Genet Resour Crop Evol (2011) 58:1177–1186
DOI 10.1007/s10722-010-9651-8

RESEARCH ARTICLE

Cloning of genome-specific repetitive DNA sequences in wild rice (*O. rufipogon* Griff.), and the development of *Ty3-gypsy* retrotransposon-based SSAP marker for distinguishing rice (*O. sativa* L.) *indica* and *japonica* subspecies

Ning Xiao · Genlou Sun · Yihuan Hong · Ruixiang Xia · Chao Zhang · Yan Su · Jianmin Chen

Received: 20 March 2010 / Accepted: 13 December 2010 / Published online: 20 January 2011
© Springer Science+Business Media B.V. 2011

Abstract In the rice genome, insertions and eliminations of transposable elements have generated numerous transposon insertion polymorphisms (TIPs). Common wild rice (*O. rufipogon* Griff.), the ancestor of Asian cultivated rice (*O. sativa* L.), carries abundant genetic variations. To find subspecies-specific (SS) markers that can distinguish *O. sativa* ssp. *indica* and ssp. *japonica*, some long terminal repeat (LTR) sequences (sc1-14) of AA genome-specific *RIRE* retrotransposon were isolated from *O. rufipogon* genome. Sequences sc1 and sc12 were successfully utilized to develop the SS marker system based on retrotransposon inserted position polymorphisms. Twenty-two SS markers (ssi1-9, ssj1-13) were developed, where ssi1-9 are the *indica*-specific types, and ssj1-13 the *japonica*-specific types. The average accuracy of these markers in distinguishing the two subspecies is over 85%. SS marker ssj-10 can distinguish the two subspecies at 100% accuracy. Principal component analysis (PCA) showed that these markers could successfully distinguish *indica* from *japonica* varieties, regardless of their geographical origin.

Keywords *O. rufipogon* · Rice · Retrotransposon · *RIRE* · Subspecies-specific markers

Abbreviations

SS marker	Subspecies-specific markers
LTR	Retrotransposon long terminal repeat
PCA	Principal component analysis
RBIPs	Retrotransposon-based insertion polymorphisms
ICD	Integrase coding domain
TEs	Transposable elements
RT	Reverse transcriptase
RFLP	Restriction fragment length polymorphism
RAPD	Random amplified polymorphic DNA
SSR	Simple sequence repeat
SNP	Single nucleotide polymorphism

Introduction

Rice (*Oryza sativa*) is one of the most important crops worldwide. Asian cultivated rice can be classified into two subspecies, *indica* and *japonica*, based

N. Xiao · Y. Hong · R. Xia · C. Zhang · Y. Su · J. Chen (✉)
College of Bioscience and Biotechnology, Yangzhou University, 88 Daxue Road, 225009 Yangzhou, China
e-mail: jmchen@yzu.edu.cn

N. Xiao
Lixiahe Region Agricultural Research Institute of Jiangsu, 225000 Yangzhou, China

G. Sun
Department of Biology, Saint Mary's University, Halifax, NS, Canada

on their morphology and growth habitats (Oka and Chang 1959, 1962). Recently, the accurate identification of *indica* and *japonica* at the molecular level has been a hot topic in rice research, because it would not only provide essential information for selecting appropriate parents in inter-subspecies hybrid rice research (Zhang et al. 1994; Liu et al. 2002), but it would also increase our understanding of the evolution of cultivated rice. Some DNA polymorphic markers have been utilized for this purpose, such as RFLPs (Zhang et al. 1992; Qian et al. 1995), RAPDs (Yu and Nguyen 1994), SSRs (Monna et al. 2006), and SNPs (McNally et al. 2009).

In higher plants, repetitive DNA sequences can be divided into two major groups based on their genomic organization and localization on the chromosomes. The first group includes satellite DNAs, telomeric repeats, and rDNA, which are preferentially located at pericentromeric, subtelomeric, telomeric, or intercalary regions of the chromosomes, in tandem repeat units. The second group of repetitive DNA sequences comprises dispersed elements, including transposable elements (TEs), which are distributed along the chromosomes of the genome and are further divided into two classes (Flavell et al. 1994). Class I includes long terminal repeat (LTR) retrotransposons. They represent 20% genome in the rice (i.e. 80 Mb) and are distributed genome-wide with a high copy number (Mao et al. 2009). It is believed that they integrated into host chromosomes through a "copy-and-paste" mechanism of RNA intermediates (Vitte et al. 2004). The class II transposable elements transpose via DNA intermediates, usually resulting in relatively low copy numbers (usually <100 copies per genome) (Kunze et al. 1997). The LTR retrotransposons inserted into the rice genome less than 15 million years ago and completed the transposition events within the last three million years (SanMiguel et al. 1996; Ma et al. 2004). However, unequal homologous recombination and illegitimate recombination have caused fragments of LTR retrotransposons to be removed and rearranged (Ma et al. 2004), which produce retrotransposon-based insertion polymorphisms (RBIPs). As individual RBIP insertions can be detected by PCR using a flanking primer and a retrotransposon-specific primer (Flavell et al. 1998), they are very useful in population genetics and phylogenetic analyses in plants (Vitte et al. 2004; Jing et al. 2005), especially in rice, barley, and wheat (Waugh et al. 1997; Queen et al. 2004; Schulman and Kalendar 2005). High copy number retrotransposons will provide more polymorphic information than the low copy number retrotransposons (Flavell et al. 1998); therefore retrotransposon families of high copy number should be selected. *Ty1-copia* retrotransposons exist in high copy numbers in plant genomes and have been utilized as molecular markers for genetic research in barley (Waugh et al. 1997), *Medicago sativa* L. (Porceddu et al. 2002), cashew (Syed et al. 2005). However, *Ty3-gypsy* retrotransposons have seldom been used to analyze retrotransposon specific-inserted polymorphisms.

