Theor Appl Genet (2000) 100:1257–1266 © Springer-Verlag 2000

T. Ishii · S.R. McCouch

Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species

Received: 17 March 1999 / Accepted: 11 November 1999

Abstract Primer pairs flanking ten chloroplast microsatellite loci, originally identified in *Oryza sativa* cv Nipponbare, were evaluated for amplification and allelic diversity using a panel of 13 diverse cultivars of rice *(O. sativa)*, 19 accessions of wild rice (three *O. officinalis*, five *O. latifolia*, five *O. minuta*, four *O. australiensis*, one *O. brachyantha* and one *O. ridleyi*) and eight other Gramineae species (maize, teosinte, wheat, oat, barley, pearl millet, sorghum and sugarcane). Amplified products were obtained for all samples at nine out of ten loci. Among the rice cultivars, the number of alleles per locus ranged from one to four, with monomorphic patterns observed at five loci. The average polymorphism information content (PIC) value at the other five (polymorphic) loci was 0.54 among the 13 cultivars. When wild rice and the other Gramineae species were compared based on the proportion of shared alleles, their phylogenetic relationships were in agreement with previous studies using different types of markers; however, the magnitude of the differences based on chloroplast microsatellites underestimated the genetic distance separating these divergent species and genera. A sequence-based comparison of homologous regions of the rice and maize chloroplast genomes revealed that, while a high level of microsynteny is evident, the occurrence of actively evolving microsatellite motifs in specific regions of the rice chloroplast genome appears to be mainly a species or genome-specific phenomenon. Thus the chloroplast primer pairs used in this study bracketed mutationally active microsatellite motifs in rice but degenerate, interrupted motifs or highly conserved, mutationally inert motifs in distantly related genera.

Communicated by K. Oono

T. Ishii · S.R. McCouch (✉)
Department of Plant Breeding, 252 Emerson Hall,
Cornell University, Ithaca, NY 14853-1901, USA
Fax: +1-607-255-0420, e-mail: srm4@cornell.edu

Present address: T. Ishii
Laboratory of Plant Breeding, Faculty of Agriculture,
Kobe University, Nada-ku, Kobe 657-8501, Japan

Key words Chloroplast microsatellite · Simple sequence length polymorphism (SSLP) · Allelic diversity · Rice *(Oryza sativa)* · Gramineae

Introduction

Microsatellites consist of tandem arrays of simple sequence repeats (SSRs), such as mono-, di-, tri- and tetranucleotide motifs. Simple sequence length polymorphisms (SSLPs), caused by variation in the number of repeat units, can be easily detected by PCR using pairs of primers designed from unique sequences flanking the microsatellite-containing regions. Microsatellites are widely used as genetic markers because they are abundant and well-distributed throughout the genomes of most eukaryotic species and because they have a high level of allelic diversity. Well-saturated microsatellite maps have been developed for human (Dib et al. 1996) and mouse (Dietrich et al. 1996), and are currently under development in several plant species, including rice (Temnykh et al. 2000; Chen et al. 1997; Akagi et al. 1996), maize (Senior and Heun 1993), barley (Becker and Heun 1995), wheat (Roder et al. 1998), *Arabidopsis* (Bell and Ecker 1994) and soybean (Akkaya et al. 1995). These studies indicate that (CA/GT)*n* dinucleotide repeats are relatively abundant in mammalian genomes while (GA/CT)*n* and (AT/TA)*n* are most abundant in plant genomes. Microsatellite markers are also a powerful tool for analyzing intraspecific variation and for DNA fingerprinting (Powell et al. 1996; Akagi et al. 1997; Olufowote et al. 1997). Application of SSLPs in phylogenetic studies involving different species or genera is more difficult because the sequence divergence that has accumulated over evolutionary time affects the primer-annealing regions as well as the length of the SSR motifs themselves.

SSLP analysis has targeted the nuclear genome for more than 10 years, but more recently, microsatellite markers for chloroplast genomes have been identified through computer searches of the GenBank database

(Powell et al. 1995a). A total of 237 and 229 microsatellites having more than ten repeats each were found in complete chloroplast sequences of six plant species and in partial sequences of other species, respectively. The size of the chloroplast genome in plants is much smaller than that of the nuclear genome and the AT content of chloroplast DNA is much higher. Most chloroplast microsatellites are (A/T)*n* mononucleotide repeats. Using these chloroplast microsatellites, intra- and inter-specific variations were examined in pine trees (Powell et al. 1995a), soybean (Powell et al. 1995b, 1996) and rice (Provan et al. 1996, 1997). Since the chloroplast genome is more conserved than the nuclear and mitochondrial genomes (Wolfe et al. 1987), chloroplast microsatellites have the potential to illuminate phylogenetic studies across greater genetic distances than nuclear or mitochondrial microsatellites.

The genus *Oryza* contains two cultivated species and more than 20 wild species. Their genome constitutions are AA, BB, BBCC, CC, CCDD, EE and FF, as well as some unidentified genomes (Vaughan 1989). Phylogenetic studies using these species have been limited to RFLP analyses of nuclear and chloroplast DNA (Dally and Second 1990; Wang et al. 1992) and AFLP analysis (Aggarwal et al. 1999). However, the entire chloroplast genome of *O. sativa* Japonica cv Nipponbare was sequenced by Hiratsuka et al. (1989), and 12 SSRs were identified in that sequence by Provan et al. (1996). Using seven of these chloroplast microsatellites, Provan et al. (1997) analyzed within-genome variation in *Oryza*.

