Distribution, proliferation, and transposition

of *mPing* family transposons in genus *Oryza*

2009

Shanta Karki

Kyoto University

Contents   Page

and *O. glaberrima*

Chapter 1

General introduction

1.1 A brief introduction to transposable elements or transposons

Transposable elements (TEs), commonly referred to as the "jumping genes" are mobile, interspersed repetitive DNA sequences that have no fixed place in the genome. They were discovered and characterized in maize by the novel laureate, Barbara McClintock (1950). TEs constitute significant portions of the eukaryotic genomes; for example, human genome has more than 45% (Biemont and Vieira 2006), *Arabidopsis* upto 10%, maize more than 60%, and rice has approximately 35% (IRGSP 2005). Initially they were ignored as the selfish, parasitic, or junk genetic elements (Biemont and Vieira 2006). Interest in TEs increased after the draft of the human genome became available. Recently, their importance has greatly increased as a major factor responsible for creating genetic variability due mainly to their ability to cause somatic mutations through excision, insertion, translocation, and irregular recombination. They can create new genes and increase the number of genes by duplication and exon shuffling (Cowan et al. 2005; Morgante et al. 2005; Jiang et al. 2004b). There are several kinds of TEs which vary in structure, mechanisms of transposition, and the choice of target sites. Based on their transposition intermediates, they are broadly classified into two distinct classes (Benjak et al. 2008). They are described in brief as follows:

Class I TEs (RNA elements): Class I TEs are also known as retrotransposons. They transpose by copy and paste mechanism using RNA as a transposition intermediate. Initially retrotransposons copy themselves to RNA (transcription), then the RNA is copied into DNA by a reverse transcriptase (often coded by the transposon itself) and inserted back into the genome. Such replicative mode of

1

transposition can rapidly increase their copy numbers which in turn swiftly expands the genome size (Slotkin and Martienssen 2007). The mutations induced by retrotransposons are stable because the sequence at the insertion site is retained due to the copy-and -paste mode of transposition. Most of them either have long terminal repeats (known as LTR-retrotransposons) or terminate at one end with poly A tract (known as non-LTR retrotransposons, Fig. 1). The non-LTR retrotransposons are further divided into two groups: long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). In plants, LTR retrotransposons are the most common types of TEs (Feschotte et al. 2002). Moreover, these elements can result in dramatic increases or decreases in genome size and thus are considered to be major players in the remarkable variation of genome size observed in flowering plants (Ammiraju et al. 2007). Recently, it has been reported that three LTR retrotransposon bursts doubled the size of *Oryza australiensis* (Piegu et al. 2006) and a differential proliferation of *Ty-3-gypsy* retrotransposon caused 22% increase in genome size of *O. granulata* (Ammiraju et al. 2007).

Class II TEs (DNA Transposons): The major difference of class II transposons from retrotransposons is that their transposition mechanism does not involve RNA intermediate, they transpose via DNA itself (Kanazawa et al. 2000). DNA transposons usually move using the transposase enzyme, by a mechanism analogous to cut- and -paste, rather than copy- and -paste. Different types of transposase work in different ways. Some of them can bind to any part of the DNA molecule, while others bind to specific sequences. Transposase makes a staggered cut at the target site producing sticky ends, cuts out the transposon and ligates it into the target site. A DNA polymerase fills in the resulting gaps from the sticky ends and DNA ligase closes the sugar-phosphate backbone. This results in target site duplication (TSD) and the insertion sites of DNA transposons may be identified by short direct repeats (a staggered cut in the target DNA filled by DNA polymerase) followed by inverted repeats (which are important for the transposon excision by transposase). However, although *Helitron* is a DNA

transposon, it does not generate TSDs upon transposition because of the specific transposition mechanism, called a rolling-circle mechanism.

Both types of TEs are subdivided into families on the basis of the transposases (TPases) that catalyze their movements. A TE family is composed of one or more autonomous elements and up to several thousand non-autonomous elements. Autonomous TEs code the transposase required for their transposition, whereas, the non-autonomous TEs are mobilized by the enzymes produced by their autonomous counterparts (Yang and Barbash 2008). Mobilization of TEs refers to insertion and excision of TEs which in turn give rise to a series of polymorphisms in closely related species, sub-species, and haplotypes (Huang and Dooner 2008). Such polymorphisms serve as valuable sources of genomic and genetic variations.
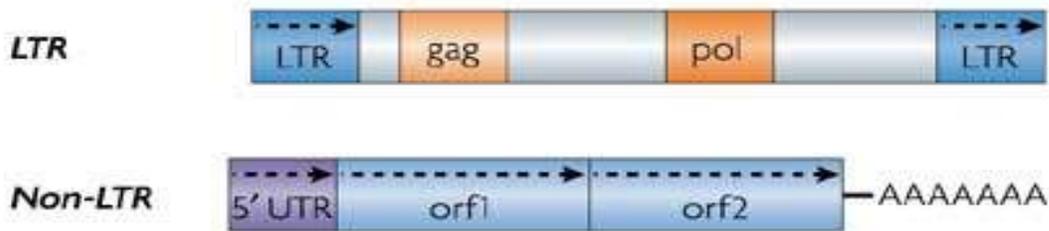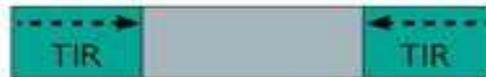
(Slotkin and Martienssen 2007)

Fig. 1. Types and structures of transposable elements.

1.2 Importance and applications of TEs

For many years, TEs were dismissed as junk DNA. Recently, however, they have been revealed as major players in molecular evolution of natural populations (Yang et al. 2005; Biemont and Vieira 2006; Bennetzen 2007). Most eukaryotic genomes, in fact, have been shaped by amplification and dispersion of TEs. TEs serve as building blocks for epigenetic phenomena and hence are responsible for epigenetic regulation of the genome (Slotkin and Martienssen 2007). They also have become useful tools in fucntional genomics and plant molecular biology, especially, for gene tagging. Several studies reported that the TEs were found to regulate expression of the neighboring genes (Kashkush et al. 2003). TEs also control the growth and development of plants with their diverse potentialities to change the genetic materials (Uozu et al. 1997; Bundock and Hooykaas 2005; Fujino et al. 2005; Kapitonov et al. 2006). TEs are considered to be major factors for causing genome variability (Benjak et al. 2008). Therefore, the evolution of transposons and their effects on host genome is currently a very active field of research. TEs are potential endogenous mutators and researchers have used transposons as a source of mutagenesis. In this context, a transposon jumps into a gene and produces a mutation. The presence of a transposon provides a straightforward means of identifying the mutant allele. Sometimes the insertion of a transposon into a gene can disrupt function of the gene in a reversible manner; that is, a transposase mediated excision of the transposon restores gene function. This feature allows researchers to distinguish between genes that must be present inside of a cell in order to function and genes that produce observable effects in cells other than those where the gene is expressed. Furthermore, TEs have been used for phylogenetic analyses due to the easily visible inter- and intra-specific insertion polymorphisms among the species and groups that have diverged from a common anscestor (Takagi et al. 2003; Park et al. 2003; Xu et al. 2007). Perhaps TEs are the most lineage-specific components of their host genomes (Ramirez et al. 2005). They have played a significant role in revealing the history of

biological evolution. They have had a major influence on the structure of genomes during evolution causing mutations and changing the expression patterns of genes. In addition, transposons have been manipulated as useful gene transfer vectors and even used in gene therapy as a non-viral vector system in medical science (Liu and Visner 2007). Therefore, applications of transposons have touched every aspect of the modern biological studies.

1.3 Miniature inverted-repeat transposable elements in plants

Miniature inverted-repeat transposable elements (MITEs) are reminiscence of non-autonomous DNA transposons. They are distinguished from other TEs because of their small size, presence of short terminal inverted repeats (TIRs), high copy numbers, genic preference for insertion, and high sequence identity among the family members (Feschotte et al. 2003; Feschotte et al. 2005). The unique properties of MITEs have been exploited as useful genetic tools for plant genome analysis. Although MITEs were first discovered in plants and are still actively reshaping plant genomes, they have been isolated from a wide range of eukaryotic organisms including mosquitoes (Tu 1997). They have been divided into tourist-like, stowaway-like, and pogo-like groups, according to similarities of their TIRs and TSDs (target site duplications). Numerically, the MITEs are the predominant type of TEs in rice genome (Turcotte et al. 2001). There are over 100,000 MITEs in the rice genome (Jiang et al. 2003). Several MITEs are found to insert in low-copy, genic regions of plant genomes. Due to their high copy numbers and insertion preferences near to or into the genic regions, they play crucial role in promoting the genetic diversity of plants during their evolutionary and domestication processes through the genomic changes that they bring about.

The understanding of the co-evolution MITEs and their host genomes, the mechanisms to regulate transposition events, and the insertion specificities is still in its infancy. Transposition and

survival mechanisms of transposons within the host genome are possibly regulated both by transposon- and host-encoded factors, to avoid deleterious effects on host and transposon as well, in a relationship that has existed for many millions of years between transposons and their hosts. The evolution of transposons and their effects on host genome is currently a very active field of research.

1.4 Overview of rice (*Oryza sativa* L.) genome

Rice (*Oryza sativa* L.), which belongs to genus *Oryza* is one of the most important food crops of the world. The genus *Oryza* has been divided into 24 species, ten genome types, four species-complexes and different ploidy levels (6 diploid and 4 polyploid genomes). The 24 species include 22 wild and two cultivated species (*O. sativa*- worldwide cultivated rice and *O. glaberrima*-cultivated rice in Africa) each of which has either 24 (diploid) or 48 (amphidiploid) chromosomes (Aggarwal et al. 1999; Park et al. 2003a). The ten genome types are: AA, BB, CC, BBCC, CCDD, EE, FF, GG, HHJJ and HHKK (Aggarwal et al. 1997; Aggarwal et al. 1999). Based on their genome structures, the Oryza species are further grouped into four complexes known as *O. sativa*, *O. officinalis*, *O. ridleyi*, and *O. granulata* (Vaughan et al. 2005). The *O. sativa* complex which includes the cultivated rice species consists of the AA genome species namely *O. barthii, O. glaberrima*, *O. glumaepatula*, *O. longistaminata*, *O. meridionalis*, *O. nivara, O. rufipogon*, and *O. sativa*, among which *O. nivara* and *O. rufipogon* have been considered as the direct wild progenitors of *O. sativa* (Wang et al. 1992; Park et al. 2003; Vaughan et al. 2005). Other species complexes include *O. officinalis* complex which comprises of *O. officinalis* (CC), *O. minuta* (BBCC), *O. punctata* (BB, BBCC), *O. rhizomatis* (CC), *O. eichingeri* (CC), *O. latifolia* (CCDD), *O. alta* (CCDD), *O. grandiglumis* (CCDD), and *O. australiensis* (EE), *O. ridleyi* complex consists of *O. ridleyi* and *O. longiglumis* both of which have HHJJ genome; and *O. granulata* complex includes *O. granulata* and *O. meyeriana* both of them have GG genome. *O. brachyantha* (FF) and *O. schlechteri* (HHKK) do not fall into any of these species complexes and are considered to represent

species on the boundary of the genus *Oryza* (Vaughan et al. 2005). The availability of these well characterized wild relatives and whole genome sequence make rice an excellent ideal experimental material to explore the genus-level transposon dynamics. On the other hand, the wild *Oryza* species are considered to be reservoirs of useful genes that can be utilized in future breeding programs to further improve the production potentials of the existing cultivated rice varieties (Ge et al. 1999; Cheng et al. 2002; Ni et al. 2002; Guo et al. 2005). Some of the important traits possessed by the wild species include: drought resistance, salt tolerance, and resistance to various diseases and pests such bacterial leaf blight, stem borer, and green leaf hopper. To make use of the important favorable traits of wild species, the comparative genomics has to be well characterized. The comprehensive study of TEs can provide suitable DNA markers for isolation of such genetic elements.

1.5 Objectives of the present study

MITEs have major roles in genome diversification and expansion. Due to their ability to self-replicate, they can proliferate and reach high copy numbers and, if fixed, can be retained in evolutionary lineages across wide taxonomic groups. However, origin, evolution, and specific biological functions of MITES remain enigmatic because very little is known about their distribution among species in different geographical regions. Moreover, the information regarding their polymorphisms both within and between the species, with various degrees of phylogenetic separation, is not sufficient. The genus *Oryza* is an ideal system to study the origin and the roles of MITEs on host genome evolution because it harbors an active MITE, *mPing* and has many wild relatives habituated to various parts of the world. Comparing the presence and absence of MITEs across species that have diverged from a common ancestor, it will be possible to identify the timing and dynamics of transposon insertions and deletions during the evolutionary history of rice. The *mPing* is an active rice MITE which is actively transposing even in intact rice plants under natural conditions (Naito et al. 2006). Autonomous elements, *Ping* and *Pong* also have been identified and were found to mobilize *mPing* (Jiang et al. 2003; Kikuchi et al. 2003; Yang et al. 2007). For these reasons, analysis of distribution, proliferation, and transposition of *mPing* family transposons in genus *Oryza* is essential to understand the roles of MITEs on the host genomes.

The present research works were undertaken to investigate the presence and distribution patterns of eight types of MITEs in genus *Oryza*. Relationships between existence of MITEs and evolutionary events of genus *Oryza* were analyzed. Furthermore, distribution and proliferation of *mPing*, *Ping*, and *Pong* (the *mPing* family transposons) in *Oryza rufipogon* accessions were studied to explore the clues to trace the origin of *mPing* family transposons. The genealogical studies of *O. rufipogon* accessions were conducted through haplotype analysis around *qSH1*, which a major QTL responsible for seed shattering

in rice. Finally, the ability of inter-specific hybridization to mobilize the *mPing* family transposons were evaluated using the interspecific inbred lines derived from cross between *O. sativa* and *O. glaberrima* to develop the transposon tagging system for wild rices.

Chapter 2

Differential distribution of miniature inverted-repeat transposable elements in wild *Oryza* species

2.1 Introduction

MITEs are a subset of DNA transposons. They are characterized by their small size (less than 600 bp) with short (10-30 bp) terminal inverted–repeats (TIRs) at each ends, flanked by short direct repeats of 2-8 bp known as the target site duplication (TSD) (Takagi et al. 2003), and are present in high copy numbers in a genome. Amongst the various types of non-autonomous class II transposons, MITEs form the major components of the rice genome (Jiang et al. 2003; Jiang et al. 2004a). For instance, sequence analysis of the rice chromosome 4 has shown that the MITEs account for nearly 50% of all the numbers of repetitive DNA (Feng et al. 2002). Moreover, MITEs reach several thousands in copy numbers and are frequently associated with the genes (Feschotte and Pritham 2007). However, very little is known about their evolution, transposition mechanisms, and specific biological functions because their distribution among species in different geographical regions is not well understood.

In the present study, presence and distribution of eight types of rice specific MITEs were investigated in 19 species of the genus *Oryza* to trace their origin during the evolutionary history of genus *Oryza*.

2.2 Materials and methods

2.2.1 Plant materials

The seeds or plantlets of 44 accessions of 18 wild *Oryza* species were provided by the National Institute of Genetics, Japan. In addition, *O. sativa* cultivars Nipponbare (*O. sativa* ssp *japonica*) and IR 36 (*O. sativa* ssp *indica*) were used. The accession number, genomic constitutions, name of the species and the country of habitats are presented in Table 1. Sterilized distilled water was used as template for negative control in every experiment.

2.2.2 DNA extraction

The DNA extraction of every plant material was carried out either by the modified CTAB method of Murray and Thompson (1980) or by DNeasy Plant Mini Kit (Qiagen, Tokyo, Japan) according to the manufacturer's protocol.