The suggestion that cultivated rice shares the same ancestor as common wild rice (Oka and Chang 1959) has been supported by a set of physical, geographical, and molecular data. Recently, Vitte et al. (2004) compared the transposition history of 110 LTR retrotransposons in the genomes of two rice varieties, 93-11 (*indica*) and Nipponbare (*japonica*), and estimated that these two genomes diverged from one another at least 200,000 years ago. That *indica* and *japonica* subspecies were independently domesticated from *O. rufipogon* was supported by the data of Rakshit et al. (2007). Thus, RBIPs might represent a record of *O. rufipogon* evolving into sub-species. Subspecies specific markers (SS) could be developed based on RBIPs. In this study, a retrotransposon inserted site amplification polymorphism system based on *RIRE* LTR fragments cloned from *O. rufipogon* (named Dongxiang wild rice in China) was used to develop SS markers.

Materials and methods

Construction of Dongxiang wild rice BAC library

Dongxiang wild rice BAC library was constructed followed by Vinatzer et al. (1998) described methods. Dongxiang wild rice DNA was partially digested with restriction enzyme *Hin*d III, subjected to double size selection, electroeluted and cloned into the pBeloBAC11 BAC vector. This library was constructed with approximately 1× coverage that is represented by 5760 BAC clones with average insert size of nearly 80 kb.

Separated AA genome specific repeats

One microliter of BAC DNA from Dongxiang wild rice (*O. rufipogon,* genome AA) was spotted on two

Hybond-N + nylon membranes (Amersham Pharmacia Biotech), and fixed at 80°C (Hong et al. 2006). These two membranes were hybridized with *O. rufipogon* and *O. officinalis* genomic DNA as probes, respectively. Four BAC clones that showed strong hybridization signals on the membrane hybridized with *O. rufipogon*, but weak in the membrane hybridized with *O. officinalis*, were used as probes. Probe preparation, membrane hybridization, and signal detection were performed following the instructions of the DIG-High Prime DNA Labeling and Detection Kit (Roche, No. 11363514910). The positive BAC clones were selected, sub-cloned, and sequenced. Blast analysis against the repeats database of NCBI and TIGR (www.ncbi.nlm.nih.gov/BLAST, www.tigr.org/tdb/e2k1/plant.repeats/) was performed.

Plant materials and DNA extraction

Eleven wild rice species representing different genomes, and 61 cultivated rice varieties (including *indica* and *japonica,* the names of cultivars were omitted), were used. The representative germplasms in Table 1 were randomly selected from rice varieties in a rice breeding program. Plant DNA was extracted from young leaves according to the method described by Edwards et al. (1991) (Table 1).

Southern hybridization

Genomic DNA samples were completely digested with *Bam*H I and *Hin*d III (Takara), and then separated by electrophoresis at 60 V overnight in a 1% agarose gel with 0.5× TBE buffer, and trans-blotted to Hybond-N+ nylon membranes. Hybridization probes from the clones were prepared using nick translation. Hybridization, washing, and signal detection were performed following the instructions of the DIG-High Prime DNA Labeling and Detection kit.

Developing molecular markers from the cloned repetitive sequences

The procedures of transposon display (Kwon et al. 2005) and sequence-specific amplification polymorphisms (Waugh et al. 1997) were modified in our study. Based on the high copy number and conserved *RIRE* family LTR, two anchored primers were designed to ensure both sides of the *RIRE* inserted position could be specifically amplified. The anchored primers, r2-1 and r2-2, were designed based on subclone sc12 to detect *RIRE*-2 sub-family retrotransposon-based insertion polymorphism (RBIP), and the RBIP of *RIRE*-3 was detected by anchored primers r3-1 r3-2, which were designed based on subclone sc1. Shcherban et al. (2000) showed that the integrase coding domain (ICD) sequences of the *RIRE-2* sub-family were highly homologous to those in *O. officinalis* species. We hypothesized that the coding regions of the reverse transcriptase (RT) of the *RIRE* family would also be highly homologous. Therefore, we designed rt-1 and rt-2 anchored primers to detect inserted polymorphisms close to the reverse transcriptase position. All adaptors sequences and pre-amplification primers were as Lee et al. (2005). The