Several major crop species belong to the family Gramineae, including rice, maize, wheat and barley. Phylogenetic relationships among these cereals have been studied based on conventional systematic data and on restriction fragment sizes of chloroplast DNA (Tateoka 1957; Vedel et al. 1980; Enomoto et al. 1985; Katayama and Ogihara 1996; Soreng and Davis 1998). Nuclear chromosomal synteny and collinearity were also analyzed using homoeologous RFLP probes (Ahn and Tanksley 1993; Van Deynze et al. 1995a, b; Devos et al. 1995; Wilson et al. 1999).

In this study, rice chloroplast microsatellites were used to examine the degree of variation or conservation at homologous loci among a diverse set of accessions from the genus *Oryza*, and eight other Gramineae species.

Materials and methods

Plant materials

Thirteen *O. sativa* cultivars were selected for the evaluation of allelic diversity of chloroplast microsatellite loci (see Table1). These cultivars were simultaneously used for nuclear microsatellite analysis (Cho et al. 2000) and are the mapping parents used in rice genome analysis in China, Korea, Philippines, Japan and the United States. For phylogenetic analysis, 19 wild rice accessions and accessions of eight other Gramineae species were used (see Table 3). Total DNA of barley, oat, millet, sorghum and sugarcane was kindly provided by William Wilson, and teosinte DNA was provided by Shannon Paintner, Cornell University, respectively.

DNA extraction

Total DNA was extracted from fresh leaves of a single plant by the potassium acetate method (Dellaporta et al. 1983). The quality and concentration of the extracted DNA were estimated using mini-gel electrophoresis (Sambrook et al. 1989).

Chloroplast microsatellites

Twelve chloroplast microsatellites having more than ten repeats were previously reported by Provan et al. (1996). However, two pairs of them were located in inverted repeat regions of the chloroplast genome, indicating that these two are duplicated within the rice chloroplast genome. In this study, ten unique pairs of primers were designed to produce specific PCR products in the range of 100–150 bp using the PRIMER program (Version 0.5, Eric Lander, Whitehead Institute, Cambridge, Mass). Because the chloroplast genome consists of AT-rich sequences which have a low melting point, long primers had to be designed, ranging from 20 to 28 nucleotides, to optimize PCR reactions (see Table 2). These primer sequences were used to amplify all the accessions reported in this study.

PCR-amplification and silver staining

PCR was performed in 50-µl reactions containing 100 ng of template DNA, 0.2 µM of each primer, 200 µM of each dNTP (dATP, dCTP, dGTP and dTTP), 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin and 1 unit of *Taq* DNA polymerase. Amplification was carried out in a PTC100 96U thermocycler (MJ Research, USA) as follows: 5min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 55°C, 2 min at 72°C, and 5 min at 72°C for final extension. Amplified products were electrophoresed in 4.0% polyacrylamide gels and the banding patterns were visualized using silver staining as described by Panaud et al. (1996).

Allele scoring

After silver staining of polyacrylamide gels, a cluster of 2–5 discrete bands was apparent for most of the markers. The size (in nucleotides) of the most-intensely amplified band for each microsatellite marker was determined for each genotype based on migration relative to a Nipponbare allele standard and molecular-weight size markers V and VIII (Boehringer Mannheim, Germany). Nipponbare was used as a reference for allele molecular-weight determination because the expected allele sizes of PCR products were available for this variety based on sequences from the GenBank database.

Data analysis

The allelic diversity of chloroplast microsatellites among the 13 cultivars was calculated according to the polymorphism information content (PIC) value described by Botstein et al. (1980) and modified by Anderson et al. (1993) for self-pollinated species as follows:

$$\mathrm{PIC}_i = 1 - \sum_{j=1}^{n} p_{ij}^2,$$

where p_{ij} is the frequency of the *j*th pattern for marker *i* and summation extends over *n* patterns.

Phylogenetic relationships among wild rice species, and among the Gramineae species were studied based on the similarity of chloroplast microsatellite allele sizes. The ratio of common amplified fragments was used as a similarity index, and calculated according to the following formula:

$$F_{ij} = 2B_{ij}/A_{ij}$$

where A_{ij} and B_{ij} are the numbers of total and common fragments observed between *i*th and *j*th varieties (Nei and Li 1979). Based

Table 1 Allelic diversity of ten chloroplast microsatellites found among 13 *O. sativa* cultivars. Allele size was designated as nucleotide length difference (+/– for longer/shorter) from Nipponbare allele size

Species	Type	Variety name	Mapping parent[a]	Locus and size in bp[b]									
				RCt1 103	RCt2 122	RCt3 129	RCt4 128	RCt5 143	RCt6 111	RCt7 126	RCt8 131	RCt9 143	RCt10 129
O. sativa	Japonica	Nipponbare	J	0	0	0	0	0	0	0	0	0	0
O. sativa	Japonica	Gihobyeo	K	0	0	0	0	0	0	0	0	0	0
O. sativa	Japonica	Jing-Xi 17	Ch	0	0	–1	0	0	0	0	0	1	0
O. sativa	Tropical Japonica	Azucena	P	0	0	–1	0	0	–1	0	–1	1	0
O. sativa	Tropical Japonica	Lemont	T	0	0	–1	0	0	0	0	–1	1	0
O. sativa	Japonica/Indica[c]	Milyang 23	K	0	0	0	0	0	0	0	0	0	0
O. sativa	Indica	IR36	Co	0	0	1	0	–1	–1	0	–1	2	0
O. sativa	Indica	IR64	P	0	0	1	0	–1	–1	0	–1	2	0
O. sativa	Indica	N22	Co	0	0	1	0	0	–1	0	–1	2	0
O. sativa	Indica	Zhai-Ye-Qing 8	Ch	0	0	–1	0	0	0	0	–1	1	0
O. sativa	Indica	Teqing	T	0	0	1	0	–1	0	0	0	2	0
O. sativa	Indica	Kasalath	J	0	0	2	0	0	–1	0	–1	2	0
O. sativa	Indica	BS125	Co	0	0	0	0	0	0	0	0	2	0
	No. of alleles			1	1	4	1	2	2	1	2	3	1
	PIC			0.000	0.000	0.710	0.000	0.355	0.473	0.000	0.497	0.639	0.000

[a] Parental variety of mapping population developed in Cornell University, USA (Co), Texas A & M University, USA (T), Rice Genome Program, Japan (J), National Institute of Science and Agricultural Technology, Korea (K), Academia Sinica, China (Ch), and International Rice Research Institute, The Philippines (P)

[b] Expected size in Nipponbare

[c] Derived from a Japonica/Indica cross

on the ratio of common fragments, a dendrogram showing phylogenetic relationships among varieties was constructed by the UPGMA method (Sneath and Sokal 1973).

