	Tanzania	AA
O. *meridionalis* Ng	W1625	Australia	AA
O. *meridionalis* Ng	W1635	Australia	AA
O. *glumaepatula* Steud.	W1169	Cuba	AA
O. *glumaepatula* Steud.	W1183	Guyana	AA
O. *glumaepatula* Steud.	W2199	Brazil	AA
O. *glumaepatula* Steud.	103812	Venezuela	AA
O. *officinalis* complex			
O. *officinalis* Wall ex Watt	W0002	Thailand	CC
O. *officinalis* Wall ex Watt	W0614	Myanmar	CC
O. *officinalis* Wall ex Watt	NIAS WBC6	Unknown	CC
O. *minuta* Presl. Et Presl.	W1213	Philippines	BBCC
O. *minuta* Presl. Et Presl.	W1337	Philippines	BBCC
O. *rhizomatis* Vaughan	2001-SL3	Sri Lanka	CC
O. *eichingeri* Peter	W1525	Uganda	CC
O. *eichingeri* Peter	2001-SL8	Sri Lanka	CC
O. *punctata* Kotschy ex Steud.	W1514	Kenya	BB
O. *punctata* Kotschy ex Steud.	103887	Tanzania	BB
O. *punctata* Kotschy ex Steud.	104154	Cameroon	BB
O. *punctata* Kotschy ex Steud.	W0015	Unknown	BBCC
O. *punctata* Kotschy ex Steud.	W0043	Unknown	BBCC
O. *punctata* Kotschy ex Steud.	W1564	Nigeria	BBCC
O. *latifolia* Desv.	W0048	Paraguay	CCDD
O. *latifolia* Desv.	W1181	Panama	CCDD
O. *alta* Swallen	W0018	Paraguay	CCDD
O. *alta* Swallen	105138	Surinam	CCDD
O. *grandiglumis* (Doell) Prod.	W0613	Brazil	CCDD
O. *grandiglumis* (Doell) Prod.	W1480(B)	Brazil	CCDD
O. *australiensis* Domin	W1639	Australia	EE
O. *australiensis* Domin	101397	Unknown	EE
O. *australiensis* Domin	101410	Australia	EE
O. *australiensis* Domin	103318	Australia	EE
O. *australiensis* Domin	105267	Australia	EE
Section Ridleyanae Tateoka			
O. *ridleyi* complex			
O. *ridleyi* Hook. f.	W0001	Thailand	HHJJ
O. *ridleyi* Hook. f.	W0604	Malaya	HHJJ
O. *longiglumis* Jansen	W1220	Indonesia	HHJJ
O. *longiglumis* Jansen	W1228	Indonesia	HHJJ
Section Granulata Roschev.			
O. *meyeriana* complex			
O. *granulata* Nees et Arn. ex Watt.	W0067(B)	Thailand	GG
O. *granulata* Nees et Arn. ex Watt.	W0022	Unknown	GG
Section Brachyantha B. R. Lu			
O. *brachyantha* Chev. et Roehr.	101232	Sierra Leone	FF
Genus *Leersia*			
L. *tisserantii* (A. Chev.) Launert	W1345	Unknown	
L. *tisserantii* (A. Chev.) Launert	W1620	Unknown	

[a] *W*- numbers from NIG; six-digit numbers from IRRI; Tog, NIAS, and *SL*- numbers from NIAS

Table 2 List of organelle single sequence repeat (*SSR*) regions and primer pairs analyzed in this study