Table 1 Rice materials used in this study

Species	Accessions	Genome	Origin
O. sativa L. ssp. *indica* Kato	31	AA	China (24), Thailand (2), Philippines (1), India (1) America (3)
O. sativa L. ssp. *japonica* Kato	30	AA	China (24), Italy (1), Philippines (1), America (1), Japan (3)
O. *rufipogon* Griff.	1	AA	China
O. spontanea Roshev.	1	AA	China
O. nivara Sharma et Shastry	1	AA	Asia subtropics
O. glaberrima Steud.	1	AA	Africa
O. barthii Cheval.	1	AA	Africa
O. longistaminata A. Chev. et Roehr.	1	AA	Asia tropics
O. perennis Moench	1	AA	Africa
Oryza officinalis Wall. ex Watt.	1	CC	Australia
O. minuta Presl	1	BBCC	Asia
O. alta Swallen	1	CCDD	Latin America
O. australiensis Domin	1	EE	Australia

Table 2 Primers used in this study

Type	Name	Sequence (5′–3′)
Anchored primer	r2-1	AGTCTCAGGGTGTTTCCTTG
	r2-2	TCTACACGGCGATGGTATTC
	r3-1	GGG AGCGTTTAGAGCGGT
	r3-2	GCTGTTTCCCATTTGTCT
	rt-1	GATTGAAGCAAGCATCTCGG
	rt-2	ACTCCGAGATGCTTGCTTCA
Adaptor sequences	MPA-1	GACGATGAGTCCTGAG
	MPA-2	TACTCAGGACTCAT
Pre-amplification primer	MP-0	GATGAGTCCTGAGTAA
Selected-amplification primer	MP-CA	GATGAGTCCTGAGTAACA
	MP-GA	GATGAGTCCTGAGTAAGA
	MP-GAA	GATGAGTCCTGAGTAAGAA
	MP-GAC	GATGAGTCCTGAGTAAGAC
	MP-GAG	GATGAGTCCTGAGTAAGAG
	MP-GAT	GATGAGTCCTGAGTAAGAT

primer and sequence information are shown in Table 2. The experimental procedures were similar to the AFLP method (Vos et al. 1995), including single enzyme digestion, adapter ligation, and PCR selective amplification.

Genomic DNA (500 ng) was fully digested with *Mse* I, and the adaptor was linked with the digested DNA in a volume of 50 μl at 16°C for 12 h. Pre-amplification was carried out with the primer (MP-0) and an anchored primer. The PCR contained 0.5 μM of each primer, 0.25 mM dNTP, 1.5 mM $MgCl_2$, and 1 unit of *Taq* DNA polymerase (Takara, Biotechnology, Japan) in a total volume of 50 μl. The PCR profile was as follows: one cycle of 94°C for 5 min; 25 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 1 min; and a final extension at 72°C for 5 min. For selective amplification, the pre-amplified products were diluted tenfold. Two microliters of the dilution was mixed with 0.5 μM of selected-amplification primer, 0.5 μM of one of the *Mse* I anchors primers, 0.2 mM dNTP each, 1.5 mM $MgCl_2$, and 0.5 unit of *Taq* DNA polymerase in a total volume of 30 μl. The PCR program was as follows: one cycle at 94°C for 5 min; ten "touchdown" cycles of 94°C for 30 s, 64°C for 30 s, and 72°C for 1 min with a decrease in annealing temperature of 1°C in each cycle; 26 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 1 min; and finally 72°C for 5 min to terminate the reaction.

Electrophoresis and fragment detection

Five microliters of the final reaction solution was mixed with 10 μl of loading-buffer (98% formamide, 0.02% BPH, 0.02% Xylene C, and 5 mM of NaOH). After being denatured and immediately cooled, 2 μl of the sample was loaded into 6% denaturing (7.5 M urea) acrylamide-bisacrylamide gel (19:1) and run at 1,600 V for 120 min. The fragments were visualized using silver staining.

SS marker analysis

It was assumed that each amplified band represented a single allele locus. The bands that were present in most *indica* varieties, but absent in *japonica* varieties were referred to as *indica*-specific markers. Conversely, *japonica*-specific markers were those bands that appeared only in *japonica* varieties. To verify the accuracy of these markers, the frequency of *indica* specific allele (*Fi*) or *japonica* specific allele (*Fj*) was calculated by the formula: $Fi = Ni/Nti$ or $Fj = Nj/Ntj$, where *Ni* is the number of cultivars with the same banding patterns as *indica*; *Nj* indicates the number of cultivars with the same banding patterns as *japonica*; and *Nti* and *Ntj* represent the total number of *indica* cultivars and *japonica* cultivars, respectively. To analyze the genetic relationships among *indica* and *japonica*, all the banding patterns were scored as 0 (absent) or 1 (present) in a data matrix and were then subjected to principal component analysis (PCA) using the program MATLAB7.1.

Results

Isolation of genome-specific repetitive DNA sequences

To isolate AA genome-specific repetitive DNA sequences from *O. rufipogon*. A total of 2668 genomic BAC clones were screened three times by Dot-Blot hybridization, and four BAC clones showed strong hybridization signals with *O. rufipogon* genomic DNA as the probe and no or faint signals with *O. officinalis* (Fig. 1). The 4 BAC clones were subcloned and 19

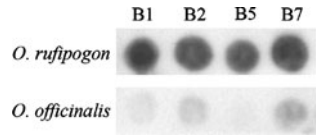

Fig. 1 Four BAC clones contains AA genome-specific repeat sequences have been identified based on the hybridization signal of *O. rufipogon* and *O. officinalis* genomic DNA as probes

subclones (sc1-19) were sequenced (Table 3). Blastn search found that sc1-16 is high homology with the LTR of *Ty3-gypsy* retrotransposon, and sc17-19 high homology with Rice RSOs81X19 insert DNA with tandem repeat (X55420). Based on the characteristics of the subfamily of highly homologous LTR sequences, the *Ty3-gypsy* family retrotransposon can be further divided into several subfamilies, such as *RIRE*, *Tos*, and *Retrosat*. Subclones sc1-16 belonged to the *RIRE* subfamily; sc1-11, sc12-14, sc15, and sc16 belonged to *RIRE3*, *RIRE2*, *RIRE*8, and *Retrosat*1-2, respectively.