In this paper, "homology" for the amplified regions and primer regions refers to the number of identical aligned bp divided by the number of total aligned bp, with a score of 1 for each gap; "homology excluding gap" refers to the number of identical aligned bp divided by the number of total bp taking no penalty for a gap.

Results

Allelic diversity of chloroplast microsatellites among rice cultivars

Amplified products were obtained for all ten chloroplast microsatellites with the 13 *O. sativa* cultivars (Table 1). To characterize the allelic diversity and the informativeness of the chloroplast microsatellites, the number of alleles and the polymorphism information content (PIC) of each were examined. The number of alleles per microsatellite ranged from 1 to 4. Those with only one allele, i.e., monomorphic loci, were observed for five markers (RCt1, 2, 4, 7 and 10), and the maximum of four alleles was detected only for RCt3. PIC values ranged from 0 to 0.710 with an average of 0.267. When PIC values of the monomorphic microsatellites (PIC=0) were excluded from the calculation, the average PIC of the five polymorphic microsatellites was 0.54.

As shown in Table 2, three, two and five microsatellites were located in coding, intron and non-coding regions, respectively. All three microsatellites in coding regions (RCt2, 4 and 7) were monomorphic among rice cultivars, as was one each in an intron (RCt1) and in a non-coding region (RCt10). Using the other five polymorphic microsatellite markers, the chloroplast genomes of the 13 cultivars could be classified into nine types. Of these, three types represented at least two cultivars; identical banding patterns were observed between IR36 and IR64, between Zhai-Ye-Qing 8 and Lemont, and between Nipponbare, Gihobyeo and Milyang 23. The chloroplast genomes of the other six cultivars were unique and could be individually identified. Kasalath was the most distinct, in that it contained a unique allele at RCt3 that was not detected in any of the other *Oryza* or other Gramineae accessions evaluated in this study.

Allelic diversity of chloroplast microsatellites among wild rice species and Gramineae species

Chloroplast microsatellites of 19 accessions of wild rice and eight accessions of other Gramineae species were examined using all ten pairs of primers (Table 3, Fig. 1). Null alleles were observed at RCt8 in *O. ridleyi* and in the eight Gramineae species. At the other nine loci, DNA fragments containing microsatellite regions were amplified with the primers derived from *O. sativa* cv Nipponbare chloroplast sequences. Monomorphic patterns across all germplasm examined were observed at two loci, RCt2 and RCt7, which were located in the coding regions of chloroplast genes.

In order to clarify the phylogenetic relationships among species and genera, the banding patterns of common chloroplast microsatellites were compared among rice species and Gramineae species. In this study, the genomes of the 19 accessions of wild rice were classified as CC, CCDD, BBCC, EE, FF and an unidentified tetraploid. In addition, three *O. sativa* cultivars having the

Table 2 Rice chloroplast microsatellite information

Locus	Other name[a]	Location[b] (gene)	Repeat	Primer sequence (5′–3′)	Size (bp) in Nipponbare
RCt1	Oscp3536	Intron (*trn*K)	$(A)_{10}$	CATCCTTTTCAATCCAAAATCA TGCCTGATGTAGGGAAAAGC	103
RCt2	Oscp29138	Coding region (*rpo*C2)	$(A)_{11}$	CTGGGGGGGATTATACCTGT ATATCTCTCATTTCCGACGCA	122
RCt3	Oscp43899	Intergenic region	$(A)_{10}$	TAGGCATAATTCCCAACCCA CTTATCCATTTGGAGCATAGGG	129
RCt4	Oscp49274	Coding region (*psb*G)	$(T)_{12}$	ACGGAATTGGAACTTCTTTGG AAAAGGAGCCTTGGAATGGT	128
RCt5	Oscp75969	Intergenic region	$(T)_{10}$	ATTTGGAATTTGGACATTTTCG ACTGATTCGTAGGCGTGGAC	143
RCt6	Oscp76221	Intergenic region	$(A)_{10}$	GAATTTTAGAACTTTGAATTTTTTACCC AAGCGTACCGAAGACTCGAA	111
RCt7	Oscp76527	Coding region (*inf*A)	$(T)_{10}$	GTGTCATTCTCTAGGCGAAC AAATATGACAGAAAAGAAAAATAGG	126
RCt8	Oscp78412	Intron (*rpl*16)	$(T)_{17}$	ATAGTCAAGAAAGAGGATCTAGAAT ACCGCGATTCAATAAGAGTA	131
RCt9	Oscp80599	Intergenic region	$(T)_{10}$	ATAAGGTTATTCCCCGCTTACC AAATTGGGGGAATTCGTACC	144
RCt10	Oscp89568	Intergenic region	$(T)_{10}$	TCTTCATTTGGAATCTGGGC CTATTGATGCAAACGCTGTACC	129
RCt11 (=RCt10)	Oscp124451	Intergenic region	$(A)_{10}$	– –	–
RCt12 (=RCt9)	Oscp134510	Intergenic region	$(A)_{10}$	– –	–