Locus	Location[a] (gene)	Repeat[b] (5′–3′)	Primer sequence	Size (bp) in *O. sativa* cv. Nipponbare
Mitochondrial SSR				
RMt3	Intergenic region	$(T)_{10}$	GGTCCGCCTTTCTCTACTAT TGTCTTTCTTGCATACATCG	447
RMt9	Coding region (*rps* 2)	$(T)_{12}$	ACTGCTTTTAAGCCTGTTTG TAGGATCTCCCATTCGTAAA	451
RMt10	Intron (*nad* 4)	$(G)_{12}$	TAACGGCTACAAGGGATAAA GTGATGTGAGCGGTTCTATT	455
RMt11	Intron (*cox* II)	$(T)_{10}$	CCAAGAGAGGACAACCTGT ATTCCTCACCTATCCTGTCA	461
RMt12	Intergenic region	$(T)_{10}$	CAAACTCAAGATGCCTATCC CAACCCGGAATATTGATTTA	436
RMt15	Intergenic region	$(A)_{10}$	GATAAAAGATGATCCCCACA AGTCCTTTTTCTGCTTGTTG	452
RMt17	Intergenic region	$(G)_{11}$	GAAAGCACCCTCTTTTTGTA CGTGGTTATCTGAAGTGGAT	460
Chloroplast SSR				
RCt3[c]	Intergenic region	$(A)_{10}$	GGGTGTAGGTAGGGCTAAAA GACTTTTGAAAATGCGAAAT	448
RCt5[c]	Intergenic region	$(T)_{10}$	TTCCTACGTGAACCAATTTT TTCAAAGGGTTAGGTTTTTCT	450
RCt6[c]	Intergenic region	$(A)_{10}$	CTTTGTTTATGCTTCGGATT GTTCGCCTAGAGAATGACAC	437
RCt10[c]	Intergenic region	$(T)_{10}$	AAACATATGCGGATCAAATC CAACACAACATAGGTCATCG	451
RCt12[c]	Intergenic region	$(A)_{10}$	TGAAGGAGGAGAAAGAAACA TGATATCATCAACCGTGCTA	449

[a] Information on location was from Notsu et al. (2002) for mitochondrial SSR, and Ishii and McCouch (2000) for chloroplast SSR

[b] Repeat length in Nipponbare

[c] Locus names followed by Ishii and McCouch (2000)

Results and discussion

SSR in rice mitochondrial genome

The SSR in the entire mitochondrial genome of *O. sativa* cv. Nipponbare were screened using the computer software Genetyx (Genetyx, Japan). Although there were many SSR of different sizes, SSR sequences equal or longer than ten uninterrupted base repeats were used to develop primers. There are 18 mononucleotide, 19 dinucleotide, and 21 trinucleotide repeats in the rice mitochondrial genome (Table 3). Of these 58 SSR, six were identified in the protein coding regions, five in introns, and 47 in intergenic regions. The longest mononucleotide repeat was 13 bases, and it was identified in the coding region of the *rps1* gene. Mononucleotide, dinucleotide, and trinucleotide repeats were distributed throughout the entire mitochondrial genome of rice, and no densely populated pockets of specific nucleotide repeats were observed.

Regarding the distribution of SSR in other plant mitochondrial genomes, a computer search was conducted using the published databases from liverwort, Arabidopsis, and sugar beet. Mononucleotide repeats longer than nine units were identified in these genomes and are shown for comparison with rice SSR (Table 4). Most of the repeats were in different positions, indicating that SSR were generated independently during the course of flowering plant evolution and subsequently diverged due to lack of sequence conservation in function.

Diversity of SSR sequences in the mitochondrial genome of genus *Oryza*

Although 18 mitochondrial mononucleotide repeat SSR loci were identified, only 14 were unique loci after elimination of duplicated loci. After examining the 14 mitochondrial loci in different accessions, DNA fragments could be amplified and sequenced from only seven loci. A possible explanation for why the other seven loci could not be amplified is that recombination events might have happened in the flanking region of the mitochondrial SSR and disrupted sequence structure. Dynamic recombination events have been reported in the mitochondrial genome of closely related species (Palmer and Herbon 1988). The seven amplified loci were used for further analysis. The DNA sequence of each of the amplified products was determined. All the SSR loci of mononucleotide repeats showed polymorphism among the species of genus *Oryza*, although the extent of polymorphism differed depending on the amplified region (Table 5). The highest polymorphism was observed in RMt17, for which repeats ranged from 8 to 14. RMt17 is located in the intergenic region of *nad9* and *trnY*. In contrast, RMt9 located in coding region of *rps2* showed no length polymorphism.

The range of G repeats in RMt17 was from 8 to 14 whereas it was from 7 to 12 in the case of RMt10. In RMt15 repeat of A ranged from 9 to 11, while in RMt3, RMt11, and RMt12, T repeats ranged from 9 to 12. Overall, variation in G repeat is higher than the other nucleotide repeats, suggesting that a G/C repeat is less stable than T/A-repeat.