Table 1. List of the plant materials used in the present study

S.N.	Genome	Acc. no.	Species	Country
1	AA	Nipponbare	*O. sativa (japonica)*	Japan
2	AA	IR36	*O. sativa (indica)*	Philippines
3	AA	W0106	*O. rufipogon*	India
4	AA	W0120	*O. rufipogon*	India
5	AA	W1294	*O. rufipogon*	Philippines
6	AA	W1866	*O. rufipogon*	Thailand
7	AA	W1921	*O. rufipogon*	Thailand
8	AA	W2003	*O. rufipogon*	India
9	AA	W1625	*O. meridionalis*	Australia
10	AA	W1635	*O. meridionalis*	Australia
11	AA	W0652	*O. barthii*	Sierra Leone
12	AA	W1588	*O. barthii*	Cameroon
13	AA	W1169=W1165	*O. glumaepatula*	Cuba
14	AA	W2145	*O. glumaepatula*	Brazil
15	AA	W2199	*O. glumaepatula*	Brazil
16	AA	W1413	*O. longistaminata*	Sierra Leone
17	AA	W1508	*O. longistaminata*	Madagascar
18	BB	W1514	*O. punctata*	Kenya
19	BBCC	W1024	*O. punctata*	Ghana
20	BBCC	W1213	*O. minuta*	Philippines
21	BBCC	W1331	*O. minuta*	Philippines
22	CC	W1527	*O. eichingeri*	Uganda
23	CC	W1805	*O. eichingeri*	Sri Lanka
24	CC	W0002	*O. officinalis*	Thailand
25	CC	W1361	*O. officinalis*	Malaysia
26	CC	W1830	*O. officinalis*	Unknown
27	CCDD	W0017	*O. alta*	Surinam
28	CCDD	W1182	*O. alta or O. latifolia*	British Guinea
29	CCDD	W0613	*O. grandiglumis*	Brazil
30	CCDD	W1194	*O. grandiglumis*	Brazil
31	CCDD	W2220	*O. grandiglumis*	Brazil
32	CCDD	W1166	*O. latifolia*	Mexico
33	CCDD	W1197	*O. latifolia*	Colombia
34	CCDD	W2200	*O. latifolia*	Brazil
35	EE	W0008	*O. australiensis*	Australia
36	EE	W1628	*O. australiensis*	Australia
37	FF	W1401	*O. brachyantha*	Sierra Leone
38	FF	W1711	*O. brachyantha*	Cameroon
39	GG	W0003	*O. granulata*	India
40	GG	W0067(B)	*O. granulata*	Thailand
41	GG	W1356	*O. meyeriana*	Malaysia
42	HHJJ	W1220	*O. longiglumis*	Dutch New Guinea
43	HHJJ	W0001	*O. ridleyi*	Thailand
44	HHJJ	W0604	*O. ridleyi*	Malaya

2.2.3 Selection of MITEs

A total of eight rice (*O. sativa*) specific MITEs: *Castaway_Os1*, *Ditto_Os1* (Yang et al. 2007), *Kiddo* (Yang et al. 2001), *Mashu* (Takagi et al. 2003), *mPing*, *Stowaway2_Os* like (Kanazawa et al. 2000), *Stowaway17_Os* like, and *Wanderer* were selected from the database based on their sequence homogeneity and high copy numbers (Table 2).

Table 2. Name of the MITEs analyzed in this study and the PCR primers used to amplify them

Name of MITEs	NCBI accession no.	Forward primers (5' to 3')	Reverse primers (5' to 3')	Size (bp)
Castaway-Os1	AC021891	GTCCCTTTGAATCATAGGGTTG	GCGCCATTTGAATGAAATGA	360
Ditto-Os1	AF488413	GAGCAAGTTTAATAGTATAGCCA	GAGCAGGTACAATAGCATGCTA	244
Kiddo	AF484680	GGGGCTGTTTGGTTCCCAGCCA	TTTGGTTGCAAGCTACACTTTG	269
Mashu	AB077839	AATGGTAAAGTAAGGTGCTCTC	GGGCACCCRCAATGGTTATCTA	263
mPing	BK000588	GGGATGAGAGAGAAGGAAAGAG	AACAATCCCCACAGTGGAG	430
Stowaway2_Os like	AF488413	TCCATATTTTAATATATAACGC	CTCCCTCCGTATTTTAATGTATG	235
Stowaway17_Os like	AB092509	CTCCCTCCATACTCATAAAGGA	CTCCGTACTTATAAATGAAATCG	259
Wanderer	AC134517	TCTCGTTTTCCGCGCGCATGC	GTCTGAGGAGAAGGGGATTG	208

2.2.4 PCR amplification

To investigate the presence or absence of above mentioned MITEs, PCR was carried out using element specific primers (Table 2). The genomic DNA of each accession was used as the templates for PCR using Ex Taq DNA polymerase (TaKaRa, Shiga, Japan). The PCR conditions were as follows: pre-denaturation for 3 min at 94 ℃ followed by 30 cycles of polymerization reaction, each consisting of a denaturation step for 10 s at 98 ℃, an annealing step for 1 min at 60 ℃ and an extension step for 1 min at 72 ℃, with a final extension step for 7 min at 72 ℃. PCR products were visualized under UV light following electrophoresis on 1.4% agarose gel in 0.5x TBE buffer.

2.2.5 Cloning, sequencing, and data analysis

For each MITE, 5 to 10 PCR amplicons were cloned. Cloning was done using TOPO TA Cloning Kit Version R (Invitrogen, Nihonbashi, Japan) according to its instruction manual. Plasmid DNA consisting of PCR product was transformed into *Escherichia coli* DH5α strain. The extraction of plasmid DNA was done using the Plasmid DNA Purification Kit (MACHEREY-NAGEL, Duren, Germany) following the procedures mentioned in its user manual. The cloned fragments were sequenced. DNA sequencing was performed following the protocol of CEQ 8000 Genetic Analysis System (BECKMAN COULTER, Tokyo, Japan). The sequence data were analyzed by nucleotide-nucleotide BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and CLUSTALW (http://align.genome.jp/) to confirm the presence of each MITE.

2.3 Results

2.3.1 Differential distribution of eight MITEs in *Oryza* species

When PCR using specific primers for *Castaway-Os1* was carried out, only AA genome species excepting *O. meridionalis* yielded PCR product of appropriate size (ca. 360 bp), whereas other species yielded larger or no product (Fig. 2). Sequencing revealed that they were of same sizes and had identical sequences. Thus, it was confirmed that *Castaway-Os1* was present in all the AA genome species excluding *O. meridionalis*. In the same way, it was found that *Mashu* (Fig. 6), *Stowaway17_Os like* (Fig. 8), and *Wanderer* (Fig. 9) are present in all the species of AA genome. On the other hand, *Kiddo* (Fig. 5) and *Stowaway2-Os* (Fig. 7) like MITE have wider distribute in genus *Oryza* than the four MITEs mentioned above. *Kiddo* is present in all the AA, BB, BBCC, and CC genome species (Fig. 5), and *Stowaway2-Os like* is present in *O. punctata* (BB, BBCC), *O. minuta* (BBCC) besides all of the AA genome species (Fig. 7). *Ditto_Os1* showed the widest distribution among the eight MITEs assessed, because it is present in the species with BB, BBCC, CC, CCDD, EE, and FF genomes in addition to all of the AA genomic species (Fig. 3). MITE *mPing* is present only in two of the AA genome species, *O. rufipogon* and *O. sativa*, among the 19 species used in this study (Fig. 6), indicating that *mPing* has the narrowest distribution. These results are summarized in Table 3.
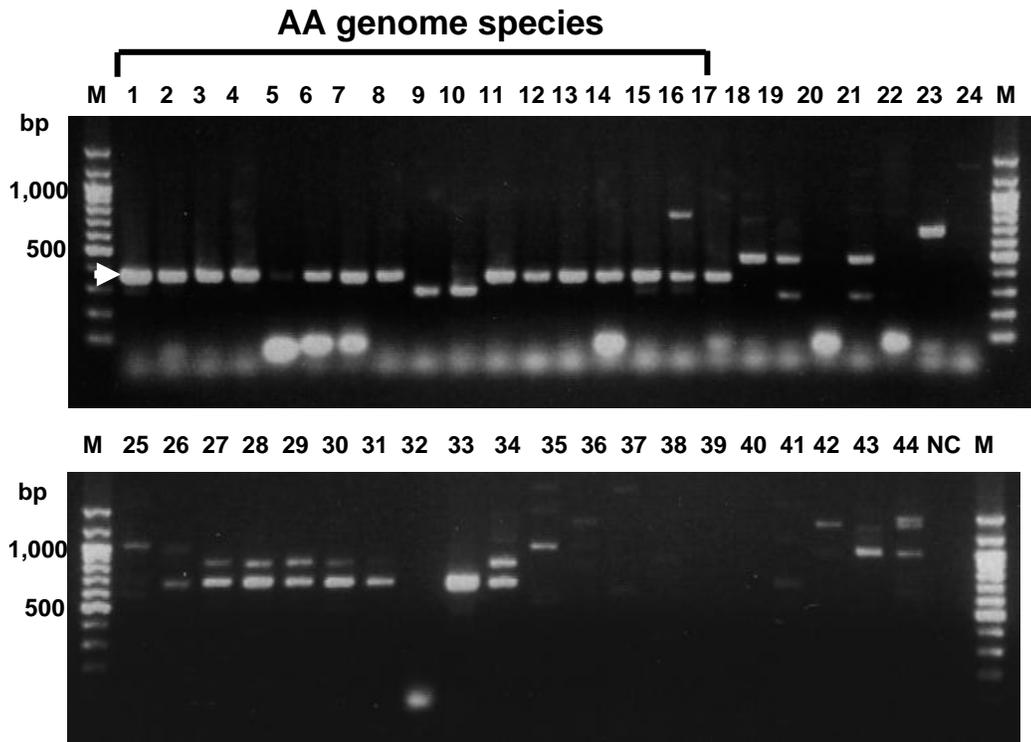
Fig. 2. Investigation of presence or absence of *Castaway-Os1*. M: DNA size marker (100 bp ladder, Nacalai, Japan); Lanes 1-2: *Oryza sativa*; lanes 3-8: *O. rufipogon*; lanes 9-10: *O. meridionalis*; lanes 11-12: *O. barthii*; lanes 13-15: *O. glumaepatula*; lanes 16-17: *O. longistaminata*; lanes 18-19: *O. punctata*; lanes 20-21: *O. minuta*; lanes 22-23: *O. eichingeri*; lanes 24-26: *O. officinalis*; lanes 27-28: *O. alta*; lanes 29-31: *O. grandiglumis*; lanes 32-34: *O. latifolia*; lanes 35-36: *O. australiensis*; lanes 37-38: *O. brachyantha*; lanes 39-40: *O. granulate*; lane 41: *O. meyeriana*; lane 42: *O. longiglumis*; lanes 43-44: *O. ridleyi*; lane NC: negative control (distilled water). Lane numbers represent the serial numbers of the accessions in Table 1. Arrow shows the perspective band (ca. 360-bp). This MITE is present only in the AA genome species.

18

Fig. 3. Examination of *Ditto_Os 1* MITE. Lane numbers represent the serial numbers of the accessions in Table 1. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan); lane NC: negative control (distilled water was used as the template). Arrowheads show the band of expected size (ca. 244- bp). The amplicons of accurate size were amplified in species belonging to AA, BB, BBCC, CC, CCDD, EE, and FF genomes.