[a] Locus name after Provan et al. (1996)
[b] Information on microsatellite location is from Hiratsuka et al. (1989) and the GenBank database (CHOSXX)

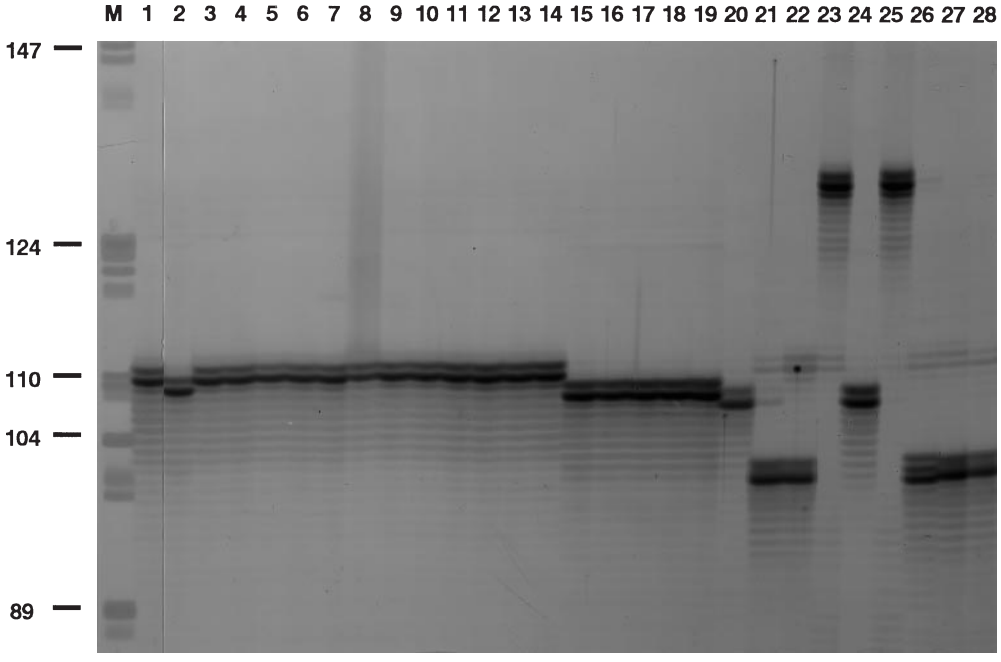

Fig. 1 Silver-stained polyacrylamide-gel electrophoresis patterns showing simple sequence length polymorphism at the RCt6 locus among rice species and Gramineae species. *M*=molecular-size marker; *1* IR36; *2–4 O. officinalis*; *5–9 O. latifolia*; *10–14 O. minuta*; *15–18 O. australiensis*; *19 O. brachyantha*; *20 O. ridleyi*; *21* maize; *22* teosinte; *23* wheat; *24* oat; *25* barley; *26* pearl millet; *27* sorghum; *28* sugarcane

AA genome, i.e., Nipponbare (Japonica), Lemont (tropical Japonica) and IR36 (Indica), were included in the comparison. The proportion of common bands among all these germplasm samples was used as a similarity index, and a dendrogram was constructed (Fig. 2). As exemplified by the pattern at RCt6 in Fig. 1, alleles of similar size ranges tended to be observed among members of the same clade in this survey of grasses; a 100-bp allele is observable in the Panicoideae, maize, teosinte, sorghum, sugarcane and pearl millet; a 107-, 108-, 109- or 110-bp allele is present in *Oryza* species; and a 128-bp allele, suggestive of an insertion event, is observable in two of the Pooideae, wheat and barley. Oat represented an exception at this locus, because it shared a 107-bp allele

Table 3 Allelic diversity of ten chloroplast microsatellites found among19 accessions of wild rice species and eight accessions of Gramineae species[a]

Species	Genome	Acc. no.	Origin	Locus and size in bp									
				RCt1 103	RCt2 122	RCt3 129	RCt4 128	RCt5 143	RCt6 111	RCt7 126	RCt8 131	RCt9 144	RCt10 129
O. officinalis	CC	100896	Thailand	–1	0	–10	0	–1	–2	0	0	3	–1
O. officinalis	CC	105315	India	0	0	–1	0	–2	–1	0	–3	0	–1
O. officinalis	CC	105365	Thailand	–1	0	–10	0	–1	–1	0	–2	4	–1
O. latifolia	CCDD	100914	Mexico	–1	0	–14	0	–1	–1	0	–2	2	–1
O. latifolia	CCDD	100963	Guatemala	–1	0	–12	0	–1	–1	0	–2	3	–1
O. latifolia	CCDD	103787	Colombia	–1	0	–12	0	–1	–1	0	–2	1	–1
O. latifolia	CCDD	104985	Cuba	–1	0	–12	0	–1	–1	0	–1	1	–1
O. latifolia	CCDD	105142	Costa Rica	–1	0	–12	0	–1	–1	0	–2	1	–1
O. minuta	BBCC	101082	Philippines	0	0	0	0	–2	–1	0	–2	0	–1
O. minuta	BBCC	101089	Philippines	0	0	0	0	–2	–1	0	–1	0	–1
O. minuta	BBCC	101094	Philippines	0	0	0	0	–2	–1	0	–2	0	–1
O. minuta	BBCC	101125	Philippines	0	0	0	0	–2	–1	0	–2	0	–1
O. minuta	BBCC	101141	Philippines	0	0	0	0	–2	–1	0	–2	0	–1
O. australiensis	EE	101397	Australia	–1	0	1	0	–2	–3	0	–5	4	–1
O. australiensis	EE	101410	Australia	–1	0	1	0	–2	–3	0	–5	4	–1
O. australiensis	EE	103318	Australia	–1	0	1	0	–2	–3	0	–5	4	–1
O. australiensis	EE	105267	Australia	–1	0	1	0	–2	–3	0	–5	4	–1
O. brachyantha	FF	101232	Sierra Leone	–2	0	–1	0	–3	–3	0	–7	4	–2
O. ridleyi	–[b]	100877	Thailand	–2	0	–3	0	–3	–4	0	N	–1	–2