The total length of sc1 (EU282824) is 889 bp, and it showed 90% homology with the LTR of *RIRE*3. The total length of sc12 (EU282834) is 1,406 bp, and the sequence from 25 to 466 bp showed 92.3% homology with the LTR of *RIRE*2. Sc1 and sc12 were used as Southern probes to further analyze their patterns in different rice genomes. Six rice species representing AA, BBCC, CC, CCDD and EE genomes were digested with *Hind* III and *Bam* HI and hybridized with the two clones. Both subclones sc1 and sc12 produced strong hybridization signal in the AA genomes (*O. sativa* and *O. rufipogon*), but weak or no hybridization signal in *O. minuta* (BBCC), *O. officinalis* (CC), *O. alta* (CCDD), and *O. australiensis* (EE), suggesting that subclones sc1 and sc12 exist in high copy numbers in the AA genome, and at very low copy numbers in other genomes. The presence of trailing hybridization bands indicated that they were scattered along the AA genome. To test the sequence variation in the AA genomes, three cultivated rice and seven wild rice species were digested with *Hind* III and hybridized with sc1 and sc12. The Southern result showed that the hybridization profiles were similar between cultivated and wild rice. The main difference was the copy number, sc1 and sc12 showed lower copy numbers in *O. perennis* and *O. barthii* than in other species (Figs. 2, 3).

Developing Ty3-gypsy retrotransposon-based SSAP marker for distinguishing indica and japonica

As subclones sc1 and sc12 showed dispersed high copy number character, we designed two types of anchored primers. Anchored primers r2-1 and r2-2 were designed based on sc-12 to detect *RIRE* 2 RBIP. Primers r3-1 and r3-2 were designed based on sc-1 to detect *RIRE* 3 RBIP. The sequence of reverse transcriptase was conserved; therefore, primers rt-1 and rt-2 were designed to detect reverse transcriptase IPP. Depending on the different primer combinations (36 primer combinations), about 20–40 amplified profiles (from 150 to 800 bp) could be detected in a PCR reaction.

In this study, a total of 22 SS markers were identified, including 19 from anchored primers r2-1/r2-2 and three from rt-1/rt-2. However, no SS markers were found from anchored primers r3-1/r3-2. Nine markers (ssi1-9) were specific to subspecies *indica*, and thirteen markers (ssj1-13) were specific to subspecies *japonica*. Only seven primer combinations

Table 3 Summary of subclones based on Blast analysis result

Name of subclones	Retrotransposon subfamily	Distribute pattern	NCBI ID
sc1-11	*RIRE*3	Interspersed repeat	EU282824, EF527879, EF513614, EU282825, EU282826, EU282827, EU282828, EF527880, EU282829, EU282830, EF513615
sc12-14	*RIRE*2	Interspersed repeat	EU282834, EU282835, EU282836,
sc15	*RIRE*8	Interspersed repeat	EU282833
sc16	*Retrosat*1-2	Interspersed repeat	EU282837
sc17-19	R1-R6	Tandem repeat	EU282838, EF513613, EU282832

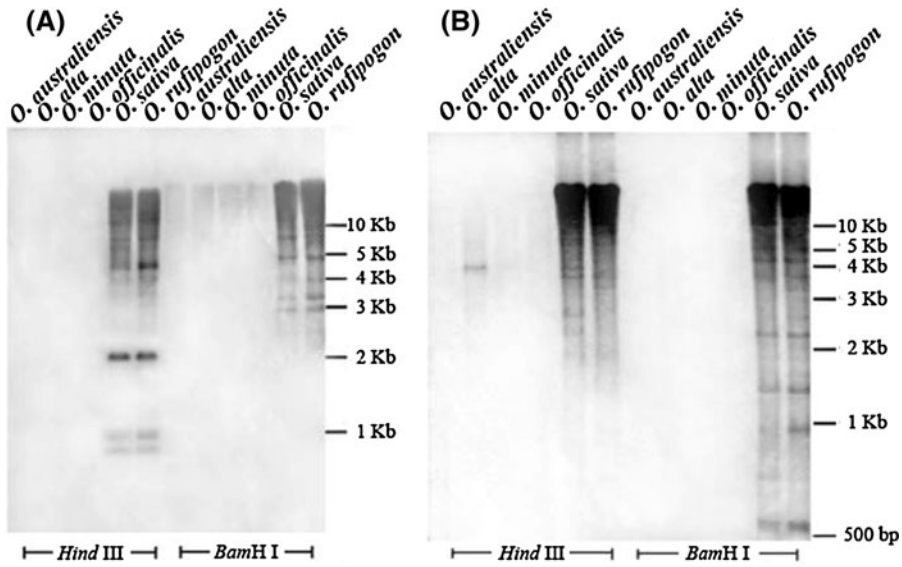

Fig. 2 Southern hybridization result of subclone sc1 (**a**) and sc12 (**b**) hybridized with AA, BBCC, CC, CCDD, EE rice genomes, respectively