Fig. 4. Determination of presence or absence of *Kiddo* MITE. Lane numbers represent the serial numbers of the accessions in Table 1. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan. Arrowhead represents the band of expected size (ca. 269- bp). Specific bands were amplified just in AA, BB, BBCC, and CC genome species.
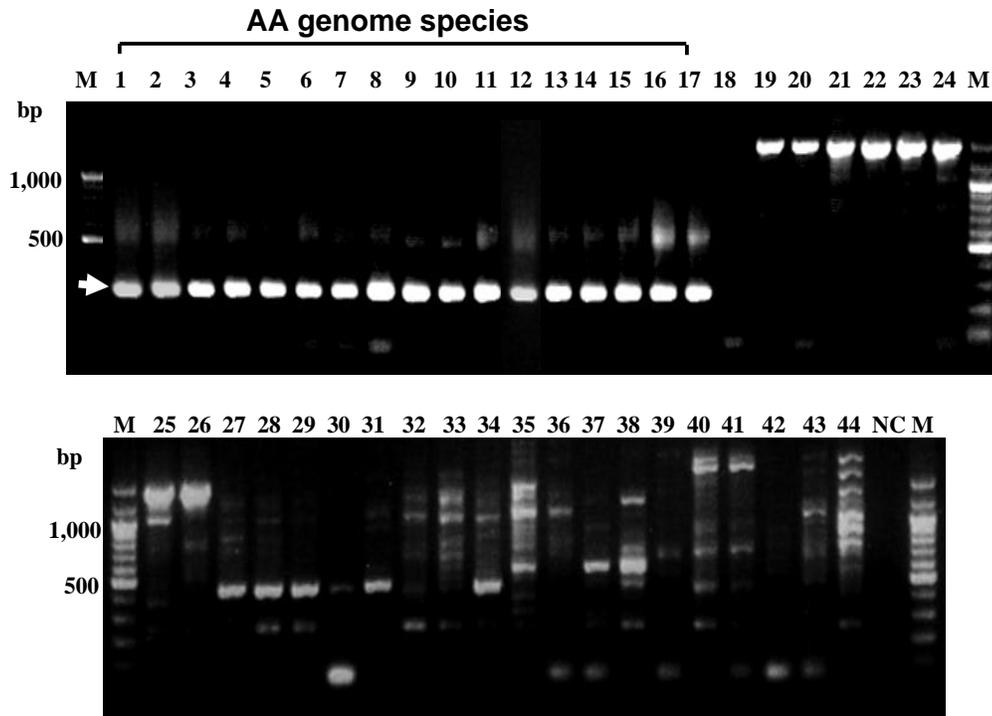
Fig. 5. Determination of presence or absence of *Mashu* MITE. Lane numbers represent the serial numbers of the accessions in Table 1. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan. Arrow shows the band of expected size (ca. 263- bp). Specific bands were amplified just in AA genome species.
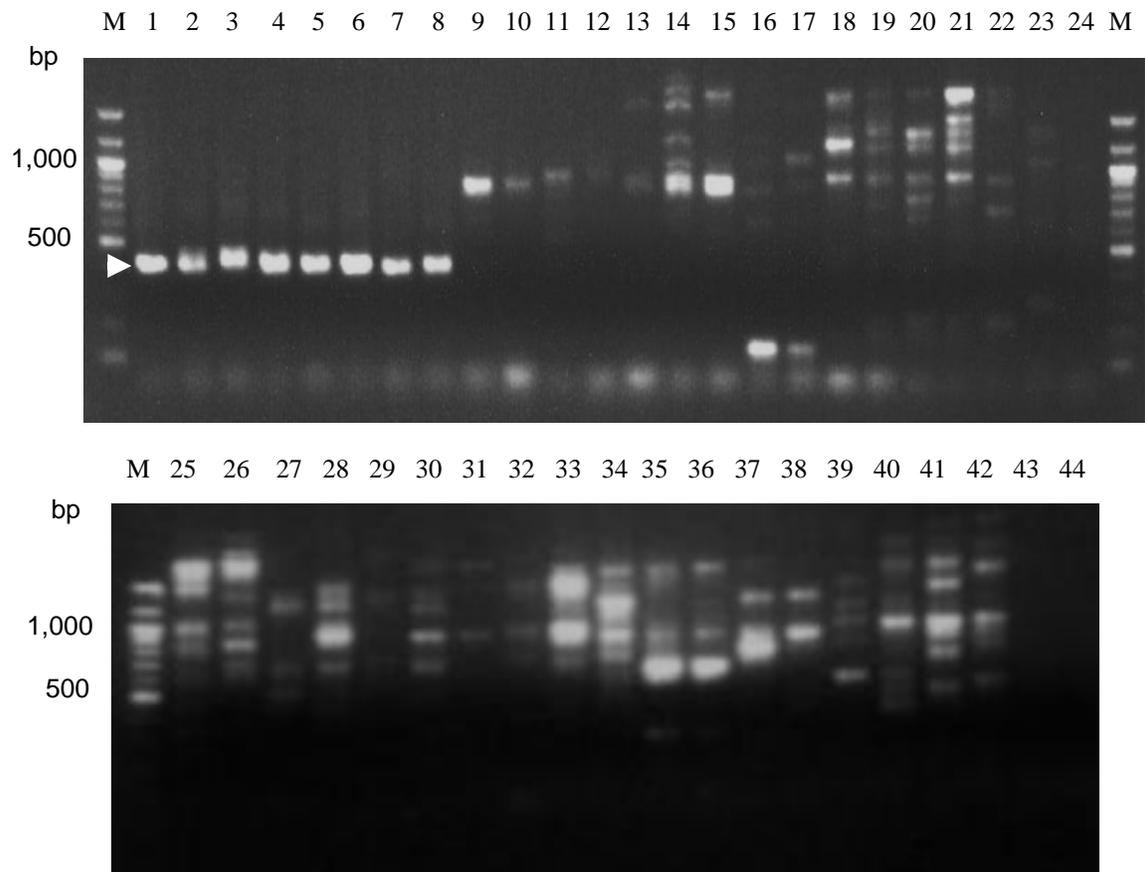
Fig. 6. Examination of *mPing* MITE. Lane numbers represent the serial numbers of the accessions in Table 1. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan. Arrowhead shows the band of expected size (ca. 430- bp). Specific bands were amplified just in *O. sativa* and *O. rufipogon* accessions.
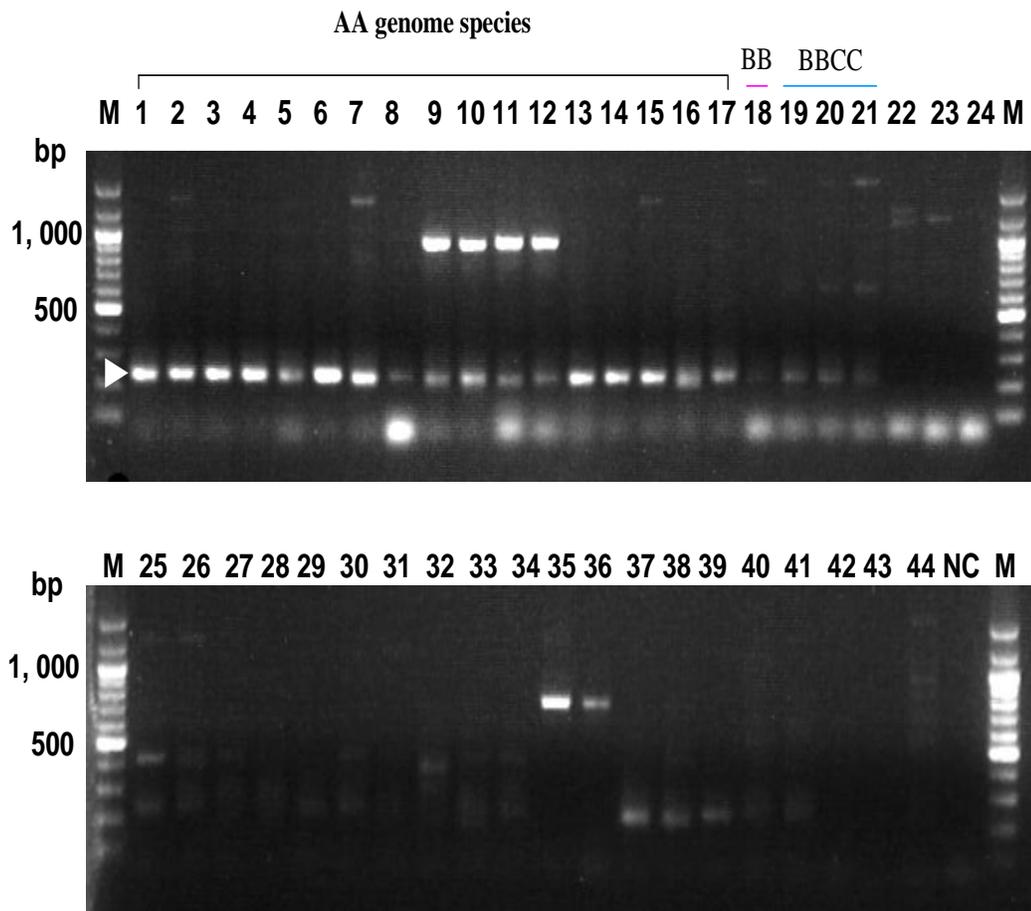
Fig. 7. Determination of the distribution of *Stowaway2_Os* like element in genus *Oryza*. Lane numbers represent the serial number of the accessions in Table 1. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan. Arrowhead represents the band of expected size (ca. 235- bp). Specific bands were amplified just in AA genome accessions.
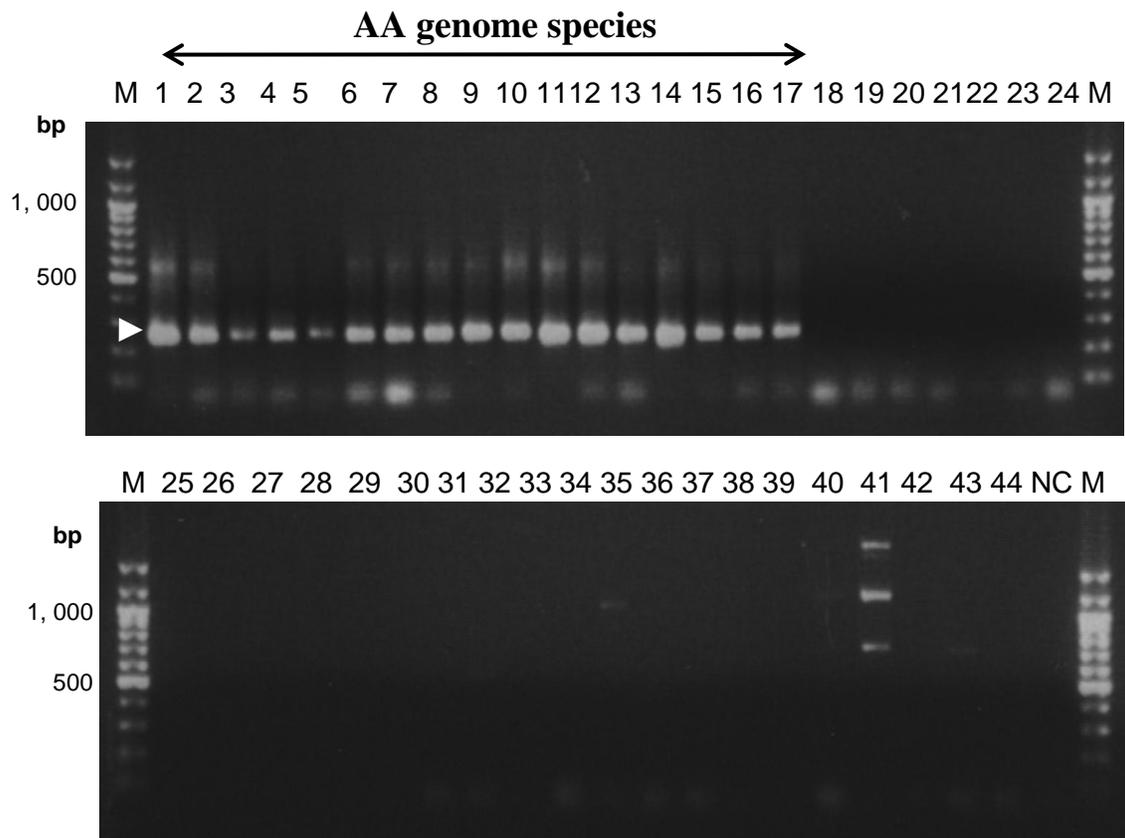
Fig. 8. Determination of the distribution of *Stowaway17_Os like* element in genus *Oryza*. Lane numbers represent the serial number of the accessions in Table 1. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan. Arrowhead represents the band of expected size (ca. 259- bp). Specific bands were amplified just in AA genome accessions.
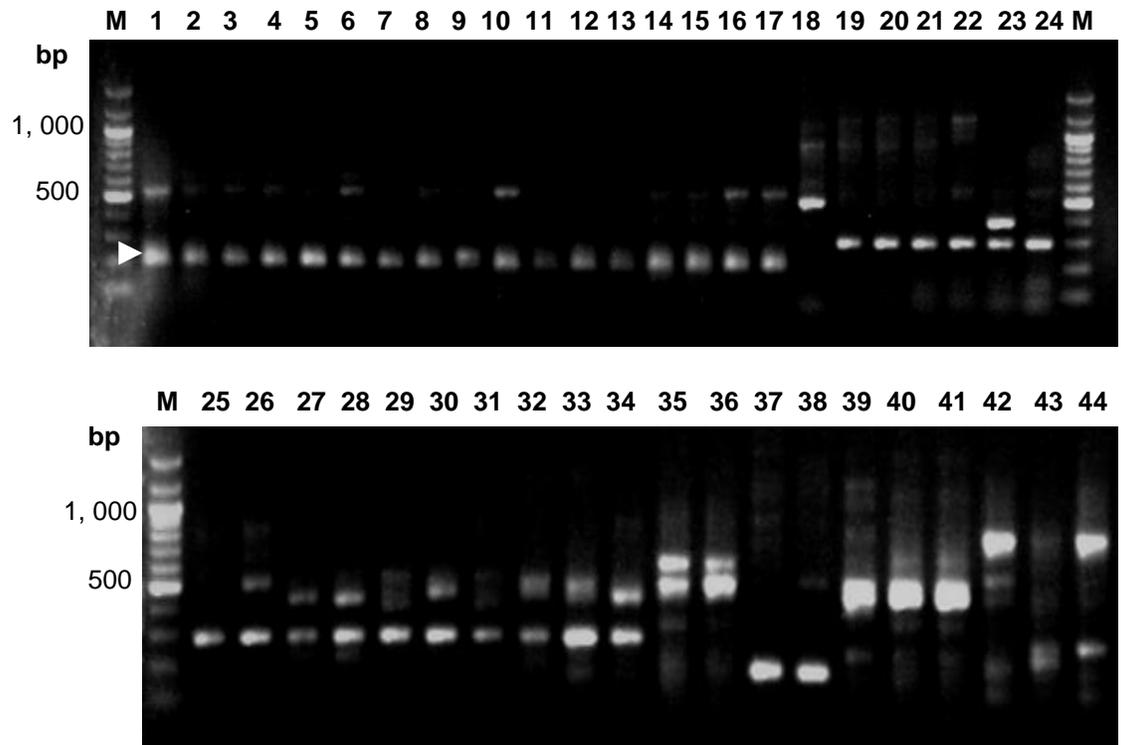
Fig. 9. Determination of the distribution of *Wanderer* MITE in genus *Oryza*. Lane numbers represent the serial number of the accessions in Table 1. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan. Arrowhead represents the band of expected size (ca. 208- bp). Specific bands were amplified just in AA genome accessions.

Table 3. Summary of the results for investigation of MITEs

Species	Presence or absence of the MITEs analyzed							
	mPing	Kiddo	Ditto-Os1	Stowaway1 7_Os like	Castaway-Os1	Mashu	Stowaway2 _Os like	Wanderer
Oryza sativa	+	+	+	+	+	+	+	+
O. rufipogon	+	+	+	+	+	+	+	+
O. barthii	-	+	+	+	+	+	+	+
O. glumaepatula	-	+	+	+	+	+	+	+
O. longistaminata	-	+	+	+	+	+	+	+
O. meridionalis	-	+	+	+	-	+	+	+
O. punctata	-	+	+	-	-	-	+	-
O. minuta	-	+	+	-	-	-	+	-
O. eichingeri	-	+	+	-	-	-	-	-
O. officinalis	-	+	+	-	-	-	-	-
O. alta	-	-	+	-	-	-	-	-
O. grandiglumis	-	-	+	-	-	-	-	-
O. latifolia	-	-	+	-	-	-	-	-
O. australiensis	-	-	+	-	-	-	-	-
O. brachyantha	-	-	+	-	-	-	-	-
O. granulata	-	-	-	-	-	-	-	-
O. meyeriana	-	-	-	-	-	-	-	-
O. longiglumis	-	-	-	-	-	-	-	-
O. ridleyi	-	-	-	-	-	-	-	-

+: Presence and -: Absence

2.3.2 Relationships between the presence of MITEs and speciation in genus *Oryza*

To verify the evolutionary history of each MITE, the existence of each MITE was applied to the phylogenetic tree of genus *Oryza* adopted from Zhu and Ge (2005) (Fig. 10). All of the eight *Oryza sativa* specific MITEs were present in AA genome *Oryza* species, which is the most recent divergent among the 10 genomes of genus *Oryza*. Furthermore, five MITEs were present in all the AA genome species. *Ditto_Os1* was present even in ancient species such as *O. brachyantha*. However, *mPing* was present in only two species, namely *O. sativa* and *O. rufipogon*, all of which are genetically closely related to each other. It indicates that *mPing* had evolved in those species after their differentiation and speciation from rest of the other AA genome species (Fig. 10). *Castaway-Os1* is detected in all of the AA genome species excepting *O. meridionalis*, the earliest divergent lineage in AA genome (Zhu and Ge 2005). Thus, it was considered that *Castaway-Os1* MITE has inserted into the AA genome species after *O. meridionalis* separated from the other AA genome species. Similar results were obtained in earlier studies in which three of the MITE insertions were detected in all the AA genome species except for *O. meridionalis* (Zhu and Ge 2005). These results indicate that the specific MITEs originated at different times during the divergence and evolutionary phases of genus *Oryza*.
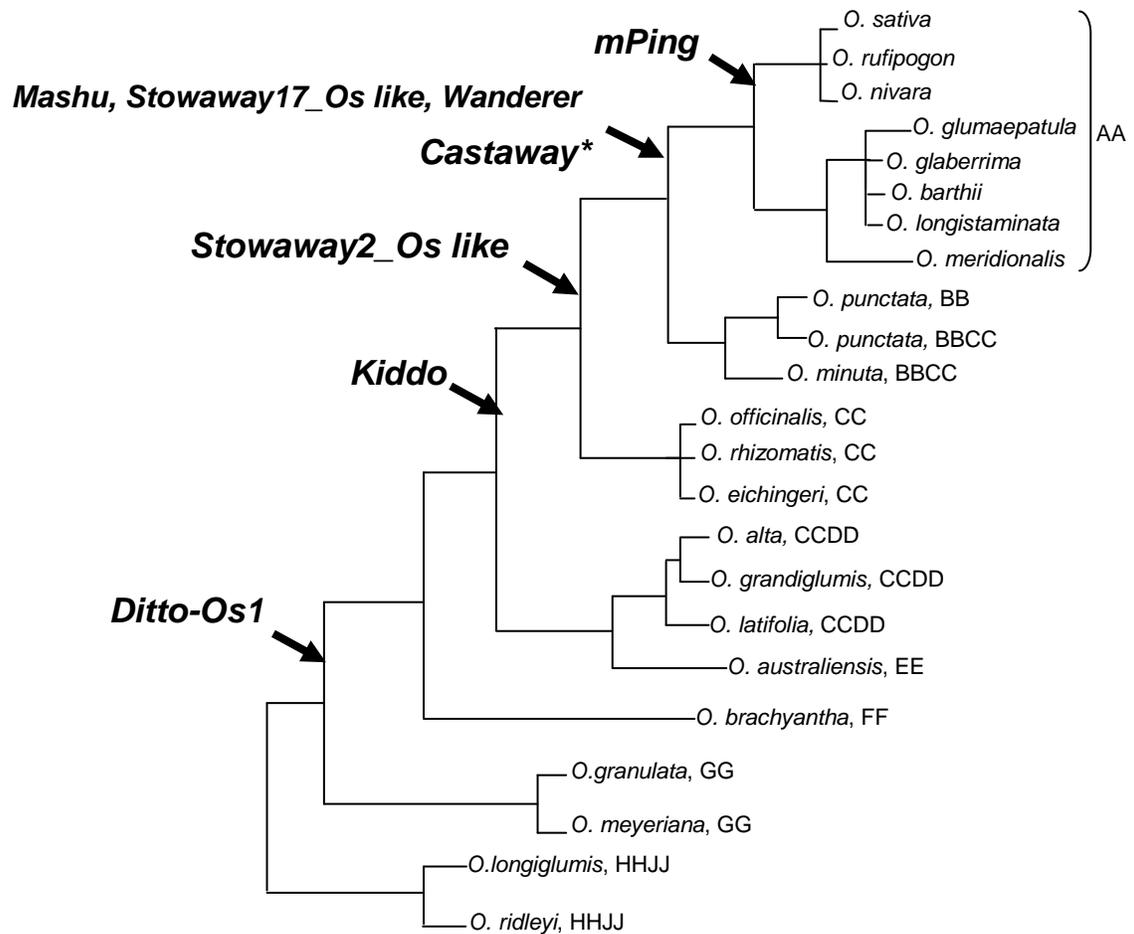
Fig. 10. An evolutionary model for eight MITEs analyzed. Phylogenetic

tree of *Oryza* species is based on Zhu and Ge (2005).