Species	Common name	RCt1	RCt2	RCt3	RCt4	RCt5	RCt6	RCt7	RCt8	RCt9	RCt10
Zea mays	Maize	0	0	–15	0	(13)	–11	0	N	(24)	–2
Zea diploperennis	Teosinte	0	0	–15	–3	(13)	–11	0	N	(24)	–2
Triticum aestivum	Wheat	0	0	1	–3	–1	(17)	0	N	(12)	–2
Avena sativa	Oat	0	0	–10	–3	–5	–4	0	N	(12)	–2
Hordeum vulgare	Barley	0	0	0	–3	–2	(17)	0	N	(12)	–2
Pennisetum glaucum	Pearl millet	0	0	–15	3	–3	–11	0	N	(12)	–2
Sorghum bicolor	Sorghum	0	0	0	9	0	–11	0	N	0	–2
Saccharum officinarum	Sugarcane	0	0	–15	9	(15)	–11	0	N	2	–2

[a] Allele size was designated as nucleotide length difference (+/- for longer/shorter) from Nipponbare allele size. Number in parenthesis indicates approximate nucleotide length. N: null allele (no amplification)

[b] Unidentified tetraploid

with *O. ridleyi*, differentiating it from the other Pooideae surveyed here.

Microsynteny between rice and maize in microsatellite-containing regions of chloroplasts

The entire sequence of the maize chloroplast genome is available in the GenBank database (CHZMXX) (Maier et al. 1995). This provided an opportunity to compare the regions of the rice and maize chloroplast genomes where the rice microsatellites had been detected. Table 4 shows the nucleotide sequences of microsatellite regions in maize, and from this information we were able to calculate the expected maize allele sizes (in nucleotides) corresponding to the ten homologous microsatellites in rice. At eight out of ten loci, the predicted sizes of maize alleles coincided precisely with those estimated by polyacrylamide-gel electrophoresis. At RCt10, a 1-bp difference was observed between expected and detected allele sizes. This might be explained by the intraspecific variation in *Zea mays*, as different accessions were used for the sequence analysis by Maier et al. (1995) and the microsatellite analysis presented in this study. At the RCt8 locus, allele sizes could not be compared because maize showed a null allele. Though a high level of sequence similarity (93.3%) was observed in primer sequences at this locus, one of the primers had a mismatch at the third nucleotide from the 3′-end, with the last two nucleotides containing thymine and adenine bases which produced weaker hydrogen bonds between DNA strands than the guanine and cytosine bases. This might explain why no PCR-amplification occurred at this locus.

According to the nucleotide sequences of rice and maize, small insertions/deletions were observed in all seven loci that were located in non-coding regions or introns. In order to examine the proportion of base substitutions occurring in the amplified microsatellite-containing regions of these two grass genera, sequence similarity was calculated excluding the regions differentiated by insertion/deletion events. Sequence similarity ranged from 87.8 to 98.4%, with an average of 92.6% (Table 4). Especially high similarity values were observed at the three loci in coding regions (RCt2, 4 and 7) and at the

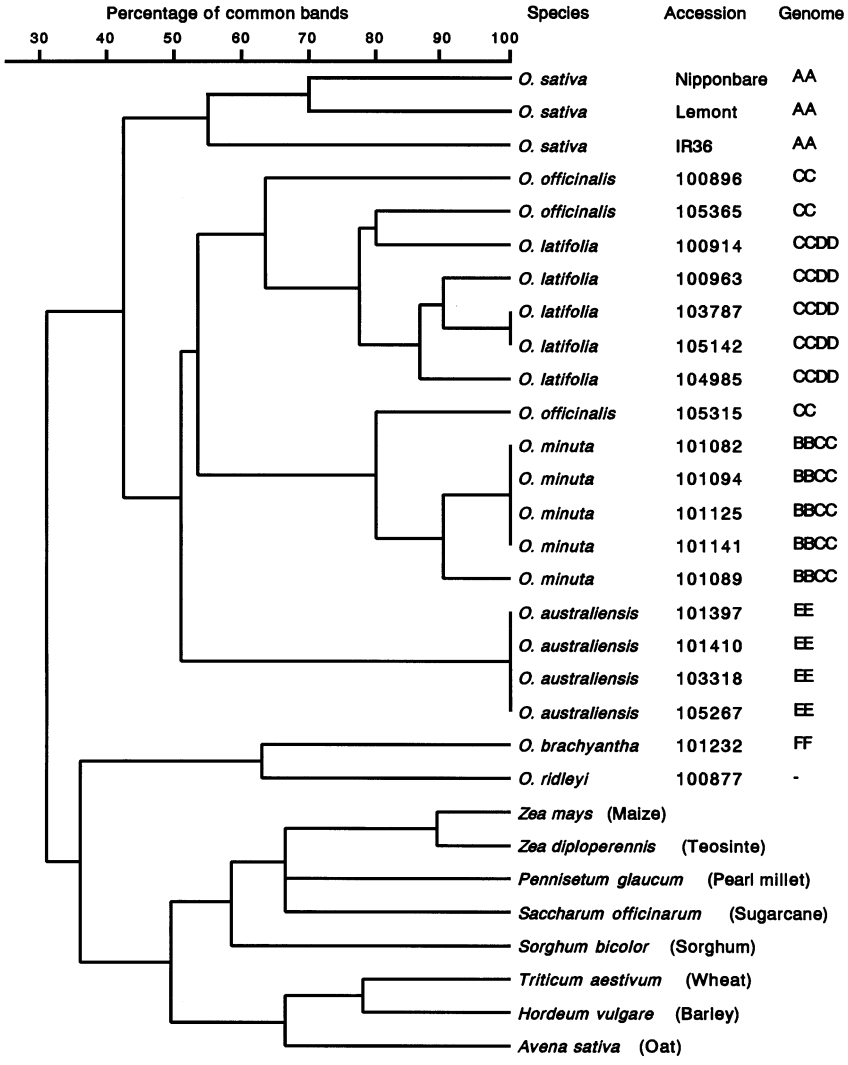

Fig. 2 A dendrogram showing genetic relationships among rice species and Gramineae species based on the chloroplast microsatellite analysis