Table 3 Mitochondrial SSR equal to or larger than 10 bp in *Oryza sativa*

	Mononucleotide repeat		Dinucleotide repeat		Trinucleotide repeat	
	Number of loci	Max. size (bp)	Number of loci	Max. size (bp)	Number of loci	Max. size (bp)
Coding region	3	13	1	10	2	12
Intron	2	12	2	10	1	12
Intergenic region	13	11	16	12	18	12

Table 4 Mitochondrial SSR differentiation among four plants

	Rice		Liverwort		Arabidopsis		Sugarbeet	
	Number of loci	Max. size (bp)	Number of loci	Max. size (bp)	Number of loci	Max. size (bp)	Number of loci	Max. size (bp)
A/T	15	13	33	17	16	14	36	13
G/C	3	12	8	12	7	12	1	10

Identification of size homoplasy caused by SSR and its flanking sequence

In addition to the seven mitochondrial loci, five chloroplast SSR loci from the rice chloroplast genome were used (Ishii and McCouch 2000). The twelve SSR loci, seven from the mitochondrial genome and five from chloroplast genome were sequenced (DDBJ/ EMBL/ GenBank accession nos. AB128166-AB128767) and used to examine intra- and inter-specific variation in accessions of the genus *Oryza* (Table 2).

RMt9 had a repeat length of 12 T nucleotides, which were present in the coding region of the *rps2* gene. Examination of the locus RMt9 revealed no variation in its length in any accessions. This is probably due to this locus being present in the coding region, and a change in the repeat number could have a deleterious affect. Addition or deletion of three nucleotides that result in addition or deletion of an amino acid without disturbing the sequence of others is unlikely. Maintaining 12 nucleotides in the SSR of the coding region seems to be strongly favored compared to a repeat in the intergenic region, even though substitution of nucleotides in the SSR of any region is possible in theory. One to three substitutions of A in the 12 T repeats of RMt9 were found in the accessions examined (Table 6). Thus, the length of sequence corresponding to this SSR region was the same, but its sequence was altered, indicating the presence of size homoplasy at this locus.

The opposite situation was observed where the SSR sequence remained the same, but the length of sequence amplified differed. Deletion or insertion was found in the flanking sequence of the SSR, resulting in size differences of the amplified DNA sequence. Hence, interpretation of whether polymorphism is caused by the SSR itself or its flanking sequence requires sequence analysis. Such deletion/insertion events were identified in the sequence of the flanking region of RCt3, RCt5, RCt6, RCt12, RMt3, RMt9, RMt10, RMt11, RMt12, and RMt17 (Table 7).

Most previous studies of length polymorphism in SSR sequences are based on the hypothesis that size differences are caused mainly by SSR length polymorphism. Our results show that there are several insertions and/or deletions in the flanking regions in addition to the SSR polymorphism. Thus, length polymorphism is a combination of variation in SSR as well as in the flanking sequence. Similar results are observed in the chloroplast SSR of the genus *Clusia* (Hale et al. 2004).

RCt5 and RCt6 have T and A repeats, respectively, and they are only 342 bp apart. The length of the T repeat of RCt5 ranged from seven to ten bases, whereas the length of the A repeat of RCt6 ranged from six to ten. A combination of different units of A and T were also identified in accessions of *Oryza* examined. For example, *O. australiensis* (EE genome) had eight Ts and seven As, whereas *O. granulata* (GG genome) had seven Ts and eight As in the RCt5 and RCt6 locus, respectively. The reference plant *O. sativa* cv. Nipponbare had ten Ts and ten As, respectively, at both these loci, suggesting that length mutation at the two sites occurred independently. This confirms that each of SSR loci represents a single "locus" for taxonomic analysis.

Phylogenetic trees for *Oryza*, based on SSR and flanking region sequences

A phylogenetic tree was constructed separately based on three kinds of information: (1) SSR region (data not shown), (2) flanking sequence to the SSR (data not shown), and (3) the whole SSR and flanking region sequence amplified (Fig. 1). The results are in close agreement with previous studies. Separate dendrograms were constructed from SSR and flanking sequences. The dendrogram based on the chloroplast SSR was similar to that of the chloroplast flanking sequences although the latter was more informative.