*: Not present in *O. meridionalis*.

28

2.4 Discussion

Mashu, *Castaway_Os1*, *mPing*, and *Stowaway17_Os like* MITEs were present in only the AA genome species, whereas other MITEs were present in the other genome species besides the AA genome species. This indicates that these four MITEs have been originated relatively later than other MITEs. Furthermore, *mPing* was found to have originated latest than the other MITEs during the evolutionary process of genus *Oryza*, because *mPing* was present only in *O. sativa* and *O. rufipogon* among 19 wild species surveyed. Hence, it was concluded that *mPing* is comparatively a young MITE in genus *Oryza*. On the other hand, *Kiddo* and *Stowaway17_Os like* were present not only in AA genome species, but also in BB, CC, and BBCC genome species. Similarly, *Ditto_Os1* was present even in CCDD, EE, and FF genome species besides AA, BB, BBCC, and CC genomes, indicating that *Ditto_Os1* was an old MITE family due to its early origin during the differentiation history of genus *Oryza* into several genome types and species. It was reported that *Tourist C* element, which is one of the MITEs in rice, was present in a functional domain in 5' region of *CatA* gene and existed in all four *Oryza* species complexes (Iwamoto et al. 1999). Moreover, *Pangrangja* elements were found to be present in AA, BB, CC, BBCC, CCDD, and EE genome species, and were proven to be useful to study the genetic variations and species relationships in *Oryza* (Park et al. 2003a; 2003b). Thus, it was considered that *Ditto_Os1* might be present in an important chromosomal region of genus *Oryza*.

Transpositions of TEs result in addition or removal of the DNA sequences in the host genome, thereby giving rise to sequence polymorphisms. Such indels cause inter-

and intra-specific variations which serve as useful markers to infer phylogenetic relationships (Kanazawa et al. 2000; Park et al. 2003). Likewise, the presence or absence of MITEs is one of the major sources for promoting genetic variations in natural populations. Thus, genome and/or species-specific MITEs might have contributed significantly to the adaptation, differentiation, speciation, and diversification processes of the genus *Oryza*.

Chapter 3

Analysis of distribution and proliferation of *mPing* family transposons in a wild rice species (*Oryza rufipogon* Griff.)

3.1 Introduction

The genus *Oryza* contains 22 wild and two cultivated species, each of which has either 24 (diploid) or 48 (amphidiploid) chromosomes, and has been divided into ten genome types, viz., AA, BB, CC, BBCC, CCDD, EE, FF, GG, HHJJ and HHKK (Aggarwal et al. 1997; 1999). Based on their genome structures, the *Oryza* species are further grouped into four complexes known as *O. sativa*, *O. officinalis*, *O. ridleyi* and *O. granulata* (Vaughan et al. 2005). The *O. sativa* complex consists of seven species including *O. sativa* and *O. rufipogon*. Although *O. rufipogon* is undoubtedly the ancestor of *O. sativa*, the origin and the evolutionary history of cultivated rice remain unclear.

Transposable elements (TEs) are mobile DNA sequences which can change their positions in the genome by inserting themselves into new sites. For many years, TEs were dismissed as selfish DNA; recently, however, TEs have been revealed as major players in genomic evolution, because they cause genome rearrangements and alter the structure and regulation of individual genes (Bennetzen et al. 2000; Biemont and Vieira 2006). Most eukaryotic genomes, in fact, have been shaped by amplification and dispersion of TEs (Le et al. 2000). In rice (*O. sativa* L.), TEs account for at least 35% of the genome (International Rice Genome Sequencing Project 2005); the numerically predominant type of TE is the miniature inverted-repeat transposable element (MITE) (Bureau et al. 1996; Mao et al. 2000; Turcotte et al. 2001; Feng et al. 2002; Jiang et al. 2004a). Taken together, these facts suggest that MITEs have significantly contributed to rice genome evolution.

Miniature *Ping* (*mPing*) is the first active MITE discovered in the rice genome (Jiang et al. 2003; Kikuchi et al. 2003; Nakazaki et al. 2003). Because *mPing* is a 430-bp sequence with 15-bp terminal inverted repeats (TIRs) which lacks an open reading frame (ORF), its mobilization depends on the transposase encoded by the ORF2 in the autonomous elements *Ping* and *Pong* (Jiang et al. 2003; Kikuchi et al. 2003; Yang et al. 2007). Both *Ping* and *Pong* also have the ORF1, in addition to the ORF2, coding a protein with weak similarity to the DNA-binding domain of *myb* transcription factor (Jiang et al. 2003). Although the copy number of *mPing* is low compared with those of other characterized MITEs in plants (Feschotte et al. 2002), it varies among the rice species; temperate *japonica* cultivars possess large numbers of *mPing* (≈50 copies), whereas tropical *japonica* and *indica* cultivars possess small numbers of *mPing* (<10 copies) (Jiang et al. 2003; Kikuchi et al. 2003). Moreover, it has been reported that the *mPing* family, which consists of *mPing*, *Ping* and *Pong*, is present only in *O. sativa* and its direct ancestor *O. rufipogon* (Hu et al. 2006). Based on this observation, *mPing* family transposons have been considered to play important roles in the genome evolution of *Oryza sativa*. Very little is known, however, about the evolutionary history through which the *mPing* family was distributed and proliferated in the genus *Oryza*.

Here, the insertional polymorphisms and the diversity in copy number of *mPing* in *O. rufipogon* accessions were shown. Furthermore, a possible evolutionary history of the genomic constitution of the *mPing* family was described.

3.2 Materials and methods

3.2.1 Plant materials

A total of 78 accessions of 19 wild *Oryza* species together with two *O. sativa* varieties (ssp *japonica* Nipponbare and ssp *indica* IR36) were used. The genomic constitutions, species name, and the habitat of each accession are presented in Table 4.

Table 4. List of the plant materials used in the present study

S.N.	Genome	Acc. no.	Species	Country	SINE Code
1	AA	Nipponbare	*O. sativa (japonica)*	Japan	
2	AA	IR36	*O. sativa (indica)*	Philippines	
3	AA	W0106	*O. rufipogon*	India	AA110000
4	AA	W0120	*O. rufipogon*	India	AA110000
5	AA	W1294	*O. rufipogon*	Philippines	AA110000
6	AA	W1866	*O. rufipogon*	Thailand	AA110000
7	AA	W1921	*O. rufipogon*	Thailand	AA110000
8	AA	W2003	*O. rufipogon*	India	AA110000
9	AA	W1625	*O. meridionalis*	Australia	AA100010
10	AA	W1635	*O. meridionalis*	Australia	AA100010
11	AA	W0652	*O. barthii*	Sierra Leone	AA101000
12	AA	W1588	*O. barthii*	Cameroon	AA101000
13	AA	W1169=W1165	*O. glumaepatula*	Cuba	AA101100
14	AA	W2145	*O. glumaepatula*	Brazil	AA101100
15	AA	W2199	*O. glumaepatula*	Brazil	AA101100
16	AA	W1413	*O. longistaminata*	Sierra Leone	AA100001
17	AA	W1508	*O. longistaminata*	Madagascar	AA100001
18	BB	W1514	*O. punctata*	Kenya	
19	BBCC	W1024	*O. punctata*	Ghana	
20	BBCC	W1213	*O. minuta*	Philippines	
21	BBCC	W1331	*O. minuta*	Philippines	
22	CC	W1527	*O. eichingeri*	Uganda	
23	CC	W1805	*O. eichingeri*	Sri Lanka	
24	CC	W0002	*O. officinalis*	Thailand	
25	CC	W1361	*O. officinalis*	Malaysia	
26	CC	W1830	*O. officinalis*	Unknown	
27	CCDD	W0017	*O. alta*	Surinam	
28	CCDD	W1182	*O. alta or O. latifolia*	British Guinea	
29	CCDD	W0613	*O. grandiglumis*	Brazil	
30	CCDD	W1194	*O. grandiglumis*	Brazil	
31	CCDD	W2220	*O. grandiglumis*	Brazil	
32	CCDD	W1166	*O. latifolia*	Mexico	
33	CCDD	W1197	*O. latifolia*	Colombia	
34	CCDD	W2200	*O. latifolia*	Brazil	
35	EE	W0008	*O. australiensis*	Australia	
36	EE	W1628	*O. australiensis*	Australia	
37	FF	W1401	*O. brachyantha*	Sierra Leone	
38	FF	W1711	*O. brachyantha*	Cameroon	
39	GG	W0003	*O. granulata*	India	
40	GG	W0067(B)	*O. granulata*	Thailand	
41	GG	W1356	*O. meyeriana*	Malaysia	
42	HHJJ	W1220	*O. longiglumis*	Dutch New Guinea	
43	HHJJ	W0001	*O. ridleyi*	Thailand	
44	HHJJ	W0604	*O. ridleyi*	Malaya	
45	AA	W0630	*O. rufipogon*	Burma	AA110000
46	AA	W1236	*O. rufipogon*	Australian New Guinea	AA110000
47	AA	W1807	*O. rufipogon*	Sri Lanka	AA110000
48	AA	W1945	*O. rufipogon*	China	AA110000
49	AA	W2051	*O. rufipogon*	Bangladesh	AA110000
50	AA	W2078	*O. rufipogon*	Australia	AA110000
51	AA	W2263	*O. rufipogon*	Cambodia	AA110000
52	AA	W0107	*O. rufipogon*	India	AA110000
53	AA	W0108	*O. rufipogon*	India	AA110000
54	AA	W0137	*O. rufipogon*	India	AA110000
55	AA	W0180	*O. rufipogon*	Thailand	AA*10000
56	AA	W0593	*O. rufipogon*	Malaya	AA110000
57	AA	W0610	*O. rufipogon*	Burma	AA110000
58	AA	W1230	*O. rufipogon*	Dutch New Guinea	AA110000
59	AA	W1235	*O. rufipogon*	New Guinea	AA100010
60	AA	W1238	*O. rufipogon*	New Guinea	AA110000
61	AA	W1239	*O. rufipogon*	New Guinea	AA100010
62	AA	W1551	*O. rufipogon*	Thailand	AA110000
63	AA	W1666	*O. rufipogon*	India	AA110000
64	AA	W1669	*O. rufipogon*	India	AA110000
65	AA	W1681	*O. rufipogon*	India	AA110000
66	AA	W1685	*O. rufipogon*	India	AA110000
67	AA	W1690	*O. rufipogon*	Thailand	AA110000
68	AA	W1715	*O. rufipogon*	China	AA110000
69	AA	W1852	*O. rufipogon*	Thailand	AA110000
70	AA	W1865	*O. rufipogon*	Thailand	AA110000
71	AA	W1939	*O. rufipogon*	Thailand	AA110000
72	AA	W1981	*O. rufipogon*	Indonesia	AA110000
73	AA	W2014	*O. rufipogon*	India	AA110000
74	AA	W2080	*O. rufipogon*	Australia	AA100010
75	AA	W2109	*O. rufipogon*	Australia	AA110000
76	AA	W2265	*O. rufipogon*	Laos	AA110000
77	AA	W2266	*O. rufipogon*	Laos	AA110000
78	AA	W2267	*O. rufipogon*	Laos	AA110000

3.2.2 DNA extraction

Genomic DNA was extracted from the young leaves, either by the modified Cetyltrimethylammonium bromide (CTAB) method of Murray and Thompson (1980) or with a DNeasy Plant Mini Kit (Qiagen, Tokyo, Japan) according to the manufacturer's protocol.

3.2.3 PCR amplification

To determine whether *mPing* was present in each accession, PCR was carried out using the following two *mPing*-specific primer pairs: *mPing*-F1 (5′-GGCCAGTCACAATGGGGGTT-3′) and *mPing*-R1 (5′- GGCCAGTCACAATGGCTAGTG-3′) and *mPing*-F2 (5′-GGGATGAGAGAGAAGGAAAGAG-3′) and *mPing*-R2 (5′-AACAATCCCCACAGTGGAG-3′). The *mPing*-F1 and *mPing*-R1 were designed on the 5′ and 3′ ends of *mPing* and thus to amplify the whole *m-Ping* sequence (430-bp), whereas the *mPing*-F2 and *mPing*-R2 were designed in the internal regions of *mPing*.

To determine whether *Ping* and *Pong* were present in each genome, we used the following specific primer pairs designed to complement *Ping*-ORF1, *Ping*-ORF2, *Pong*-ORF1, and *Pong*-ORF2, respectively : *Ping*-ORF1-F1 (5′-ACAAGCGGATACTCCGAC-3′) and *Ping*-OFR1-R1 (5′-GAGGACAATGCCTTCCATAAC-3′); *Ping*-ORF2-F1 (5′-AAGGTATTTGCTAAGTCATAT-3′) and *Ping*-ORF2-R1 (5′-ATGAGCCAAACGATCTCTCATACTA-3′); *Pong*-ORF1-F1 (5′-TCTACTCCACCACCAACACC-3′) and *Pong*-ORF1-R1 (5′-TCAGCCTTGTTTTTGTCCTTC-3′); and *Pong*-ORF2-F1 (5′-CTCATACGAAGACCTCCTCC-3′) and *Pong*-ORF2-R1 (5′-CGCCTAAGATACCTCTCACC-3′).

35

To detect the MITE *Kiddo*, the following primer sequences were used: *Kiddo*-F (5′-GGGGCTGTTTGGTTCCCAGCCA-3′) and *Kiddo*-R (5′-TTTGGTTGCAAGCTACACTTTG-3′).

The genomic DNA of each accession was used as the template for PCR using Ex-*Taq* DNA polymerase (TaKaRa, Shiga, Japan). The PCR conditions consisted of: pre-denaturation for 3 min at 94 ℃ followed by 30 cycles of polymerization reaction, each consisting of a denaturation step for 10 s at 98 ℃, an annealing step for 45 s at 60 ℃ and an extension step for 45 s at 72 ℃, with a final extension step for 7 min at 72 ℃. PCR products were visualized under UV light following electrophoresis on 1.5 % agarose gel in 0.5x TBE buffer.

3.2.4 Transposon display

The copy numbers of *mPing* and *Kiddo* were determined by transposon display (TD) according to the protocols of Casa et al. (2000) and Takata et al. (2005), respectively. The genomic DNA of each accession of *O. rufipogon*, Nipponbare, and IR36 was digested with *Mse*I or *Msp*I, and ligated to an adapter.

Primary PCR was carried out using an adapter-specific primer (*Mse*I-1: 5'-GACGATGAGTCCTGAGTAA-3' or *Msp*I-1: 5'-GTAATACGACTCACTATAGGGC-3') and an MITE-specific primer (*mPing*-P1: 5'-AATGTGCATGACACACCAG-3' or *Kiddo*-P1: 5'-CAAGTAATGGTAAAGTGTGGCTGGG-3'). Secondary PCR was carried out using a one-base selective primer (*Mse*I-2: 5'-GACGATGAGTCCTGAGTAAN-3' or *Msp*I-2: 5'-GTGCTTGATGCTTGAAAACGG-3') and another MITE-specific primer (*mPing*-P2: 5'-CAGTGAAACCCCCATTGTGAC-3' or *Kiddo*-P2: 5'-AATGGTAAAGTGTGGCTGGGAAC-3') labeled with TAMRA fluorescence reagent.