two loci occurring in inverted repeat regions of the chloroplast genome (RCt9 and 10). When the microsatellite primer-annealing regions (forward and reverse primer regions; 41–48 bp) were compared specifically between rice and maize, the homology was similar to that of the amplified regions calculated without the insertions/deletions.

Discussion

Using the same 13 cultivars examined here, Cho et al. (2000) studied 300 rice nuclear microsatellites, 171 developed from a library of random genomic clones and 129 from DNA sequences extracted from GenBank. They revealed that the average number of alleles per locus was 5.13 for microsatellites from the genomic library and 2.78 for those from GenBank, with an average PIC value of 0.68 and 0.39, respectively. These values were higher than those of chloroplast microsatellites; 1.80 and 0.267 for the average allele number and the average PIC value, respectively. These figures confirm that chloroplast SSLPs detect less diversity than nuclear microsatellites. Provan et al. (1996) reported that chloroplast microsatellites were more informative than nuclear microsatellites. In their calculations of diversity, they pooled information from six polymorphic chloroplast microsatellites using 20 A-genome accessions (14 cultivars and six wild accessions). However, the average diversity value for an individual chloroplast locus was lower than the average of eight nuclear microsatellites reported by Wu and Tanksley (1993) using the same set of plant materials.

A low PIC value for an individual chloroplast microsatellite is the consequence of the conservative nature of the chloroplast genome which shows uniparental (maternal) inheritance, a lower mutation rate, and contains a higher proportion of coding regions than the nuclear genome (Wolfe et al. 1987; Hiratsuka et al. 1989). The rice chloroplast genome consists of a circular DNA molecule. The size of the molecule is about 135 kbp (Hirai et al. 1985) and there are 20–200 molecules in a single chloroplast (Palmer 1987). Since the chloroplast genome has a prokaryotic structure, polymorphisms occurring in mi-

Table 4 Homology of nucleotide sequences between rice and maize[a] in chloroplast microsatellite regions

Locus		Size (bp)	Microsatellite constitution	Homology (%)	No. of gaps	Total size of gap (bp)	Homology excluding gap	Homology of primer sequences
RCt1	Rice	103	$(A)_{10}G$		3	3		
	Maize	103	$(A)_{11}$	83.0	2	3	88.0	84.1
RCt2	Rice	122	$(A)_{11}$		0	0		
	Maize	122	$(A)_{10}T$	95.1	0	0	95.1	97.6
RCt3	Rice	129	$(A)_{10}$		0	0		
	Maize	114	$(A)_5C(A)_4$	83.7	3	15	94.7	95.2
RCt4	Rice	128	$(T)_{12}GT$		0	0		
	Maize	128	$(T)_{14}$	93.8	0	0	93.8	97.6
RCt5	Rice	143	$(T)_{10}$		1	13		
	Maize	156	$(T)_5AC(T)_3$	81.4	0	0	88.8	88.1
RCt6	Rice	111	$TAG(A)_{10}$		0	0		
	Maize	100	CA[b]	82.0	1	11	91.0	87.5
RCt7	Rice	126	$(T)_{10}$		0	0		
	Maize	126	$(T)_{10}$	94.4	0	0	94.4	93.3
RCt8	Rice	131	$(T)_{17}$		0	0		
	Maize	107	$(T)_2(C)_2(T)_6(G)_2$[c]	71.8	4	24	87.8	93.3
RCt9	Rice	144	$(T)_{10}$		3	24		
	Maize	168	$(A)_2(T)_4G(T)_3$	80.4	0	0	93.8	93.2
RCt10	Rice	129	$(T)_{10}$		1	1		
	Maize	128	$(T)_7G$[d]	96.2	1	2	98.4	97.6

[a] Complete nucleotide sequences of maize chloroplast genome are from Maier et al. (1995) and GenBank database (CHZMXX)

[b, c] and [d]: 11-, 5- and 2-bp gaps were observed in maize homologous regions, respectively

crosatellite regions are likely to be the result of mis-replication or intermolecular recombination in repeated sequences. In this study, ten chloroplast microsatellite loci were surveyed on 13 rice cultivars and no heterogeneous genome constitution was found.

As shown in Table 2, three, two and five microsatellites are located in coding, intron and non-coding regions, respectively. All three microsatellites in coding regions (RCt2, 4 and 7), one in an intron (RCt1) and one in a non-coding inverted repeat sequence (RCt10), were monomorphic in the 13 rice cultivars. The mechanism underlying the observed sequence conservation is unknown, but is probably due to the fact that variation in the length of a simple sequence repeat motif would cause a change in the protein product if it occurred in the open reading frame. Similarly, such changes might affect the splicing activity if they occurred in an intron, and would likely disrupt the symmetry and function of the inverted repeat sequences, and therefore would tend to be suppressed in the chloroplast genome. Further, the three microsatellites in coding regions (RCt2, 4 and 7) showed monomorphism among 19 accessions of wild rice species, indicating a strong selective pressure against length variation at these loci in the genus *Oryza*.