The level of nucleotide variation per loci is 20% higher in the chloroplast than the mitochondrial sequences across accessions of *Oryza* analyzed. This supports the view that the rate of sequence mitochondrial substitution is slower than that of the chloroplast (Wolfe et al. 1987).

Table 5 Organelle SSR polymorphism in rice

Species	Accession	Loci[a]											
		RMt3	RMt9	RMt10	RMt11	RMt12	RMt15	RMt17	RCt3	RCt5	RCt6	RCt10	RCt12
O. sativa cv. Nipponbare		10	12	12	10	10	10	11	10	10	10	10	10
O. sativa cv. Kasalath		10	12	12	10	10	10	11	12	10	9	10	12
O. nivara	SL10	10	12(1)	12	10	10	10	11	10	10	9	10	11
O. rufipogon	W0120	10	12	12	10	10	10	11	11	9	9	10	12
O. rufipogon	W0157	10	12	12	10	10	10	11	10	10	9	10	11
O. rufipogon	W2078	10	12(1)	11	10	10	10	11	10	8	9	10	10
O. rufipogon	W2099	10	12(1)	12	10	10	10	11	10	8	9	10	10
O. glaberrima	Tog5923	10	12(1)	12	10	10	10	11	10	9	9	11	11
O. barthii	W0652	10	12(1)	12	10	10	10	11	10	9	9	11	11
O. longistaminata	W1421	10	12(1)	11	11	10	10	11	9	8	9	10	13
O. longistaminata	W1499	10	12(1)	12	11	10	10	11	9	8	9	10	13
O. longistaminata	W1650	10	12(1)	11	10	10	10	12	9	8	9	10	13
O. meridionalis	W1625	10	12(1)	12	10	10	10	11	11	8	9	10	10
O. meridionalis	W1635	10	12(1)	12	10	10	10	11	10	8	9	10	12
O. glumaepatula	W1169	10	12(1)	12	10	10	10	11	10	9	9	10	10
O. glumaepatula	W1183	10	12(1)	11	11	10	10	11	9	8	9	10	12
O. glumaepatula	W2199	10	12(1)	12	11	10	10	11	9	8	9	10	12
O. glumaepatula	103812	10	12(1)	12	10	10	10	11	10	9	9	11	10
O. officinalis	W0002	10	12(2)	9	9	10	11	14	9	9	9	9	14
O. officinalis	W0614	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. officinalis	NIAS WBC6	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. minuta	W1213	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. minuta	W1337	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. rhizomatis	2001-SL3	10	12(2)	8	9	10	11	13	9	9	8	9	14
O. eichingeri	W1525	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. eichingeri	2001-SL8	10	12(2)	8	9	10	11	13	9	9	9	9	13
O. punctata	W1514	10	12(2)	8	9	11	10	del	9	8	9	9	9
O. punctata	103887	10	12(3)	8	9	11(1)	10	del	9	8	9	9	9
O. punctata	104154	10	12(2)	8	9	11(1)	10	del	11	8	9	9	9
O. punctata	W0015	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. punctata	W0043	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. punctata	W1564	10	12(2)	8	9	10	11	14	9	9(s)	9	9	13
O. latifolia	W0048	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. latifolia	W1181	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. alta	W0018	10	12(2)	8	9	10	11	13	9	9	9	9	14
O. alta	105138	10	12(2)	8	9	11	10	14	10	9	9	9	11
O. grandiglumis	W0613	10	12(2)	8	9	11	10	14	11	9	10	9	11
O. grandiglumis	W1480B	10	12(2)	8	9	11	10	14	11	9	10	9	11
O. australiensis	W1639	11	12(2)	8	10	10	9	11	11	8	7	9	14
O. australiensis	101397	11	12(2)	8	10	10	9	11	11	8	7	9	14
O. australiensis	101410	11	12(2)	8	10	10	9	11	11	8	7	9	14
O. australiensis	103318	11	12(2)	8	10	10	9	11	11	8	7	9	14
O. australiensis	105267	11	12(2)	8	10	10	9	11	11	8	7	9	14
O. ridleyi	W0001	10	12(2)	7	12	–	10	8	7	7	6	7	9
O. ridleyi	W0604	10	12(2)	7	12	–	11	8	7	7	6	7	9
O. longiglumis	W1220	10	12(1)	7	12	–	11	8	7	7	6	7	9
O. longiglumis	W1228	10	12(1)	7	12	–	10	8	7	7	6	7	9
O. granulata	W0022	9	12(3)	7	9	–	9	11	8	7	8	7	8
O. granulata	W0067	9	12(3)	7	9	–	10	11	8	7	8	7	8
O. brachyantha	101232	10	12(2)	7	9	–	–	8	9	8(1)	7	7	8
L. tisserantii	W1345	10(1)	12	7	10	–	11	14	7	7	9	8	8
L. tisserantii	W1620	10(1)	12	7	10	–	10	13	7	7	9	8	8