Primary PCR conditions were as follows: pre-denaturation for 2 min at 72 °C and subsequent pre-denaturation for 3 min at 94 °C, followed by 30 cycles of polymerization reaction, each consisting of a denaturation step for 30 s at 95 °C, an annealing step for 45 s at 58 °C, and an extension step for 45 s at 72 °C, with a final extension step for 7 min at 72 °C. PCR products were diluted between 20- and 50-fold with 1/10 TE buffer and used as templates for secondary PCR (selective amplification). Selective amplification cycling started with pre-denaturation for 3 min at 94 °C followed by 7 cycles of touchdown PCR. The first touchdown cycle consisted of a denaturation step for 30 s at 95 °C, an annealing step for 45 s at 64 °C, and an extension step for 45 s at 72 °C; starting with the second cycle, temperature at the annealing step was lowered by 1.0 °C per cycle until it reached 58 °C. Touchdown PCR was followed by 30 cycles of polymerization reaction, each consisting of a denaturation step for 30 s at 95 °C, an annealing step for 45 s at 58 °C, and an extension step for 45 s at 72 °C, with a final extension step for 7 min at 72 °C.

3.2.5 Phylogenetic analysis

Nucleotide polymorphisms around *qSH1*, a major quantitative trait locus for seed shattering in rice (Konishi et al. 2006), were investigated in *O. rufipogon* accessions using the following four primer pairs: Con24-13U: 5′-CAATGGAAAAGCTGATG-3′ and Con24-13L: 5′-CGTTGCATGAATTGTAGCAC-3′, RBEL-E1-U: 5′-ATCATGCAGCAAGTGACCAC-3′ and RBEL-E1-L2: 5′-TCACAACCTAGAGATGAGGC-3′, 106.5k-U: 5′-CAATACATGCATGGATGCGT-3′ and 107.2k-L: 5′-TCCATATGTGTGTAGGACCC-3′, and 108.9k-U: 5′-ACAGGGTGATCCCAACAGTT-3′ and 109.8k-L: 5′-TAACCGGTGATGGTTGTGCA-3′. The PCR conditions were the same as described above.

PCR products were sequenced using ABI PRISM DNA sequencing system (Grainer Japan, Tokyo, Japan). After the sequence data obtained from the four primer pairs were combined, the sequence alignment was determined using the online computer program CLUSTAL W (http://align.genome.jp/). A non-rooted phylogenetic tree was constructed using MEGA version 3.1 (Kumar et al. 2004). The neighbor-joining method (Saitou and Nei 1987) was conducted with Kimura's 2-parameter distances (Kimura 1980).

3.3 Results

3.3.1 *mPing* is a specific MITE in *O. sativa* and *O. rufipogon*

Recently, the *mPing* family was shown to be distributed only in the cultivated rice *O. sativa* and the wild rice *O. rufipogon* (Hu et al. 2006). Because its distribution and proliferation have not previously been substantially discussed, we set out to investigate the distribution of *mPing* among wild rice species by applying PCR to 78 accessions covering nine of the ten genomic constitutions of rice. According to our PCR analysis using *mPing*-F1 and *mPing*-R1 primers which were designed on the 5′ and 3′ ends of *mPing*, respectively, and thus to amplify the whole *m-Ping* sequence (430 bp), only *O. sativa* and *O. rufipogon* accessions yielded PCR products of appropriate size (approximately 430 bp), whereas other species yielded products that were too long to be *mPing* (Fig. 11). Sequencing analysis showed that the PCR products with sequences of approximately 430 bp all shared very high similarity with the *mPing* sequence (98-100%). In previous studies, the sequences nearly identical to *mPing* but with a few nucleotide polymorphisms have been regarded as subtypes of *mPing* (Jiang et al. 2003; Kikuchi et al. 2003; Hu et al. 2006); it was, on the other hand, concluded that these PCR products obtained using

38

mPing-F1 and *mPing*-R1 primers were in fact *mPing*, although they were not completely identical with the typical *mPing* sequence. Similar results were obtained from the PCR analysis with *mPing*-F2 and *mPing*-R2 primers, which were designed at the internal position of *mPing* (Fig. 12). These results confirm that *mPing* is deployed only in *O. sativa* and *O. rufipogon*. Four (W1235, W1239, W2014, and W2080) of the forty *O. rufipogon* accessions did not yield any PCR products corresponding to *mPing*, indicating that these accessions do not harbor *mPing* in their genomes. These four accessions are classified as *O. rufipogon* based on their morphological characteristics, but three of them, W1235, W1239, and W2080, have been assigned the SINE code of *O. meridionalis* (Table 4). Thus they are considered to have been misclassified as *O. rufipogon* based solely on their morphological characteristics; based on their genetic characteristics, these three accessions have treated as *O. meridionalis* in further analyses.
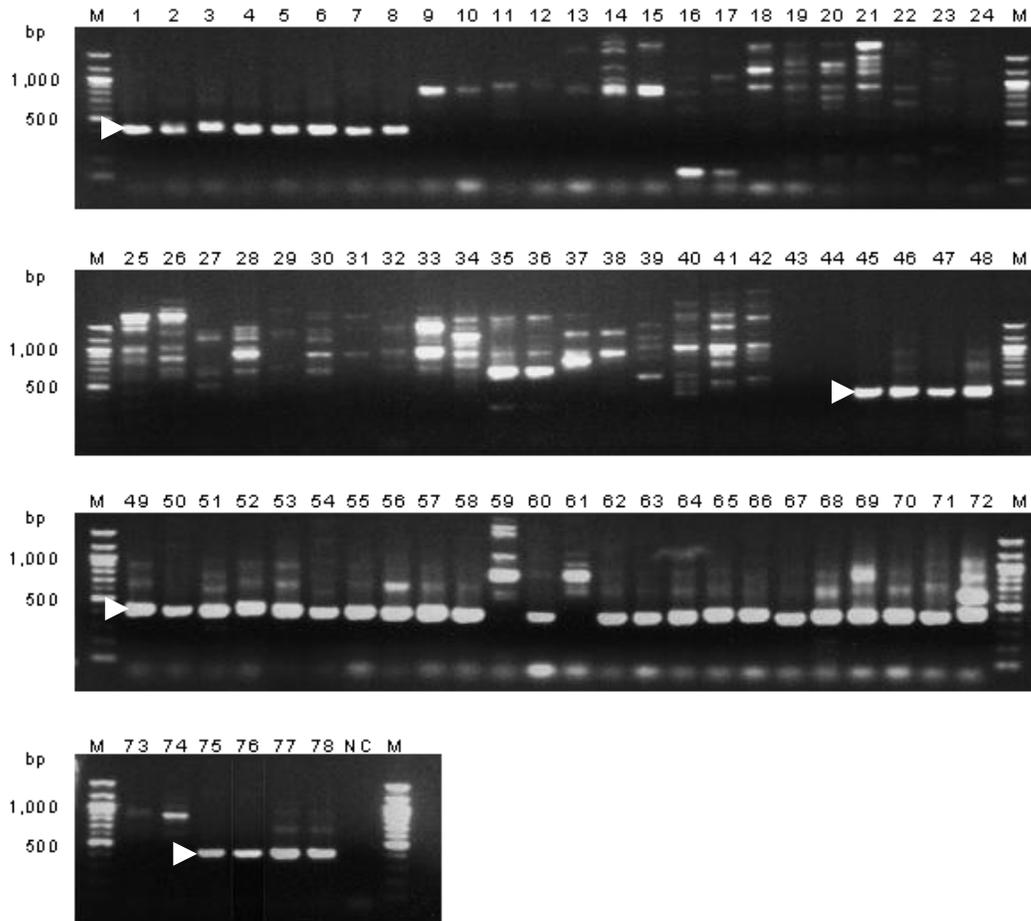
Fig. 11. Determining the presence or absence of *mPing* in wild *Oryza* species by PCR analysis using *mPing*-F1 and *mPing*-R1 primers. M, DNA size marker (100 bp ladder, Nacalai, Japan); Lanes 1-2, *Oryza sativa*; lanes 3-8, *O. rufipogon*; lanes 9-10, *O. meridionalis*; lanes 11-12, *O. barthii*; lanes 13-15, *O. glumaepatula*; lanes 16-17, *O. longistaminata*; lanes 18-19, *O. punctata*; lanes 20-21, *O. minuta*; lanes 22-23, *O. eichingeri*; lanes 24-26, *O. officinalis*; lanes 27-28, *O. alta*; lanes 29-31, *O. grandiglumis*; lanes 32-34, *O. latifolia*; lanes 35-36, *O. australiensis*; lanes 37-38, *O. brachyantha*; lanes 39-40, *O. granulate*; lane 41, *O. meyeriana*; lane 42, *O. longiglumis*; lanes 43-44, *O. ridleyi*; lanes 45-78, *O. rufipogon*; lane NC, negative control (distilled water). Arrow heads indicate the PCR amplicons of correct size (ca. 430-bp).
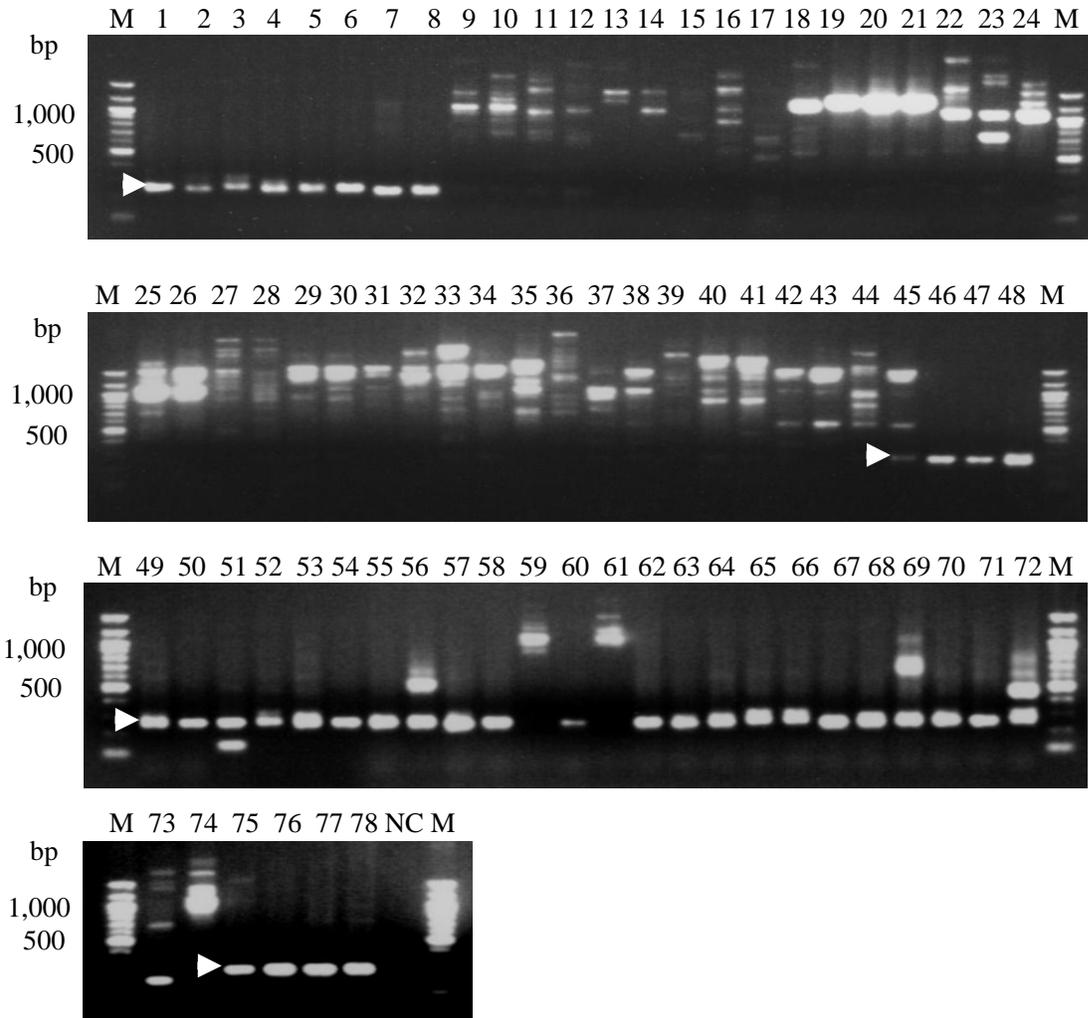
Fig. 12. Confirmation of presence or absence of *mPing* in wild *Oryza* species by PCR analysis using the internal primer pairs (*mPing*-F2 and *mPing*-R2). The lane labels are as same as in Fig. 11. Arrow heads represent the PCR products of correct sizes (ca. 230-bp).

3.3.2 Distribution of the autonomous elements *Ping* and *Pong* in genus *Oryza*

mPing has been thought to be a deletion derivative of *Ping*, because the terminal sequences of *mPing* are identical to those of *Ping* except for a single base-pair mismatch. It is known, however, that not only *Ping* alone but also *Pong* alone can mobilize *mPing* (Jiang et al. 2003; Kikuchi et al. 2003; Yang et al. 2007). This observation underscores the necessity of investigating whether the wild species with *mPing* also possess *Ping* and/or *Pong* elements.