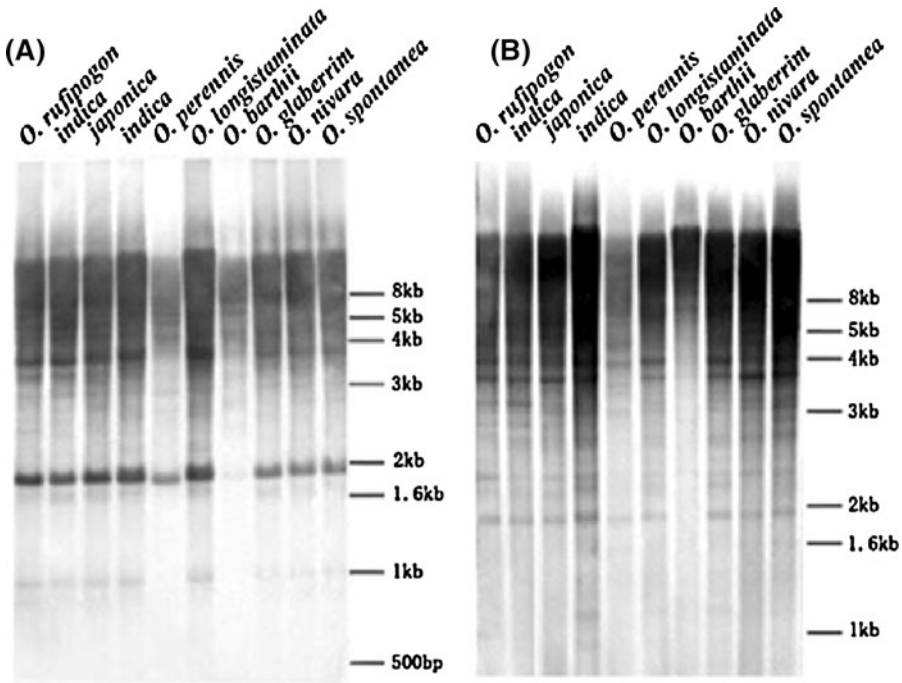

Fig. 3 Southern hybridization result of subclone sc1 (**a**) and sc12 (**b**) hybridized with the cultivated and wild rice of AA genomes

displayed SS markers, accounting for 23.3% of the total combinations (Table 4). The average accuracy in distinguishing between the two subspecies was more than 85%, including ten SS markers with accuracies over 95% (ssj-4, 10, 12, and ssi-1, 4, 5, 6, 7, 8, and 9). Notably, the accuracy of ssj-10 was 100%, and it can distinguish *indica* from *japonica* (Table 4, Fig. 4).

Principal component analysis (PCA) was performed to reveal the relationships of 61 cultivated rice varieties. The 61 cultivated rice varieties were obviously divided into the *indica* and *japonica* groups. Along the first principal component, 31 *indica* varieties from different geographical regions (24 from China, 2 from Thailand, 1 from the Philippines, 1 from India, and 3 from America)

Table 4 SS markers identified in this study

Name	Primer combinations	*Ni*	*Nj*	*Fi*	*Fj*
ssi-1	r2-1/MPGAG	29	2	0.977	0.067
ssi-2	r2-1/MPGAG	23	2	0.767	0.067
ssi-3	r2-1/MPGAA	28	1	0.933	0.033
ssi-4	r2-1/MPGAT	30	2	1.000	0.067
ssi-5	r2-1/MPGAT	30	3	1.000	0.1
ssi-6	r2-1/MPGAT	30	3	1.000	0.1
ssi-7	r2-2/MPGAT	30	1	1.000	0.033
ssi-8	r2-2/MPGAT	29	5	0.977	0.167
ssi-9	rt-1/MPGAA	30	5	1.000	0.167
ssj-1	r2-1/MPCA	4	27	0.133	0.9
ssj-2	r2-1/MPGAA	5	27	0.167	0.9
ssj-3	r2-1/MPGAA	4	27	0.133	0.9
ssj-4	r2-1/MPGAG	5	30	0.167	1.000
ssj-5	r2-1/MPGAG	4	27	0.133	0.9
ssj-6	r2-1/MPGAG	4	27	0.133	0.9
ssj-7	r2-1/MPGAT	1	27	0.033	0.9
ssj-8	r2-1/MPGAT	1	27	0.033	0.9
ssj-9	r2-1/MPGAT	0	17	0	0.567
ssj-10	r2-2/MPCA	0	30	0	1.000
ssj-11	r2-2/MPCA	0	27	0	0.9
ssj-12	rt-1/MPGAA	4	30	0.133	1.000
ssj-13	rt-1/MPGAA	0	25	0	0.833