[a]Number in *parentheses* indicates a site of substitution on SSR, *del* indicates deletion event of SSR, *dash* means no amplification

Results deduced from the amplified chloroplast SSR plus flanking region sequence, and the amplified mitochondrial SSR plus flanking region sequence were compared. Both showed a similar pattern, although the level of variation was greater in the chloroplast than the mitochondrial SSR sequences. Regarding the position of GG genome in the dendrogram, results obtained from mitochondrial SSR analysis were different from the results on chloroplast SSR analysis.

Phylogenetic relationships among genus *Oryza* species

Based on the complete data set of DNA sequences from the chloroplast and mitochondrial SSR regions analyzed, the most parsimonious tree was obtained (Fig. 1). The overall result is in close agreement with current understanding of *Oryza* taxonomy. However, results shed new light on species relationships within and among species complexes.

Table 6 Internal substitutions on SSR in RMt9

Substitution	Sequence	Genome type
0	5′-TTTTTTTTTTTT-3′	AA[a]
1	5′-TTTTTTTTTATT-3′	AA, HHJJ[b]
2	5′-TTTTATTTTATT-3′	BB[c], CC, EE, FF, BBCC, CCDD, HHJJ
3	5′-TTTGATTTTATT-3′	BB, GG

[a]AA includes *O. sativa* and two accessions of *O. rufipogon*
[b]HHJJ includes *O. longiglumis*
[c]BB includes two accessions of *O. punctata*

O. sativa complex specific features

The *O. sativa* complex is monophyletic based on the 18 AA genome accessions analyzed. SSR sequences were generally conserved, with a few exceptions despite the different origins of accessions. The G repeat of RMt17 was 12 bases long in *O. longistaminata* (W1650), whereas the rest of AA genome accessions had 11 repeats. In RMt10, AA genome accessions had 10–12 repeats of G, and these numbers are only found in AA genome accessions analyzed. Mitochondrial SSR are more highly conserved among the AA genome accessions, suggesting they have spread pan-tropically more recently than species of the *O. officinalis* complex (see below). Two *O. rufipogon* accessions, W2078 and W0157, showed 11 and 12 G-repeats, respectively, in RMt10, indicating genetic diversification within the same species. Nucleotide sequence 14 bp downstream of SSR region is 5′-TGAACATGATGAACATG-3′ in RMt3. There is a tandem repeat at an 8-bp motif (5′-TGAACATG-3′) in the 17-bp sequence, and there is an A between the two TGAACATGs. A 9-bp deletion was observed in all accessions having the AA and BB (diploid) genomes, and the sequence deleted was the 8-bp repeat and the A. Therefore, it is likely that the tandem repeat of the 8-bp motif (5′-TGAACATG-3′) may have been involved in this deletion event.

O. glumaepatula consists of two groups showing a paraphyletic relationship (Fig. 1). Our results supports earlier research that the AA genome *Oryza* in Latin America consists of at least two groups one that is indigenous, *O. glumaepatula*, and one that is introduced, probably in historic times from Asia, *O. rufipogon* (Juliano et al. 1998; Vaughan et al. 2003).

O. officinalis complex specific features

O. australiensis (EE genome) has a paraphyletic position in relation to *O. punctata* (BB genome) and other species of the *O. officinalis* complex (CC, CCBB, and CCDD genome species). The results support earlier studies that show BB and EE genome species are well diverged from other species of the *O. officinalis* complex (Dally and Second 1990). Our results also suggest that chloroplast genome of *O. punctata* (BB genome) in the *O. officinalis* complex is more closely related to AA genome species than most of the other *O. officinalis* complex species.