PCR analyses were carried out using *Ping*- and *Pong*-specific primer pairs. Nipponbare (*O. sativa* ssp. *japonica*) and 17 of 37 *O. rufipogon* accessions yielded *Ping*-ORF1 (264 bp) and *Ping*-ORF2 (369 bp) as PCR products (Fig. 13). W2078 and W1690 of *O. rufipogon* yielded *Ping*-ORF2 but not *Ping*-ORF1. In PCR of *Ping*-ORF1, *O. eichingeri* (W1527 and W1805) and *O. officinalis* (W0002, W1361, and W1830) each produced a single band around 500 bp, but sequencing of these bands confirmed that they were not the specific product of *Pong*-ORF1 (data not shown). Along with Nipponbare and IR36 (*O. sativa* ssp. *indica*), all *O. rufipogon* accessions produced both *Pong*-ORF1 (500 bp) and *Pong*-ORF2 (586 bp) as PCR products (Fig. 14). These results show that *Ping* and *Pong* are present only in *O. sativa* and *O. rufipogon*, but also that the distribution of *Ping* is limited, even within *O. rufipogon,* to certain accessions (Table 5).
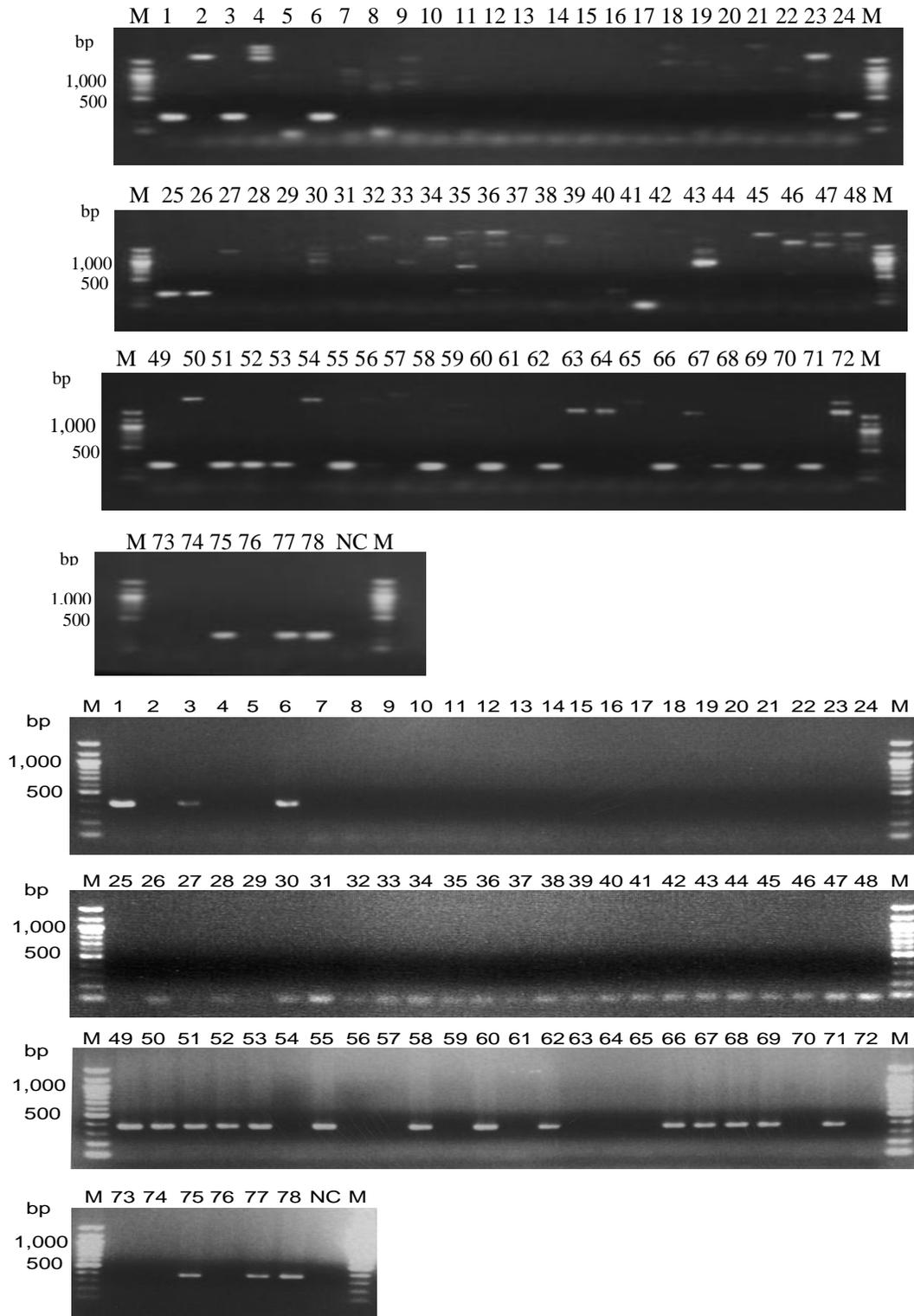
Fig. 13. PCR amplifications of *Ping*-ORF1 (upper picture, expected size= 264 bp) and *Ping*-ORF2 (lower picture, expected size= 369 bp). Lane numbers are same as in Fig. 11.
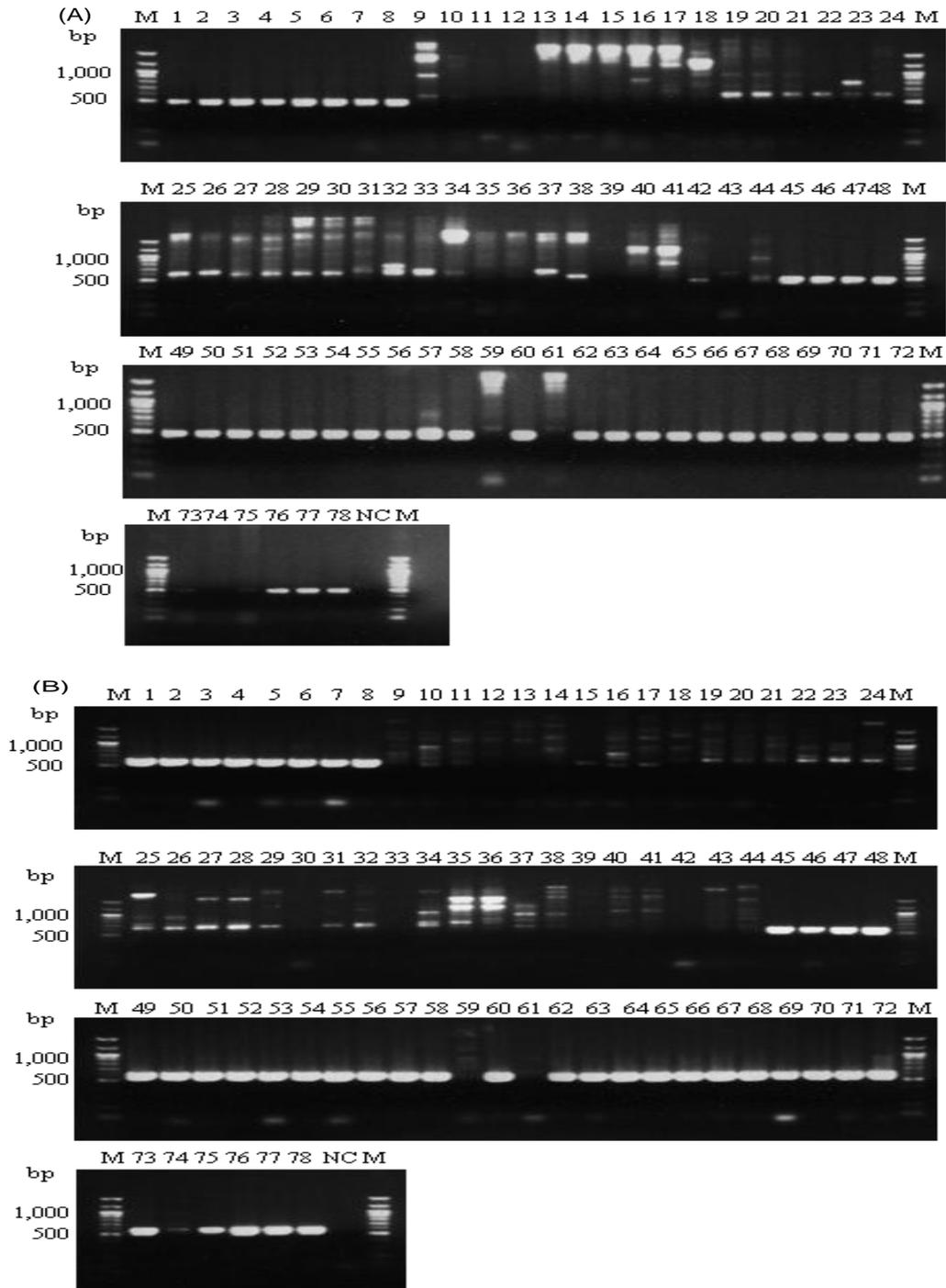
Fig. 14. PCR amplifications of *Pong*-ORF1 (A, expected size= 500 bp) and *Pong*-ORF2. (B, expected size= 586 bp). Lane numbers are same as that of Fig. 11. Expected sized amplicons were produced only in *O. sativa* and *O. rufipogon* accessions.

3.3.3 Insertional polymorphism and copy number of *mPing* in *O. rufipogon* genomes

TD was employed to elucidate the insertional polymorphisms and copy numbers of *mPing* in *O. rufipogon* genomes. The banding patterns of *mPing* varied considerably between the accessions, and no common band to all accessions was detected (Fig. 15). The copy numbers of *mPing* were estimated by counting all the bands amplified when using combinations of the *mPing*- and adapter- specific primers. The copy number of *mPing* in Nipponbare estimated by this method (50 copies) was consistent with the actual number of *mPing* according to the database of the complete genome sequence. Thus, the copy numbers of *mPing* in *O. rufipogon* were also accurately evaluated by this method. The copy numbers of *mPing* in *O. rufipogon* accessions ranged from zero to 30 (Table 5). Nine accessions had less than 10 copies of *mPing*, 20 accessions had 10-20 copies, and 8 accessions had 21-30 copies. The average copy numbers of *mPing* among all accessions, among accessions with *Ping*, and among accessions without *Ping* were 14.3, 14.9, and 13.7, respectively. According to Jiang et al. (2003), almost all *O. sativa* ssp. *japonica* varieties have ca. 50 copies. It is concluded, therefore, that the copy numbers of *mPing* in *O. rufipogon* is lower than that in *O. sativa* ssp. *japonica*.
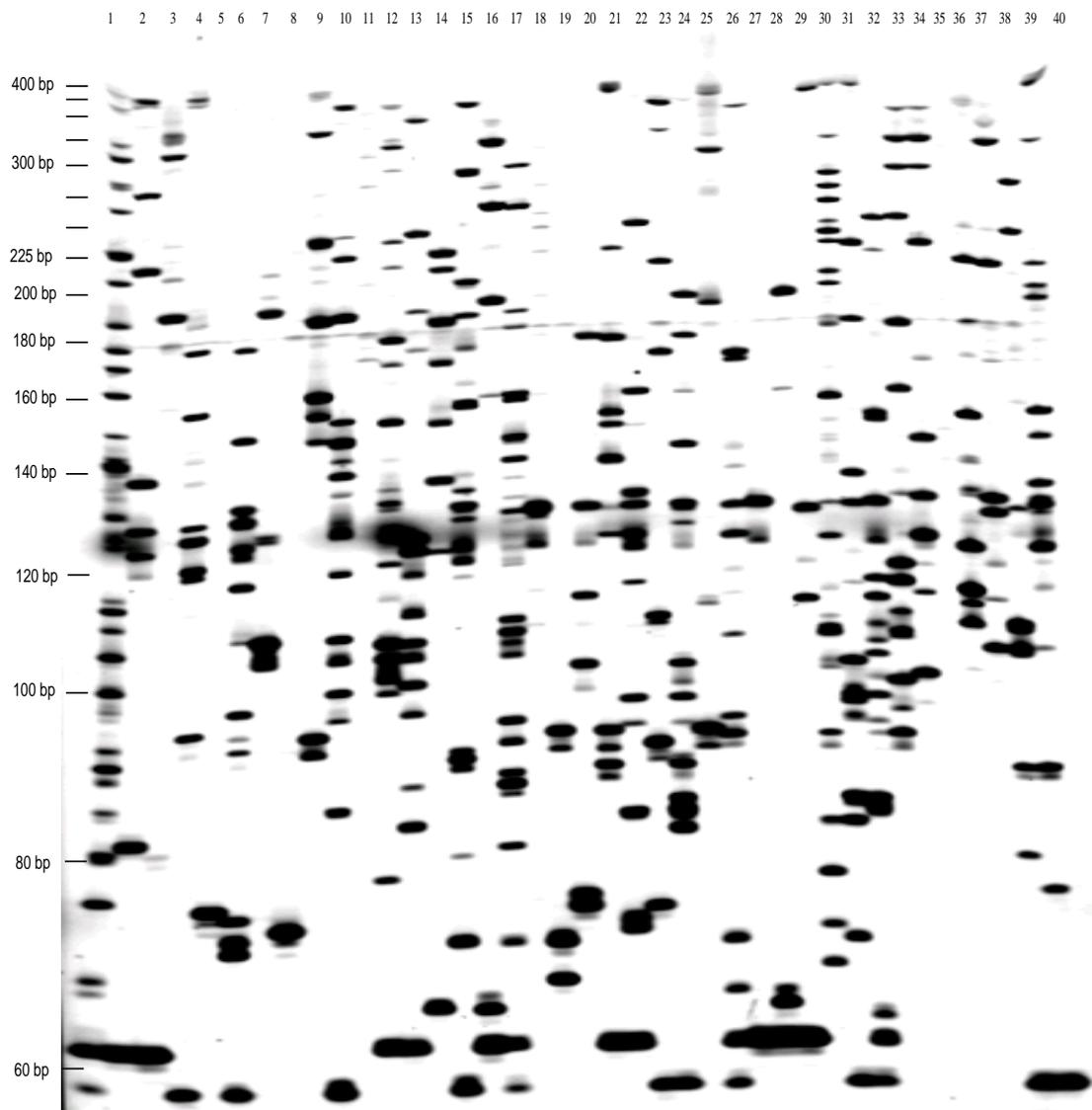
Fig. 15. Transposon display for *mPing* in *O. rufipogon* accessions. The migration of the DNA marker is indicated on the left in units of base pairs (bp). *Mse*I-digested DNA samples were ligated to an adapter and used as template for PCR with an *mPing* specific primer and an adapter primer. Lanes 1-2: *Oryza sativa*; lanes 3-39: *O. rufipogon*; and lane 40: negative control (distilled water). Both interspecific and intraspecific band polymorphisms are clearly visible.

Table 5. Distribution and copy number of *mPing* family transposons in *O. rufipogon* accessions

S.N.	Acc. no.	Copy number of *mPing*	Ping		Pong	
			ORF1	ORF2	ORF1	ORF2
1	Nipponbare	48	+	+	+	+
2	IR36	11			+	+
3	W0106	12	+	+	+	+
4	W0120	15			+	+
5	W1294	4			+	+
6	W1866	16	+	+	+	+
7	W1921	10			+	+
8	W2003	8			+	+
9	W0630	13			+	+
10	W1236	19			+	+
11	W1807	9			+	+
12	W1945	25			+	+
13	W2051	15	+	+	+	+
14	W2078	8		+	+	+
15	W2263	21	+	+	+	+
16	W0107	13	+	+	+	+
17	W0108	28	+	+	+	+
18	W0137	14			+	+
19	W0180	6	+	+	+	+
20	W0593	16			+	+
21	W0610	15			+	+
22	W1230	16	+	+	+	+
23	W1238	11	+	+	+	+
24	W1551	21	+	+	+	+
25	W1666	13			+	+
26	W1669	18			+	+
27	W1681	5			+	+
28	W1685	6	+	+	+	+
29	W1690	6		+	+	+
30	W1715	30	+	+	+	+
31	W1852	16	+	+	+	+
32	W1865	22			+	+
33	W1939	20	+	+	+	+
34	W1981	13			+	+
35	W2014	0			+	+
36	W2109	15	+	+	+	+
37	W2265	14			+	+
38	W2266	14	+	+	+	+
39	W2267	22	+	+	+	+

3.3.4 Distribution pattern of *Kiddo*

No species other than *O. sativa* and *O. rufipogon* harbored *mPing* family transposons. To determine whether *mPing* is the only MITE in genus *Oryza* to show this exclusive distribution pattern, it is necessary to investigate the distribution of other rice-specific MITEs. *Kiddo* is one of the rice-specific MITEs that have been identified in *O. sativa* ssp. *indica* variety (Yang et al. 2001). It is known that the copy number of *Kiddo* in the Nipponbare genome (42 copies), like that of *mPing* (50 copies), is lower than the copy numbers of the other rice MITEs (Takata et al. 2007). Moreover, many *Kiddo* elements, like *mPing* elements, were inserted in the gene regions (Yang et al. 2001) and showed insertional polymorphism between *japonica* and *indica* rices (Yang et al. 2005). For these reasons the distribution, copy number, and insertional polymorphisms of *Kiddo* were investigated in *O. rufipogon*, although no autonomous element for *Kiddo* has been identified so far.