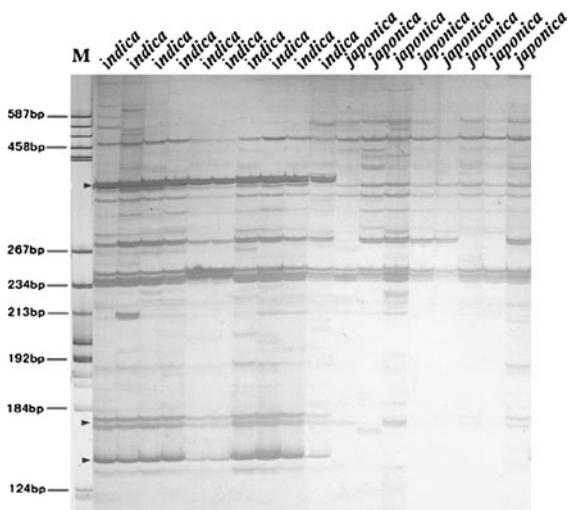

Fig. 4 The *indica*-specific markers amplified by r2-1/MPGAA primer combination. *Arrowhead* indicates the *indica*-specific marker location

formed the *indica* group, and 30 *japonica* varieties (24 from China, 1 from Italy, 1 from the Philippines, 1 from America, and 3 from Japan) also formed the

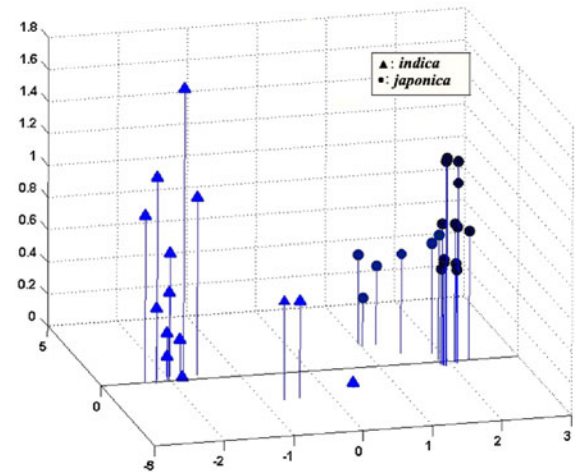

Fig. 5 Scatter plot showing the genetic relationships among the *indica* and *japonica* varieties based on PCA analysis

japonica group. Three China *indica* varieties are scattered near the *japonica* group, indicating a closer genetic relationship with the *japonica* group. The distribution of the *indica* group is more scattered than that of the *japonica* group, which indicates that the *indica* varieties have higher genetic diversity than the *japonica* varieties (Fig. 5).

Discussion

Compared with the SS markers developed using other methods, our marker system is more rapid and efficient. Neeraja et al. (2006) used 372 microsatellite primers covering the entire genome to find SS markers, but only 36 markers (9.68% of the whole set of primers) gave clear polymorphisms on 3% agarose gels. Chin et al. (2007) designed 765 primers by comparing DNA sequences at every 2–3 cM interval between *indica* and *japonica*, and identified 67 markers (8.75% of the total primers) as SS markers. In our study, we found 22 SS markers using 36 primer combinations (61.11% of total primer combinations). More importantly, some primer combinations generated several SS markers, for example, the r2-1/MPGAG primer combination generated *indica* specific markers ssi1-2 and *japonica* specific markers ssj4-6. Combination r2-1/MPGAT generated ssi4-6 and ssj7-8. In addition, the F_1 generation of an inter-species hybrid between *O. sativa* and *O. rufipogon* was used to check the stability of the

markers; 98% of bands from the F_1 were from parents, but only 4% of the dominant bands belonged to the co-dominant type. It is interesting that some non-parental bands also appeared. Their frequency ranged from 1 to 3%, depending on the primer combinations. This phenomenon was also reported by Lee et al. (2005). This suggests that *RIRE* 2 and *RIRE* 3 might be activated during the restructuring of two heterogeneous genetic materials.

The amplification of LTR retrotransposons is main cause of the expansion of the genome (SanMiguel et al. 1996). However, about 97 M LTR retrotransposon sequences have removed by illegitimate recombination and unequal homologous recombination (Ma et al. 2004). During this process, much solo-LTRs with small inserted/deleted fragments were formed and scatted in the rice genome (Vitte and Panaud 2003). The *RIRE* family retrotransposons exist in high copy numbers in *O. sativa*, and are further classified into many sub-families based on the LTR sequences, such as *RIRE* 1–10. *RIRE* 2 is a representative member of the *RIRE* family with a short LTR (440 bp) and has an extra 4-Kb downstream region from the *gag-pol* region (Kumekawa et al. 1999). *RIRE* 3 is a representative member with a very long LTR (2,316 bp) and has an ORF of unknown function in the upstream region of the *gag-pol* region. Does the presence of a longer LTR element correlate with more insertion/deletions polymorphisms and more SS markers? The *RIRE* 3 LTR is longer than *RIRE* 2, however, no SS markers were found using r3-1/r3-2 primer combinations in our results. Vitte et al. (2004) showed that SS markers could be found when the LTR of *RIRE* 8 (3,009 bp) and *Retrosat*1 (441 bp) were analyzed using inserted site-specific primers. Thus, the specificity might be not related to the length of the LTR elements, but related to which type of LTR retrotransposon is chose by inserted site-specific primers.