Origin of the maternal parent for allotetraploid

Mitochondrial and chloroplast genomes are usually maternally inherited, so results provide evidence of the donors of the maternal genome to allotetraploid species. A small fragment located 35 bp upstream of a repeat that consists of eight nucleotides (5′-CATTCTAT-3′) in RCt3, was absent in all accessions of BBCC, CC, and CCDD genome accessions. Of these BBCC, CC, and CCDD accessions, overall DNA sequence similarity was 99.7% (4,039 nucleotides out of 4,050 nucleotides). These results strongly suggest that *Oryza* with the CC

Table 7 Length mutations in flanking region of SSR

Location	Mutation	Region	Size (bp)	Genome type
RMt3	Insertion	Downstream	9	CC, EE, FF, GG, BBCC, CCDD, HHJJ
	SSR[a]	Downstream		Not specific
RMt9	Insertion	Upstream	36	AA (except for *O. sativa* cv. Nipponbare), FF
	Insertion	Downstream	9	GG
RMt10	Deletion	Downstream	7	FF
	Insertion	Downstream	3	FF
RMt11	SSR[a]	Downstream		Not specific
	Deletion	Downstream	9	GG
	Deletion	Downstream	8	GG
RMt12	Deletion	Upstream	10	AA (only one accession of *O. rufipogon*)
	Deletion	Upstream	11	EE
RMt17	Deletion	Upstream to downstream	171	BB
RCt3	Deletion	Upstream	8	CC, BBCC, CCDD
	Deletion	Downstream	39	AA (only one accession of *O. glumaepatula*)
	Deletion	Downstream	5	CC, EE, FF, GG, BBCC, CCDD, HHJJ
RCt5	SSR[a]	Upstream		Not specific
RCt6	Insertion	Upstream	1	HHJJ
RCt12	Insertion	Upstream	6	FF

[a]*SSR* indicates simple sequence repeat newly identified in addition to SSR in Table 2