Using the five polymorphic microsatellite markers, the chloroplast genomes of 13 cultivars were classified into nine different types. According to Ishii and Tsunewaki (1991) who made RFLP analysis on chloroplast DNA using six restriction enzymes, only five chloroplast genome types were observed among 68 *O. sativa* local varieties from 15 Asian countries. Compared to RFLP markers, microsatellites provide a higher resolution for identifying and classifying chloroplast genome types. This result suggests that simple sequence repeat length variation occurs more frequently than base substitutions or insertion/deletion events in the chloroplast genome.

Among the 19 accessions of wild rice and eight other Gramineae species surveyed in this study, null alleles were observed only at RCt8. All eight Gramineae accessions and one wild rice accession *(O. ridleyi)* carried the null allele at RCt8. Previously, 24 nuclear microsatellites were examined in the same set of plant materials using the primer pairs designed from the nuclear sequences of rice cultivar IR36 by Chen et al. (1997). Amplified products from all wild rice species were obtained with only six pairs of primers (25%) and no amplified products were observed for any of the Gramineae species with these nuclear-derived primers (data not shown). This suggests that many wild rice species (specifically those that do not share the AA genome of *O. sativa*) and most other Gramineae species have diverged sufficiently from cultivated rice species to have accumulated mutations in the primer regions flanking nuclear microsatellites, thus inhibiting the potential for PCR-amplification and subsequent comparative genome analysis. This is supported by the recent sequence analysis by Chen (1999). For this reason, it is difficult to use nuclear microsatellite markers to study the evolution of homologous loci in species with different genomes.

In contrast to nuclear microsatellite markers, chloroplast microsatellite primers did amplify alleles in wild rice species having different genomes and in the even

more-distantly related Gramineae species. As a result, a dendrogram could be constructed, as shown in Fig. 2. It is of interest that the magnitude of differentiation among the three *O. sativa* cultivars (based on the percentage of common bands) is almost the same as that between the four wild rice species, *O. officinalis*, *O. latifolia*, *O. minuta* and *O. australiensis*, and the other eight Gramineae species. The explanation for this apparent divergence from evolutionary reality stems from the fact that the amplified products generated with a pair of primers contain the microsatellite motif and the regions bordering the SSR, so that the length polymorphisms detected here were not always due to variation only in the number of simple sequence repeat units. In fact, several alleles were smaller than the original Nipponbare allele by more than the ten bases that comprised the entire microsatellite motif, indicating that small insertion/deletion events had occurred very near the SSR.

In the genus *Oryza*, prominent size differences within the genus were not detected among alleles at any of the loci except in the case of the CC- and CCDD-genome species at the RCt3 locus. Therefore, allele size differences in rice were inferred to be caused mainly by the microsatellite length polymorphisms. According to the dendrogram, the three wild species, *O. latifolia* (CCDD), *O. minuta* (BBCC) and *O. australiensis* (EE), showed a low level of within-species variation, and were differentiated equally from each other. *O. officinalis* (CC) showed the largest amount of within-species variation of any of the wild species evaluated. Two accessions had allele patterns similar to *O. latifolia*, while the rest of the accessions clustered with *O. minuta*. Although relatively few accessions have been analyzed here, these data suggest that the cytoplasms of *O. latifolia* and *O. minuta* might have been independently derived from *O. officinalis* (CC). The cluster consisting of *Oryza brachyantha* (FF) and *O. ridleyi* (tetraploid – genome unknown) was located far from the main cluster of other *Oryza* species, indicating that these two species had diverged widely from the other species in the genus *Oryza*. Provan et al. (1997) also analyzed relationships among rice genomes using two-dimensional plots based on the size differences at two of the most informative loci. In this study, these loci were designated RCt9 and RCt12, but as they were both located at the end of inverted repeat sequences in the chloroplast genome, their microsatellite regions and most of the flanking sequences were identical. Therefore, the sequence information derived from RCt9 and RCt12 was redundant and did not cover the whole chloroplast genome.

In the Gramineae species, allele sizes differed from the Nipponbare allele by a greater number of base pairs than those found in *Oryza* species. Although the amplified products in the Gramineae were obtained with primers derived from rice cv Nipponbare sequences, their length polymorphisms were likely to be caused by small insertion/deletion events in addition to, or possibly instead of, microsatellite expansion and contraction. Among the eight non-rice Gramineae species examined here, complete chloroplast sequence information was available only for maize (Maier et al. 1995). The rice chloroplast microsatellite regions and their homologous regions in maize were compared in Table 4. Small insertions/deletions, often involving a few base pairs at each of several locations, were found in seven out of ten loci, and base substitution events were also observed within the mononucleotide repeat regions as well as in the flanking regions. The three loci located in coding regions (RCt2, 4 and 7) were highly conserved, containing no insertions/deletions throughout the amplified regions. Further, no base substitutions were detected in the $(T)_{10}$ repeat at RCt7 and only a single base-pair substitution was observed in the microsatellite motifs at RCt2 and RCt4. In this comparison, these homologous regions in maize still contained SSRs consisting of ten or more repeats, suggesting an ancient origin of these repeats that predates the divergence of the grasses, estimated to have occurred approximately 60 million years ago. Due to the highly conserved nature of these three SSR loci, they did not contribute to the length differences between rice and maize.

If the rice-maize comparison serves as an example, it can be concluded that less than half of the rice chloroplast microsatellites (10 bp or longer) detected in Nipponbare are retained in other Gramineae species, and when they are, they tend to be highly conserved, showing little or no variation. This observation explains why the comparison among distantly related Gramineae species based on chloroplast microsatellite allele sizes compressed the magnitude of differentiation. Length polymorphisms between diverse genera represent an accumulation of different kinds of mutation events, with base substitution and insertion/deletion events representing significantly lower mutation rates than microsatellite expansion/contraction. Thus, the magnitude of the size differences of amplified products does not provide a reliable measure of evolutionary distance because it is impossible to distinguish how many or what kind of events have contributed to the size differences.