Small fragment insertion/deletions and base mutations are the main reason for the specificities of markers. Some SS markers of different primer combinations were cloned and sequenced. Two reasons might explain the marker specificity. Firstly, the presence of small fragment insertion/deletions: the sequences of *indica*-specific markers ssi-1 have a deletion of 8 bp, and ssi-7, 8 have a deletion of 11 bp compared with the sequences in *japonica*. The sequences of *japonica*-specific markers ssj-3, 5, and ssj-6 have deletions of 138, 14, and 138 bp, respectively. If an anchored primer pair had the flanking sequence of the insertion/deletion site, polymorphic bands between *indica* and *japonica* would be produced. A longer amplification band could be produced in the "inserted species", and a shorter band produced in the "deleted species". If an anchored primer matched the insertion/deletion site, an amplification band would be produced in one species but not in the other. The second reason is nucleotide base mutation. If mutation events happened in restriction enzyme sites, different restriction fragments between *indica* and *japonica* varieties after *Msel* I digestion would be resulted, e.g. the ssj-1,2 and ssi-2,3 markers.

SS markers are a series of allele sites specifically present in *indica* or *japonica* and are key markers for variety classification, and the study of origin and genetic differentiation of Asian cultivated rice. Previous studies on the differentiation of subspecies and their relationships in rice (Sun et al. 2002; Garris et al. 2005; Londo et al. 2006) used a few markers or unselected markers. LTR retrotransposons comprise of 20% of the rice genome. The SS markers developed here are good markers for distinguishing subspecies. The *japonica* SS markers showed allelic loci in most *japonica* varieties but little or none in *indica* varieties. Markers ssj1-3, and 5-8 were detected in 27 *japonica* varieties, but only in 1-5 *indica* varieties, giving an 90% chance to distinguish two subspecies. Marker ssj-9 was completely lost in all *indica* varieties, and had a low accuracy for *japonica* as it was only detected in 17 *japonica* varieties. However, ssj-10, 11, and 13 showed high specificity in *japonica*, because no allelic loci were detected in any *indica* varieties. Highly accurate distinction was also shown for *indica* SS markers. Markers ssi4, 5, 6, 7, and 9 were detected in all *indica* varieties but were only detected 2–5 in *japonica* varieties, giving an accuracy of more than 90%. Thus, these SS markers, based on *RIRE* retrotransposon LTRs, can accurately distinguish the two subspecies. In addition, marker ssj-11 exists in a high copy number in *japonica* subspecies, but only 2–5 copies of fragments missing 76 bp were detected in *indica*, suggesting that the amplified ssj-11 fragment was selectively lost in the *indica* subspecies but preserved in the *japonica* subspecies during rice genome evolution. Similar phenomena were also found in other studies (Britten et al. 2003; Vitte et al. 2004).

Further study of the flanking sequences of ssj-11 insertion sites in *indica* and *japonica* might provide information on the molecular evolutionary clock. It will not only explain the mechanism behind the missing genomic fragments, but also provide molecular evidence on the differentiation of *O. sativa.*

Rice has two genetically divergent cultivars, *indica* and *japonica*, and ecologically distinct wild progenitors, *O. nivara* and *O. rufipogon*. It is generally agreed that rice *O. sativa* was domesticated from wild Asian species belonging to the AA-genome group of the genus *Oryza* (Chang 1976; Second 1982; Wang et al. 1992; Khush 1997; Ge et al. 1999). However, controversy has persisted on two major issues. One is which wild species, *O. nivara* or *O. rufipogon*, is the direct wild progenitor of cultivated rice. The other is whether rice was domesticated once or multiple times from divergent wild populations (Syed et al. 2005; Zhang et al. 2009; Zhao et al. 2009). Rakshit et al. (2007) considered that *indica* and *japonica* were independently domesticated from *O. rufipogon*. Cluster analysis based on subspecies-specific intron length polymorphism (SSILP) markers support the hypothesis that *indica* and *japonica* were domesticated from the *O. rufipogon* in tropical Asia and in Southern China, respectively (Zhao et al. 2009). In our study, about 31% of total PCR profiles are polymorphic between the two subspecies (*indica* and *japonica*). PCA results supported that subspecies *indica* and *japonica* varieties from different geographical regions can be distinguished by our molecular marker system. Utilization of these markers to analyze Asian typical wild rice and cultivated rice will provide more molecular evidence for cultivated rice domestication. Our results will serve as an important basis for further studies of cultivated rice domestication and the differentiation of *O. sativa*.

Acknowledgments The research was supported by IAEA (12228/RO).