When the *Kiddo*-specific primer pair was used, the PCR products of prospective size (269 bp) were amplified in all the AA genome species (Fig. 16). Moreover, *O. punctata* of BB and BBCC genome, *O. minuta* of BBCC genome, *O. eichingeri* and *O. officinalis* of CC genome also yielded the PCR products of 269 bp (Fig. 16). Sequencing analysis showed that these products had the same sequence that *Kiddo* has in the AA genome species. Consequently, it was confirmed that *Kiddo* is present not only in the AA genome but also in the BB, CC, and BBCC genome species.
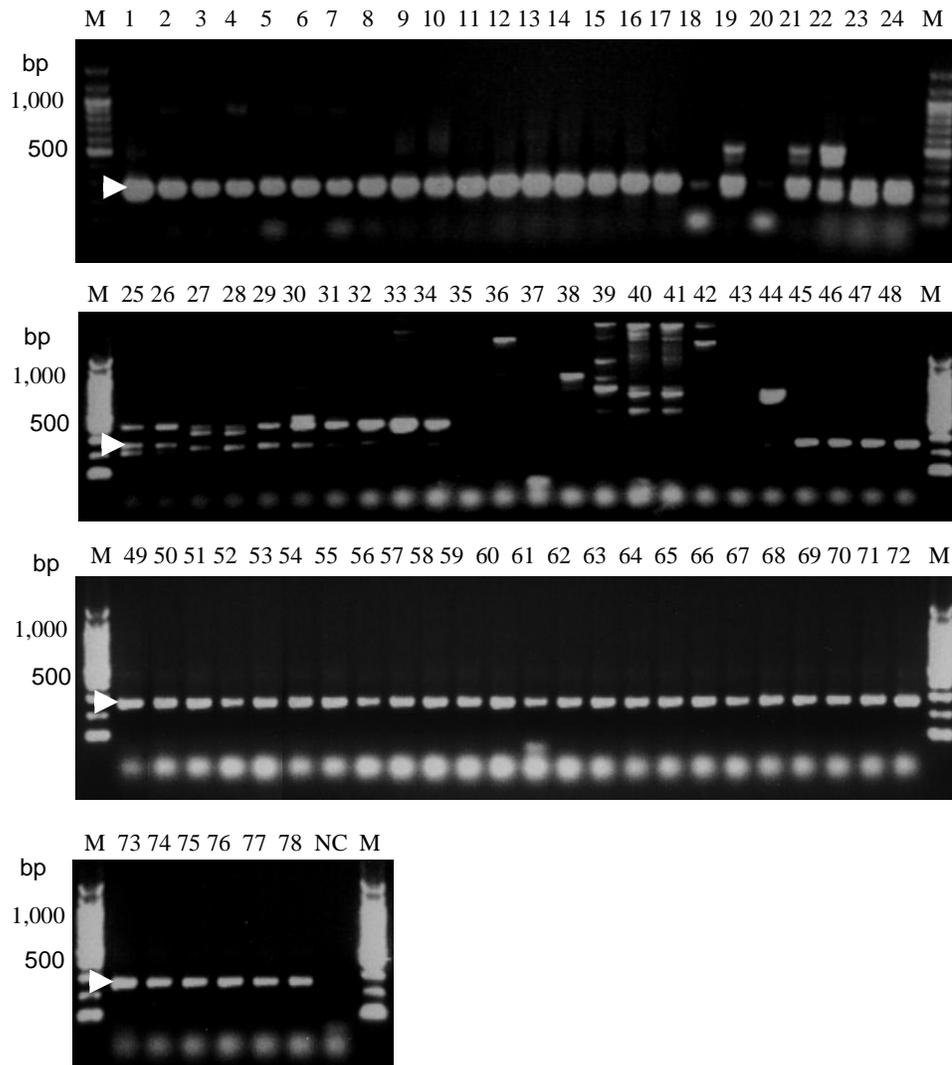
Fig. 16. Investigation of presence or absence of *Kiddo* in wild *Oryza* species. PCR was carried out using *Kiddo*-F and *Kiddo*-R primers (expected product size, 269 bp). Lane numbers are as same as that of Fig. 11. Lane M: DNA size marker (100 bp ladder, Nacalai, Japan); lane NC: negative control (distilled water). Arrowheads represent the bands of correct size.

To investigate the insertional polymorphisms and copy numbers of *Kiddo* in *O. rufipogon* genomes, TD was carried out using a combination of *Kiddo-* and adapter-specific primers. The copy number was estimated by same methods described above. One band was detected that was common to all the accessions, and many other bands were shared by several accessions (Fig. 17). This indicates that these accessions have *Kiddo* elements at the same chromosomal positions. The copy number of *Kiddo* ranged from five to 14 amid the *O. rufipogon* accessions, which is lower than that of *mPing*.
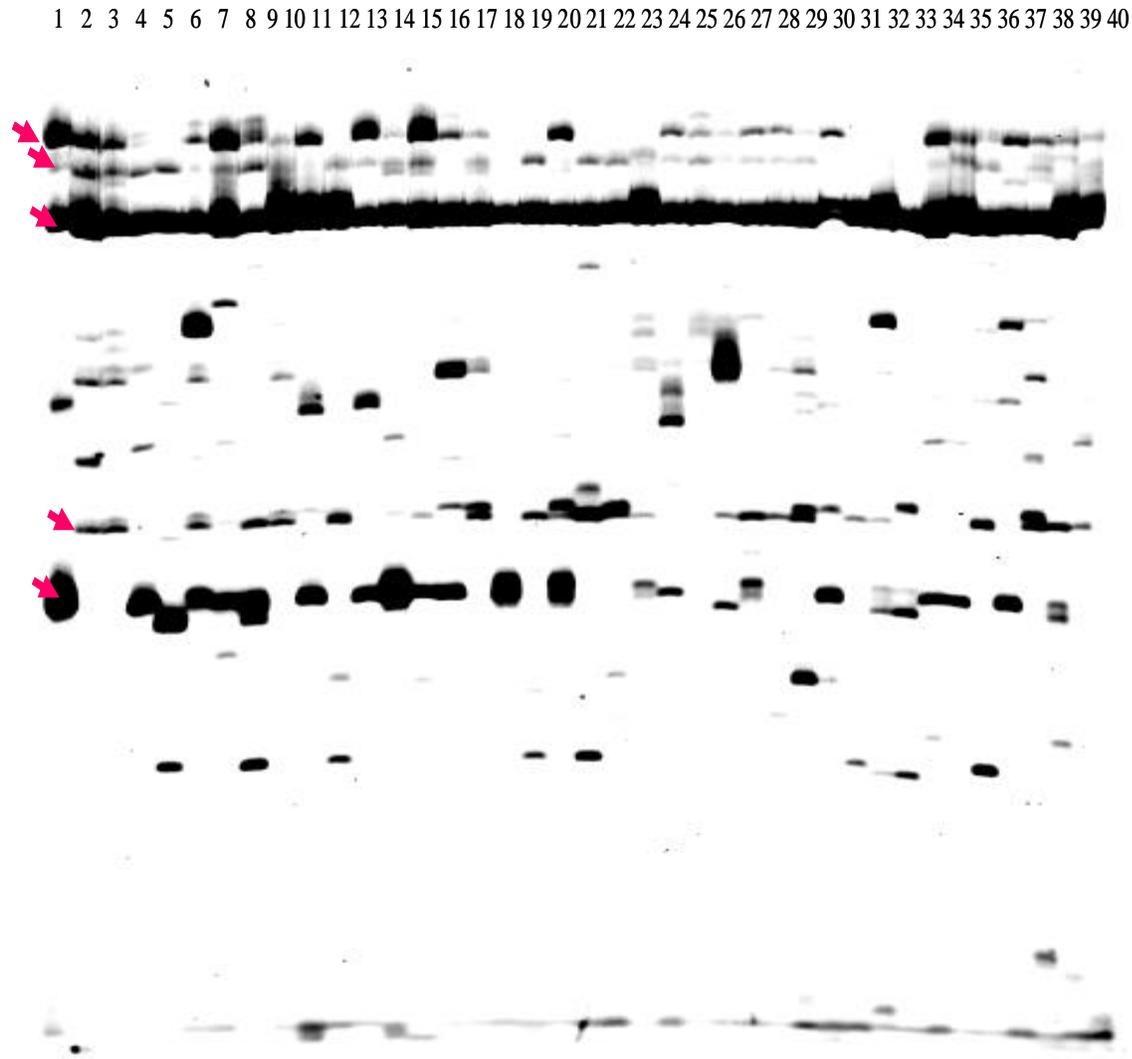
Fig. 17. Transposon display for *Kiddo* in *O. rufipogon* accessions. *Msp*I-digested DNA samples were ligated to an adapter and used as template for PCR with a *Kiddo* specific primer and an adapter primer. Lanes, 1-2: *Oryza sativa*; lanes, 3-39: *O. rufipogon*; lane, 40: negative control (distilled water). Arrows represent the bands common to all or many accessions.

3.3.5 Phylogenetic relationships between *O. rufipogon* accessions

To get further insights into the distribution and proliferation processes of the *mPing* family during the diversification of *O. rufipogon*, phylogenetic relationships among the *O. rufipogon* accessions was analyzed using DNA sequences of four genomic regions around *qSH1* (Konishi et al. 2006). Although *qSH1* is a major QTL responsible for the reduction of seed shattering that has been identified in the F_2 population of the cross between *indica* and *japonica* varieties, nucleotide polymorphisms around *qSH1* have proven to be suitable for studying the evolutionary history of Asian rice (Onishi et al. 2007). The polymorphisms in and around *qSH1* formed certain haplotypes that could distinguish *japonica* and *indica* subspecies. Haplotypes correspond to a set of closely related polymorphisms that are inherited as a unit. Therefore, a haplotype can be considered as a DNA signature of a paternal lineage. As such, haplotypes are often used to compare phylogenetic relationships of different populations derived from a common ancestor. For these reasons, haplotypes in and around *qSH1* region (Fig. 18) were thought to be suitable to study the phylogenetic relationships among various *O. rufipogon* accessions. PCR products yielded by each of the four primer pairs for each *O. rufipogon* accessions were sequenced.
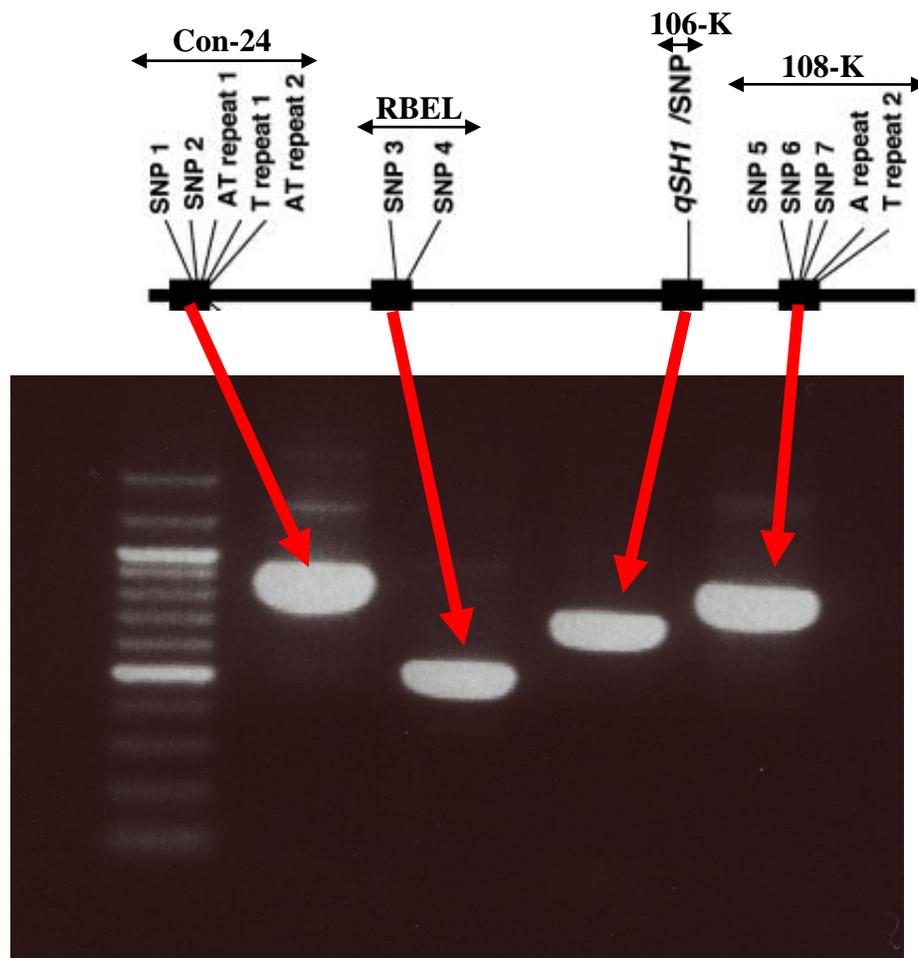
Fig. 18. Positions of primers used to amplify the *qSH1* region. The arrows point to the appropriate sized PCR products obtained from each primer sets. Genomic DNA of W0120 was used as the template.