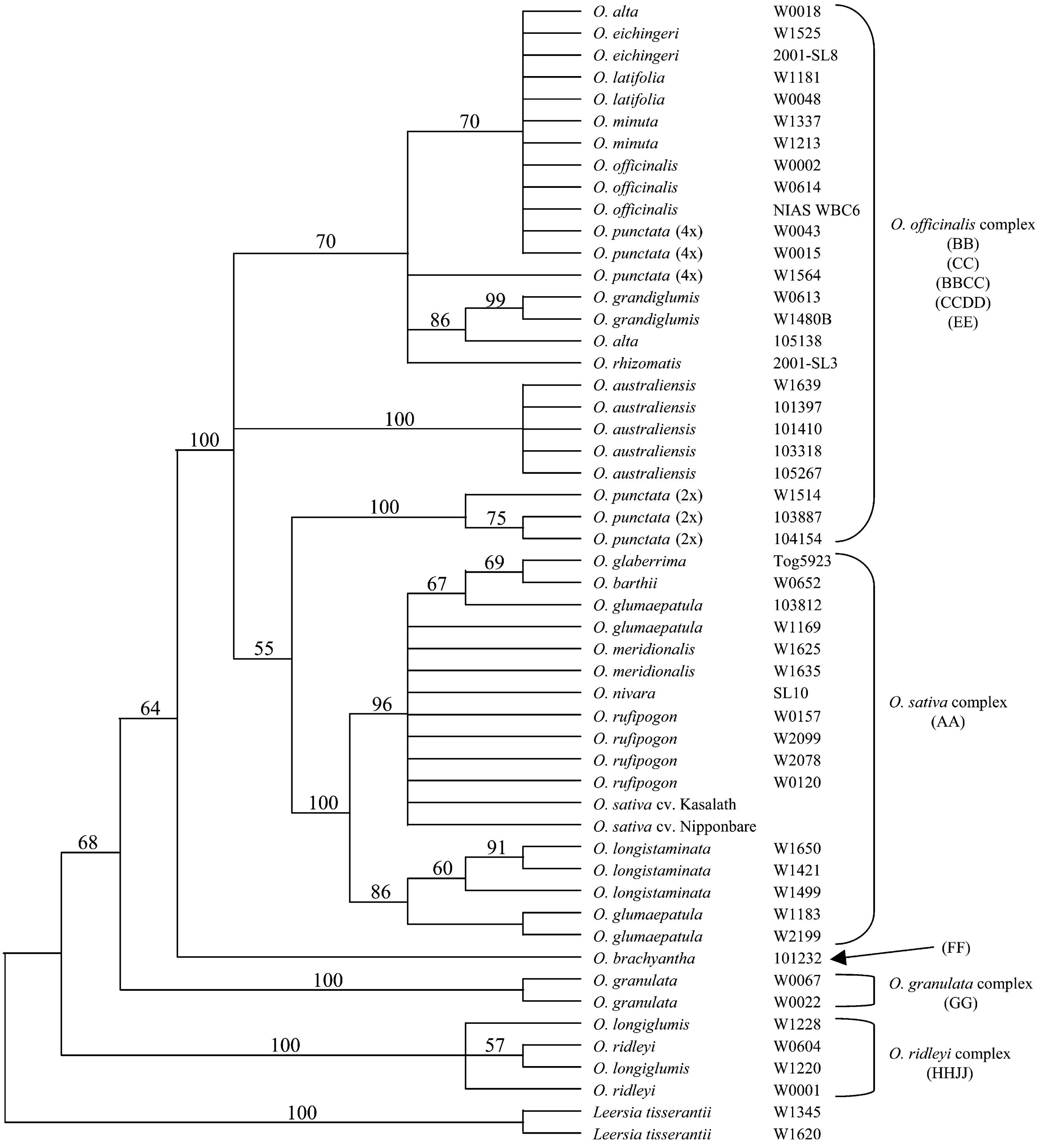

Fig. 1 Phylogenetic tree of the genus *Oryza* based on 12 organelle single sequence repeats and their flanking sequences, using the maximum parsimony method. The tree is a strict consensus of the 4,394 most parsimonious trees. Tree length = 356, consistency index = 0.6657, retention index = 0.9043, rescaled consistency index = 0.6020. The *numbers above the nodes* represent bootstrap values expressed as percentage of 1,000 bootstrap replications

genome was the maternal parent for *O. minuta* (W1213 and W1337) and *O. punctata* (W0015 and W0043) of BBCC genome as well as for *O. latifolia* (W0048 and W1181) of CCDD genome.

Our results do not support earlier results that have suggested that *O. punctata* (BB) is the maternal parent of *O. minuta* (Dally and Second 1990; Kanno and Hirai 1992). Our results clearly demonstrated that *O. minuta*

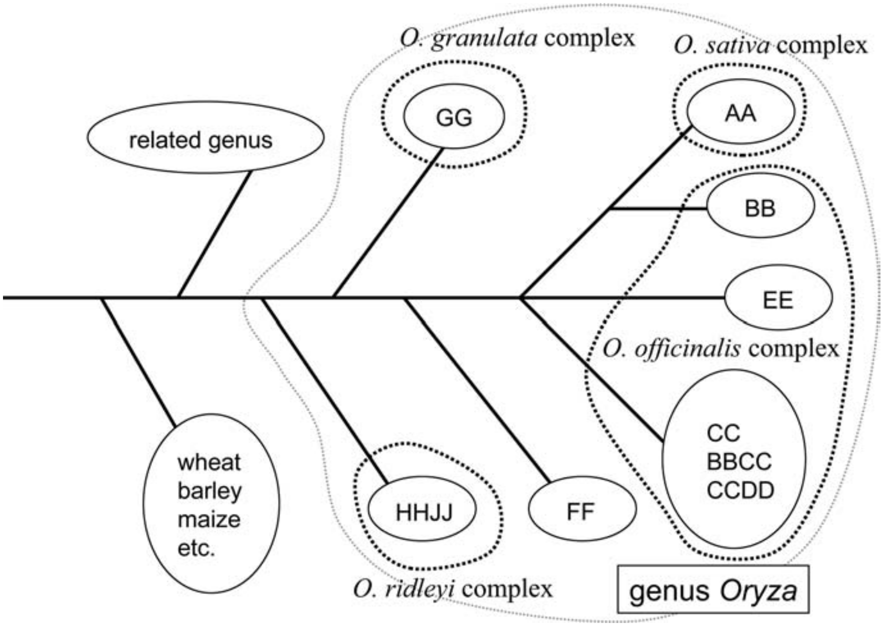

Fig. 2 A model for evolution of cytoplasmic genome in *Oryza* species based on Fig. 1

(BBCC genome) is in the same clade as CC genome accessions and has a paraphyletic relationship to *O. punctata* (BB genome). It seems necessary to check a large number of *O. minuta* accessions to determine whether the discrepancy in results between this study and previous studies results here from using different germplasm accessions.