References

Britten RJ, Rowen L, Williams J, Cameron RA (2003) Majority of divergence between closely related DNA samples is due to indels. PNAS 100:4661–4665

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

Chin JH, Kim JH, Jiang WZ, Chu SH, Woo MO, Han LZ, Brar DS, Koh HJ (2007) Identification of subspecies-specific STS markers and their association with segregation distortion in rice (*Oryza sativa* L.). J Crop Sci Biotech 10:175–184

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for plant DNA preparation. Nucleic Acids Res 19:1349

Flavell AJ, Pearce SR, Kumar A (1994) Plant transposable elements and the genome. Curr Opin Genet Dev 4:838–844

Flavell AJ, Knox MR, Pearce SR, Noel Ellis TH (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16:643–650

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch SR (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

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

Hong YH, Qian K, Qin QL, Chen JM (2006) Cloning and analysis of a repetitive DNA from Dong Xiang wild rice (*Oryza rufipogon* Griff.) and its chromosomal localization. J Yangzhou Univ 27:56–61

Jing RC, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis TH, Flavell AJ (2005) Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in Pisum species. Genetics 171:741–752

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

Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E (1999) Identification and characterization of novel retrotransposons of the gypsy type in rice. Mol Gen Genet 260:593–602

Kunze R, Saedler H, Lönnig W-E (1997) Plant transposable element. Adv Bot Res 27:331–370

Kwon SJ, Park KC, Kim JH, Lee JK, Kim NS (2005) Rim 2/Hipa CACTA transposon display: a new genetic marker technique in *Oryza* species. BMC Genet 6:1–15

Lee JK, Kwon SJ, Park KC, Kim NS (2005) Isaac-CACTA transposons: new genetic markers in maize and sorghum. Genome 48:455–460

Liu XC, Ishiki K, Wang WX (2002) Identification of AFLP markers favorable to heterosis in hybrid rice. Breed Sci 52:201–206

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. PNAS 103:9578–9583

Ma JX, Devos K, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869

Mao L, Wood TC, Yu Y, Muhammad A, Budiman MA, Tomkins J, Woo SS, Sasnowski M, Presting G, Frisch D, Goff S, Dean RA, Wing RA (2009) Rice transposable elements: a survey of 73, 000 sequence-tagged-connectors. Genome Res 10:982–990

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasram B, Bustamante CD, Weigel D, Mackill DJ,

Bruskiewich RM, Rätsch G, Buell CR, Leung H, Leach JE (2009) Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. PNAS 106:12273–12278

Monna L, Ohta R, Masuda H, Koike A, Minobe Y (2006) Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars *(Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.). DNA Res 13:43–51

Neeraja CN, Malathi S, Siddiq EA (2006) Subspecies-specific microsatellite markers for rice (*Oryza sativa* L.). J Plant Biochem Biotechnol 15

Oka HI, Chang WT (1959) The impact of cultivation on populations of wild rice, *Oryza sativa f. spontanea*. Phyton 13:105–117

Oka HI, Chang TT (1962) Rice varieties intermediate between wild and cultivated forms and the origin of the Japonica type. Bot Bull Acad Sin 3:109–131

Porceddu A, Albertini E, Barcaccia G, Marconi G, Bertoli FB, Veronesi F (2002) Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol Genet Genom 267:107–114

Qian HR, Zhuang JY, Lin HX, Lu J, Zheng KL (1995) Identification of a set of RFLP probes for subspecies differentiation in *Oryza sativa* L. Theor Appl Genet 90:878–884

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004) Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Gen Genom 271:91–97

Rakshit S, Rakshit A, Matsumura H, Takahashi Y, Hasegawa Y, Ito A, Ishii T, Miyashita NT, Terauchi R (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theor Appl Genet 114:731–743

SanMiguel P, Tikhanov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Schulman AH, Kalendar R (2005) A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. Cytogenet Genome Res 110:598–605

Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Jpn J Genet 57:25–57

Shcherban AB, Vaughan DA, Tomooka N (2000) Isolation of a new retrotransposon- like DNA sequence and its use in analysis of diversity within the *Oryza officinalis* complex. Genetica 108:145–154

Sun CQ, Wang XK, Yoshimura A, Doi K (2002) Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). Theor Appl Genet 104:1335–1345

Syed NH, Sureshsundar S, Wilkinson MJ, Bhau BS, Cavalcanti JJV, Flavell AJ (2005) Ty1-copia retrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.). Theor Appl Genet 110:1195–1202

Vinatzer BA, Zhang HB, Sansavini S (1998) Construction and characterization of a bacterial artificial chromosome library of apple. Theor Appl Genet 97:1183–1190

Vitte C, Panaud O (2003) Formation of Solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in Rice *Oryza sativa* L. Mol Biol Evol 20:528–540

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol Gen Genom 272: 504–511

Vos P, Hogers R, Bleeker M, Reijans M, Lee TVD, Hornes M, Frijters A, Pot J, Peleman J, Kupier M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

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

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W (1997) Genetic distribution of *BARE*-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

Yu LX, Nguyen HT (1994) Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). Theor Appl Genet 87:668–672

Zhang QF, Maroof MAS, Lu TY, Shen BZ (1992) Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. Theor Appl Genet 83:495–499

Zhang Q, Gao YJ, Yang SH, Ragab RA, Maroof MAS, Li ZB (1994) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet 89:185–192

Zhang D, Zhang H, Wang M, Sun J, Qi Y, Wang F, Wei X, Han L, Wang X, Li Z (2009) Genetic structure and differentiation of *Oryza sativa* L. in China revealed by microsatellites. Theor Appl Genet 119:1105–1117

Zhao X, Yang L, Zheng Y, Xu Z, Wu W (2009) Subspecies-specific intron length polymorphism markers reveal clear genetic differentiation in common wild rice (*Oryza rufipogon* L.) in relation to the domestication of cultivated rice (*O. sativa* L.). J Genet Genom 36:435–442