In conclusion, the overall stability and high level of microsynteny among chloroplast genomes in diverse members of the grass family makes it possible to amplify orthologous loci using common sets of primers across a broad spectrum of Gramineae genera. However, while the primer pairs used in this study bracketed several mutationally active A/T-rich mononucleotide microsatellite motifs in rice, they amplified degenerate, interrupted motifs in distantly related genera. Thus, the occurrence of length variation due to the expansion/contraction of microsatellite motifs in specific regions of chloroplast genomes appears to be a species- or genome-specific phenomenon and should not be confused with length variation due to insertion/deletion events in homologous regions of distantly related genera. As a result, analysis based on length variation using the chloroplast microsatellite primers presented in this study offers a useful indicator of evolutionary divergence for within-genome comparisons, but sequence-based analysis will be re-

quired to obtain reliable estimates of divergence across larger evolutionary distances.

Acknowledgements We thank Dr. Masahiro Yano, National Institute of Agribiological Resources, Japan, for providing the seed stocks of 'Kasalath'. We also thank William Wilson and Shannon Paintner, Cornell University, USA, for providing us with the total DNA of barley, oat, millet, sorghum and sugarcane, and teosinte DNA, respectively. This work was supported by the Government of Japan based on a visiting scientist fellowship to T. Ishii, and by the Rockefeller Foundation for a grant (93–37300–8703) to S. McCouch.

References

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS (1999) Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor Appl Genet 98:1320–1328

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90:7980–7984

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 93: 1071–1077

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1997) Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor Appl Genet 94:61–67

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB (1995) Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci 35:1439–1445

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Becker J, Heun M (1995) Barley microsatellites: allele variation and mapping. Plant Mol Biol 27:835–845

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Chen X (1999) Frequency, distribution and sequence divergence of rice microsatellites. PhD dissertation, Cornell University

Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor Appl Genet 95:553–567

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, Park WD, Ayres N, Cartinhour S, McCouch SR (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100: 713–722

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

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Devos KM, Moore G, Gale MD (1995) Conservation of marker synteny during evolution. Euphytica 85:367–372

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Mirissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature 380:152–154

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ, Evans CA, DeAngelis MM, Levinson DM, Kruglyak L, Goodman N, Copelang NG, Jenkins NA, Hawkins TL, Stein L, Page DC, Lander ES (1996) A comprehensive genetic map of the mouse genome. Nature 380:149–152

Enomoto S, Ogihara Y, Tsunewaki K (1985) Studies on the origin of crop species by restriction endonuclease analysis of organellar DNA. I. Phylogenetic relationships among ten cereals revealed by the restriction fragment patterns of chloroplast DNA. Jpn J Genet 60:411–424

Hirai A, Ishibashi T, Morikami A, Iwatsuki N, Shinozaki K, Sugiura M (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 kDa photosystem-II reaction center protein. Theor Appl Genet 70:117–122

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (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

Ishii T, Tsunewaki K (1991) Chloroplast genome differentiation in Asian cultivated rice. Genome 34:818–826

Katayama H, Ogihara Y (1996) Phylogenetic affinities of the grasses to other monocots as revealed by molecular analysis of chloroplast DNA. Curr Genet 29:572–581

Maier RM, Neckermann K, Igloi GL, Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614–628

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Olufowote JO, Xu Y, Chen X, Park WD, Beachell HM, Dilday RH, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40:370–378

Palmer JD (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am Nat 130:S6–S29

Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252:597–607

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995a) Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–7763

Powell W, Morgante M, Andre C, McNicol JW, Macharay GC, Doyle JJ, Tingey SV, Rafalski JA (1995b) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. Curr Biol 5:1023–1029

Powell W, Morgante M, Doyle JJ, McNicol JW, Tingey SV, Rafalski JA (1996) Genepool variation in genus *Glycine* subspecies *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. Genetics 144:793–803

Provan J, Corbett G, Waugh R, McNicol JW, Morgante M, Powell W (1996) DNA fingerprints of rice (*Oryza sativa*) obtained from hypervariable chloroplast simple sequence repeats. Proc R Soc Lond B 263:1275–1281

Provan J, Corbett G, McNicol JW, Powell W (1997) Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. Genome 40:104–110

Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884–889

Sneath PHA, Sokal RR (1973) Numerical taxonomy. W. H. Freeman and Company, San Francisco

Soreng RJ, Davis JI (1998) Phylogenetics and character evolution in the grass family (Poaceae): Simultaneous analysis of mor-

phological and chloroplast DNA restriction-site character sets. Bot Rev 64:1–85

Tateoka T (1957) Miscellaneous papers on the phylogeny of Poaceae. X. Proposition of a new phylogenetic system of Poaceae. J Jap Bot 32:275–287

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 100:698–712

Van Deynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells ME (1995a) Comparative mapping in grasses: wheat relationships. Mol Gen Genet 248:744–754

Van Deynze AE, Nelson JC, O'Donoughue LS, Ahn SN, Siripoonwiwat W, Harrington SE, Yglesias ES, Braga DP, McCouch SR, Sorrells ME (1995b) Comparative mapping in grasses: oat relationships. Mol Gen Genet 249:349–356

Vaughan D (1989) The genus *Oryza* L. Current status of taxonomy. IRRI Res Paper Ser No. 138. International Rice Research Institute, Manila, The Philippines

Vedel F, Lebacq P, Quetier F (1980) Cytoplasmic DNA variation and relationships in cereal genomes. Theor Appl Genet 58: 219–224

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

Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize, and the domesticated panicoids. Genetics 153:453–473

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

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235