On the other hand, our results do support earlier reports that a diploid CC genome species was the maternal parent of *O. punctata* (BBCC) of Africa (Dally and Second 1990; Kanno and Hirai 1992). Among diploid CC genome species, only *O. eichingeri* (CC genome) is found in Africa in shaded habitats. Based on this study, it is likely that an ancestral form of *O. eichingeri* was the maternal parent of *O. punctata* (BBCC).

Previous studies have shown that the three CCDD genome species of Latin America, *O. alta*, *O. grandiglumis*, and *O. latifolia*, are all very closely related and could even be considered one complex species (Vaughan et al. 2003). This study shows that *O. latifolia* has different chloroplast genome from *O. alta* and *O. grandiglumis*. Since a diploid DD genome species has not been found, it is not possible to know whether *O. alta* and *O. grandiglumis* have received their chloroplast genomes from a different polyploid event with CC and DD genome species or has diverged after the polyploid event that gave rise to these three species. Further studies of the chloroplast genomes of the Latin American allotetraploid species are necessary.

O. granulata complex, *O. ridleyi* complex, and *O. brachyantha*

O. brachyantha, *O. granulata*, and species of the *O. ridleyi* complex all form distinct clades (Fig. 1), as would be expected from species in different sections of the genus *Oryza*. The relationship among accessions of *O. ridleyi* and *O. longiglumis* of the *O. ridleyi* complex suggest that these two closely related species might have diverged quite recently.

In summary, the phylogenetic relationships among species of the genus *Oryza*, based on analysis of SSR and their flanking region sequences from cytoplasmic DNA, has shed new light on *Oryza* species relationships. Based on the results, a model phylogenetic tree was constructed for evolution of the cytoplasm genome in genus *Oryza* (Fig. 2). In addition, the approach used here has given new insights into the origin of *Oryza* allopolyploid species and should be extended to check the maternal origin of *O. malampuzhaensis* (BBCC). This is the first report in plants to describe phylogeny based on mitochondrial and chloroplast SSR and their flanking sequences.

Acknowledgements DNA from four *O. australiensis* (EE genome) and *O. brachyantha* (FF genome) were kindly provided by Dr. T. Ishii of Kobe University, Japan. We thank Drs. Masood and Ishii for their valuable comments on our results. We thank Ms. Nohara for her technical assistance.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Chin EC, Senior ML, Shu H, Smith JS (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866–873

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

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *glycine*). Mol Biol Evol 15:215–218

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

Hale ML, Borland AM, Gustafsson MH, Wolff K (2004) Causes of size homoplasy among chloroplast microsatellites in closely related *Clusia* species. J Mol Evol 58:182–190

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theor Appl Genet 100:1257–1266

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends in Ecol Evol 11:424–429

Juliano AB, Naredo MEB, Jackson MT (1998) Taxonomic status of *Oryza glumaepatula* Steud. I. Comparative morphological studies of New World diploids and Asian AA genome species. Genet Res Crop Evol 45:197–203

Kanno A, Hirai A (1992) Comparative studies of the structure of chloroplast DNA from four species of *Oryza*: cloning and physical maps. Theor Appl Genet 83:791–798

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35:25–34

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $\mathrm{tRNA}^{\mathrm{Cys}}$ (GCA). Nucleic Acids Res 28:2571–2576

Moxon ER, Wills C (1999) DNA microsatellites: agents of evolution? Sci Am 280:94–99

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

Nishikawa T, Salomon B, Komatsuda T, von Bothmer R, Kadowaki K (2002) Molecular phylogeny of the genus *Hordeum* using three chloroplast DNA sequences. Genome 45:1157–1166

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics 268:434–445

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol 28:87–97

Soranzo N, Provan J, Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. Genome 42:158–161

Swofford DL (1998) PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39:277–287

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet 15:57–61

Vaughan DA (1994) The wild relatives of rice. International Rice Research Institute, Manila

Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. Curr Opin Plant Biol 6:139–146

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769–779

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058