Table 6. Haplotypes detected in the present study

Acc. no.																							
											Haplotypes in four regions around *qSH1*												
	SNP1	SNP2	AT repeat1	AT repeat2	SNP3	SNP4	*qSH1*/SNP	SNP5	SNP6	SNP7	A repeat 1	T repeat 1	A repeat 2	SNP8	SNP9	C repeat	SNP10	SNP11	SNP12	SNP13	A repeat 3	A repeat 4	SNP14
Nipponbare	G	T	AT2	AT4	C	A	T	A	A	A	A3	T8	A4	C	G	C9	C	T	A	G	A7	A7	T
Kasalath	A	C	AT3	AT3	T	G	G	T	G	G	A2	T9	A8	C	A	C6	C	G	G	C	A6	A7	C
IR36	G	C	AT3	AT3	C	G	G	T	G	G	A2	T8	A4	C	A	C6	C	G	G	C	A6	A7	T
W0106	A	C	AT3	AT3	T	G	G	A	G	G	A3	T8	A4	C	A	C6	C	G	G	C	A6	A7	T
W0120	G	C	AT3	AT3	C	G	G	A	G	G	A3	T8	A4	T	G	C6	C	G	G	C	A6	A7	T
W1294	G	C	AT3	AT3	C	G	G	A	G	G	A3	T8	A4	C	G	C9	C	G	G	C	A6	A7	T
W1866	A	C	AT2	AT3	T	G	G	T	G	G	A2	T9	A4	C	A	C6	C	G	G	C	A6	A6	C
W1921	G	C	AT3	AT3	T	G	G	A	G	G	A3	T9	A4	C	G	C6	C	G	G	C	A6	A7	T
W2003	G	C	AT3	AT3	C	G	G	A	G	G	A3	T9	A4	C	G	C5	T	G	G	C	A6	A7	T
W0630	A	C	AT3	AT3	T	G	G	T	G	G	A2	T8	A4	C	A	C6	C	G	G	C	A6	A7	T
W1236	G	C	AT3	AT3	C	G	G	A	G	A	A3	T8	A4	T	G	C10	C	G	G	C	A6	A7	T
W1807	G	C	AT3	AT3	C	G	G	T	G	G	A3	T8	A4	C	G	C6	C	G	G	C	A6	A6	C
W1945	G	C	AT3	AT3	C	G	G	A	G	G	A3	T8	A4	C	G	C6	C	G	G	C	A6	A6	C
W2051	G	C	AT3	AT3	C	G	G	A	G	G	A3	T9	A4	C	G	C6	C	G	G	C	A6	A6	C
W2078	G	C	AT3	AT3	T	G	G	A	G	G	A3	T9	A4	C	G	C6	C	G	G	C	A6	A6	C
W2263	A	C	AT3	AT3	T	G	G	T	G	G	A2	T9	A8	C	A	C6	C	G	G	C	A6	A7	T
W0107	A	C	AT3	AT3	T	G	G	A	G	G	A3	T8	A8	C	A	C6	C	G	G	C	A6	A7	T
W0108	G	C	AT3	AT3	C	G	G	A	G	G	A3	T8	A4	T	G	C6	C	G	G	C	A6	A7	T
W0137	G	C	AT3	AT3	C	G	G	A	G	G	A3	T8	A4	C	G	C8	C	G	G	C	A6	A7	T
W0180	G	C	AT3	AT3	C	G	G	A	G	G	A3	T8	A4	T	G	C6	C	G	G	C	A6	A7	T
W0593	G	C	AT3	AT3	C	G	G	A	G	G	A3	T7	A4	C	G	C8	C	G	G	C	A6	A7	T
W0610	A	C	AT3	AT3	T	G	G	A	G	G	A3	T7	A4	C	A	C8	C	G	G	C	A6	A7	T
W1230	A	C	AT2	AT3	T	G	G	A	G	G	A3	T8	A4	C	A	C9	C	G	G	C	A6	A7	T
W1238	G	C	AT3	AT3	C	G	G	T	G	G	A3	T8	A4	C	G	C6	C	G	G	C	A6	A7	T
W1551	A	C	AT2	AT3	T	G	G	A	G	G	A3	T8	A4	C	A	C6	C	G	G	C	A6	A7	T
W1666	A	C	AT3	AT3	T	G	G	A	A	A	A3	T8	A4	C	A	C9	C	T	A	G	A7	A7	T
W1669	G	C	AT3	AT3	C	G	G	A	A	A	A3	T8	A4	T	G	C9	C	T	A	G	A7	A7	T
W1681	A	C	AT3	AT3	T	G	G	A	G	A	A3	T8	A4	C	A	C9	C	G	G	G	A6	A7	T
W1685	A	C	AT3	AT3	T	G	G	A	G	G	A3	T8	A4	C	A	C5	T	G	G	C	A6	A7	T
W1690	G	C	AT3	AT3	C	G	G	A	G	G	A3	T8	A8	C	G	C5	T	G	G	C	A6	A7	T
W1715	G	C	AT3	AT3	C	G	G	T	G	G	A3	T9	A4	C	G	C6	C	G	G	C	A6	A7	T
W1852	A	C	AT2	AT3	C	G	G	T	G	G	A2	T8	A8	C	A	C6	C	G	G	C	A6	A7	T
W1865	A	C	AT2	AT3	T	G	G	T	G	G	A2	T9	A4	C	A	C6	C	G	G	C	A6	A7	T
W1939	G	C	AT3	AT3	C	G	G	T	G	G	A2	T8	A4	C	G	C6	C	G	G	C	A6	A6	C
W1981	G	C	AT3	AT3	C	G	G	T	G	G	A2	T9	A4	C	G	C6	C	G	G	C	A6	A6	C
W2014	G	C	AT3	AT3	C	G	G	A	G	G	A3	T9	A4	C	G	C9	C	G	G	C	A6	A6	C
W2109	G	C	AT3	AT3	C	G	G	A	A	A	A3	T8	A4	C	G	C9	C	T	A	G	A7	A7	T
W2265	A	C	AT3	AT3	T	G	G	T	G	G	A2	T8	A8	C	A	C6	C	G	G	C	A6	A7	T
W2266	G	C	AT3	AT3	C	G	G	A	G	G	A3	T9	A4	C	G	C5	T	G	G	C	A6	A7	T
W2267	G	C	AT3	AT3	C	G	G	A	G	G	A2	T9	A8	C	G	C10	C	G	G	C	A6	A7	T
CG14	G	C	AT3	AT3	C	G	G	A	G	G	A2	T8	A4	C	G	C10	C	G	G		A6	A7	T
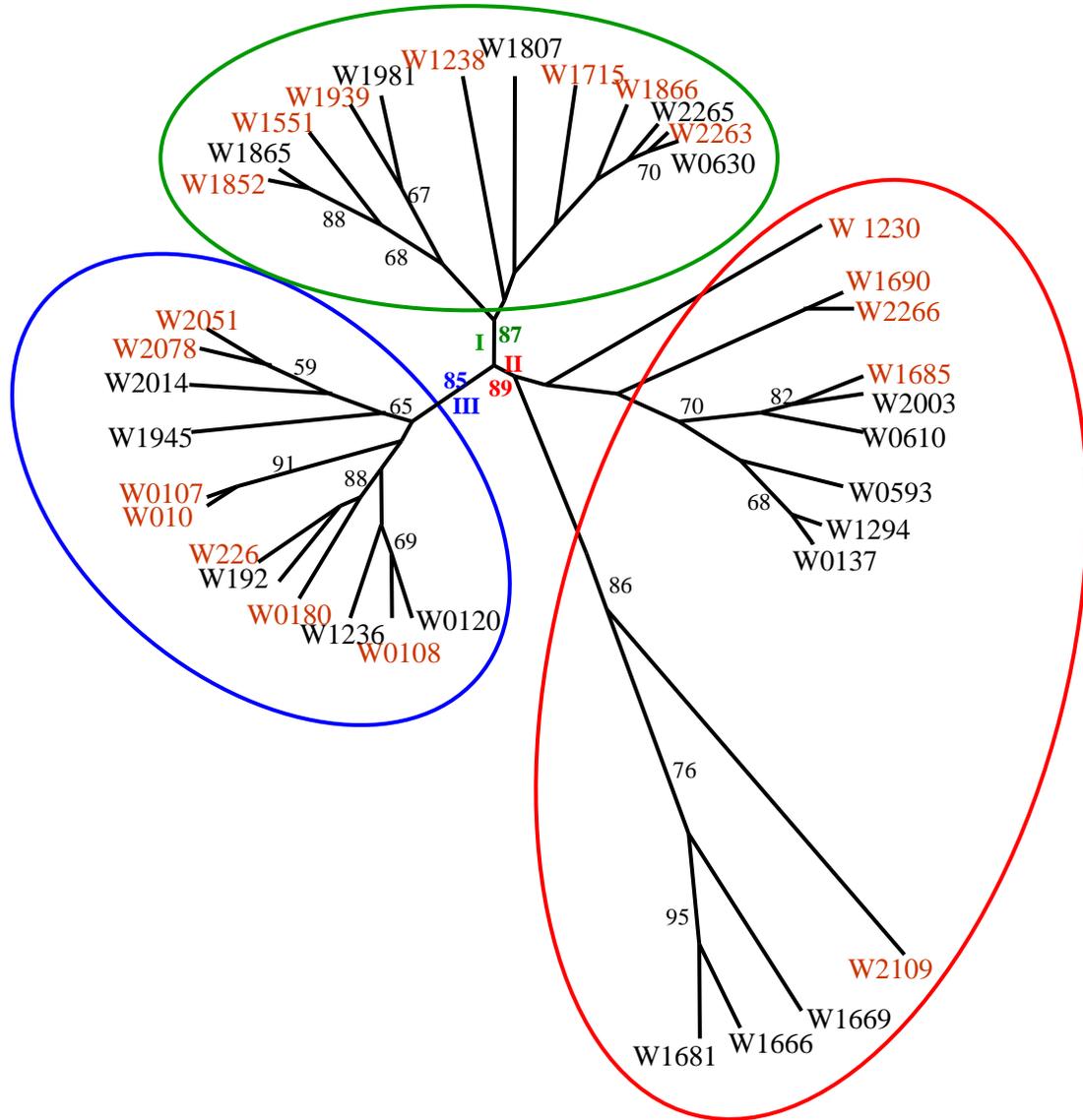
Fig. 19. Neighbor-joining phylogenetic tree of *O. rufipogon* accessions generated from the sequence data around *qSH1* region. Numbers above the branches represent the bootstrap values that are more than 50 %. Letters I, II, and III represent the three major clusters. Red characters indicate the accessions harboring *Ping*.

3.4 Discussion

The present study shows that the *mPing* elements are deployed only in *O. rufipogon*, and are not deployed in other 18 other wild *Oryza* species surveyed. *Kiddo*, in contrast, is deployed not only in AA genome species but also in BB, CC, and BBCC genome species. It has been reported that the *Tourist C* element, another MITE found in rice, is present in all the four *Oryza* species complexes (Iwamoto et al. 1999). Moreover, *Pangrangja* elements were found to be present in AA, BB, CC, BBCC, CCDD, and EE genome species, and to be applicable for studying the relationships among species in genus *Oryza* (Park et al. 2003). These observations suggest that *mPing* originated later than the other MITE elements during the evolutionary history of genus *Oryza*: in other words, *mPing* is a comparatively young MITE.

Pong is present in all of the 37 *O. rufipogon* accessions used in this study, but the presence of *Ping* was detected in only 17 of these plant materials. The phylogenetic tree showed that the accessions with *Ping* were not grouped together in the same cluster; in fact, many of them were closely related to accessions without *Ping*. It is therefore proposed that the *mPing* family had already evolved and was present in the common ancestor of the *O. rufipogon* varieties, and later on, *Ping* was eliminated from the genomes of some of the accessions by random drift and/or natural selection along with the advancement of diversification in *O. rufipogon*. In *O. sativa*, *Ping* is present in only temperate *japonica* varieties, and the copy number of *mPing* in temperate *japonica* varieties was higher than that in tropical *japonica* and *indica* varieties (Jiang et al. 2003; Kikuchi et al. 2003). *Ping* has been believed to play a dominant role in the proliferation of *mPing* in *O. sativa*. In this study, however, the average copy number of *mPing* in *O.*

rufipogon was not associated with the presence or absence of *Ping*. This indicates that *Pong*, as well as *Ping*, is responsible for the mobilization and proliferation of *mPing* in *O. rufipogon*.

O. rufipogon has been considered to be the direct progenitor of the cultivated rice species, *O. sativa* ssp. *japonica* and *indica* (Xu et al. 2007). The copy number of *mPing* in *O. sativa* ssp. *japonica* was estimated by database searches to be 60-80 copies (it is actually 50 copies) (Kikuchi et al. 2003), whereas that of *O. sativa* ssp. *indica* was estimated to be less than 10 copies (Jiang et al. 2003). Accordingly, it was suggested that the first amplification of *mPing* might have occurred during the domestication of *O. sativa*, and that, subsequently, *mPing* was activated again in *japonica* varieties, where it proliferated from 10 to 50 copies (Jiang et al. 2003). The copy numbers of *mPing* were markedly different even among *O. rufipogon* accessions. In the phylogenetic analysis, the accessions with different copy numbers of *mPing* were randomly distributed in different phylogenetic clusters. There was no direct correlation between the copy numbers of *mPing* and the geographical distribution of *O. rufipogon* accessions. Hence, it was concluded that the amplification of *mPing* occurred independently and randomly in each *O. rufipogon* accessions.

In *mPing*-TD analysis, *O. rufipogon* accessions yielded only polymorphic bands. Moreover, the copy number of *mPing* varied considerably among the accessions. Estimates of the outcrossing rates of Asian wild rice species have ranged from 5 to 60 %, though cultivated rice species are predominantly selfing (Morishima and Barbier 1990). This suggests that *O. rufipogon* strains easily outcross with each other, consequently the insertional polymorphisms of *mPing* may occur. Yet outcrossing cannot fully explain the

differences in insertional polymorphisms and copy numbers of *mPing* among *O. rufipogon* accessions, because these accessions have several *Kiddo* elements at the same chromosomal positions. The *mPing* is still actively transposing in some of *japonica* cultivars under natural conditions (Naito et al. 2006). It is, therefore, considered that *mPing* is currently active, even within *O. rufipogon* strains.

Chapter 4

Mobilization of *mPing* and *Pong* transposons by interspecific cross between *Oryza sativa* and *O. glaberrima*

4.1 Introduction

Rice is one of the most important food crops worldwide, providing a major source of sustenance to over half the world's population. It has two cultivated species: *Oryza glaberrima* Steud., which is cultivated only in West Africa (Ghesquiere 1997), and *O. sativa* L., which has been domesticated in Asia and then spread in a large parts of the world (Linares 2002). Compared with *O. sativa*, *O. glaberrima* has agronomically important traits such as weed competitiveness, drought tolerance and adaptability to low fertilizer conditions (Jones et al. 1997). Furthermore, it is resistant to pests and diseases including African rice gall midge, rice yellow mottle virus and blast disease. If these traits can be introduced into *O. sativa* cultivars, the production yield in severe cultivation conditions, such as rainfed hilly areas, deep-water floating areas and coastal mangrove areas, may be considerably improved (Sarla and Swamy, 2005). One of the successful examples is the development of NERICA (New Rice for Africa) lines by inter-specific cross between *O. glaberrima* and *O. sativa*. NERICA lines expressed high levels of heterosis or hybrid vigor combining the best attributes of both the parent species, thereby they showed the high production yield (6 t/ha in the best conditions), drought- and pest-

resistance, and short maturing period (Semagn et al. 2007). However, genes responsible for these traits are not identified yet.

Transposable elements (TEs) which were first discovered in maize by Barbara McClintock more than a half century ago, have been used as a powerful tool for gene isolation, since the transposition and insertion of TEs cause the alternation of gene structure and expression. Nowadays, transposon tagging using active TEs has become one of the efficient gene tagging systems in rice (Jeon and An 2001). After *bronze* gene (*Bz1*) was isolated from maize by using active *Ac/Dc* transposons (Fedoroff et al. 1984), transposon tagging using *Ac/Ds* and *Tos17* has been applied to rice and successfully identified many genes (Hirochika et al. 1996). However, the transformation and tissue culture procedures are necessary for the activation of these elements. Therefore, the important feature for transposon tagging is the transposition of TEs under natural conditions.

Miniature-Ping (*mPing*) is an active miniature inverted-repeat transposable element (MITE) isolated from rice genome (Jiang et al. 2003; Kikuchi et al. 2003; Nakazaki et al. 2003). Because *mPing* is a non-autonomous element without an open reading frame (ORF), its transposition depends on a transposase coded by autonomous elements *Ping* and/or *Pong* (Jiang et al. 2003; Kikuchi et al. 2003; Yang et al. 2007). The transposition of *mPing* was detected not only in cultured cells (Kikuchi et al. 2003; Jiang et al. 2003) but also in intact plants growing in a paddy field (Nakazaki et al. 2003; Naito et al. 2006). As well as other MITEs, *mPing* was preferentially found in single-copy regions and near transcribed DNA (Naito et al. 2006). Furthermore, mutant genes caused by *mPing*

insertion have been identified (Teraishi et al. 1999; Nakazaki et al. 2003). These suggest that *mPing* is a suitable transposon for gene tagging in *Oryza* species.

4.2 Materials and methods

4.2.1 Plant materials

Twenty two NERICA lines, a recurrent parent strain *Oryza sativa* (WAB450-104), and a donor parent strain *O. glaberrima* (CG14) were used. The accession numbers of each of NERICA lines and their parent strains used for the study are given in Table 7. Nipponbare (*O. sativa* ssp. *japonica*) was used in all experiments as a positive control.

Table 7. List of the plant materials used in this study

S.N.	Pedigree	Name[*]	Introgression[*] (%)
1	Nipponbare		
2	CG14 (*Oryza glaberrima*)	Donor parent	
3	WAB450-104 (*O. sativa*)	Recurrent parent	
4	WAB450-11-1-P26-2-HB	ND	ND
5	WAB450-24-3-1-P37-HB	ND	ND
6	WAB450-24-3-4-P18-3-1	20	7.0
7	WAB450-24-3-P3-1-HB	27	4.5
8	WAB450-25-2-9-1-1	ND	ND
9	WAB450-5-1-BL1-DV6	23	7.2
10	WAB450-I-B-P-129-HB	45	6.4
11	WAB450-I-B-P-133-HB	21	8.1
12	WAB450-I-B-P-135-HB	7	5.3
13	WAB450-I-B-P-157-1-1	49	2.9
14	WAB450-I-B-P-157-2-1	24	6.6
15	WAB450-I-B-P-163-2-1	22	5.3
16	WAB450-I-B-P-163-4-1	ND	ND
17	WAB450-I-B-P-20-HB	NERICA 7	7.3
18	WAB450-I-B-P-23-HB	25	5.1
19	WAB450-I-B-P-24-HB	26	5.5
20	WAB450-I-B-P-28-HB	NERICA 3	3.4
21	WAB450-I-B-P-33-HB	12	6.3
22	WAB450-I-B-P-1-1	ND	ND
23	WAB450-I-B-P-82-2-1	46	7.3
24	WAB450-I-B-P-38-HB	NERICA 1	6.7
25	WAB450-11-1-P31-1-HB	NERICA 5	11.0

*, Refer to Semagn et al. (2007).
ND, not determined

4.2.2 Genomic DNA extraction and quantification

Genomic DNA was extracted from the young leaves (3 weeks seedlings) by the modified Cetyltrimethylammonium bromide (CTAB) method of Murray and Thompson (1980). DNA was quantified by using spectrophotometer (Bio Photometer, Eppendorf, Hamburg, Germany).

4.2.3 PCR amplification

To determine whether *mPing*, *Ping*, and *Pong* were present in NELICA lines and their parents, PCR amplification using the element specific primers was carried out. *mPing*-specific primers were designed in the internal region (*mPing*-F: 5′-GGGATGAGAGAGAAGGAAAGAG-3′ and *mPing*-R: 5′-AACAATCCCCACAGTGGAG-3′). For *Ping* and *Pong*, we used the following specific primer pairs designed in to complement *Ping*-ORF1, *Ping*-ORF2, *Pong*-ORF1, and *Pong*-ORF2, respectively: *Ping*-ORF1-F1 (5′-ACAAGCGGATACTCCGAC-3′) and *Ping*-OFR1-R1 (5′-GAGGACAATGCCTTCCATAAC-3′); *Ping*-ORF2-F1 (5′-AAGGTATTTGCTAAGTCATAT-3′) and *Ping*-ORF2-R1 (5′-ATGAGCCAAACGATCTCTCATACTA-3′); *Pong*-ORF1-F1 (5′-TCTACTCCACCACCAACACC-3′) and *Pong*-ORF1-R1 (5′-TCAGCCTTGTTTTTGTCCTTC-3′), and *Pong*-ORF2-F1 (5′-CTCATACGAAGACCTCCTCC-3′) and *Pong*-ORF2-R1 (5′-CGCCTAAGATACCTCTCACC-3′). The genomic DNA was used as the template for PCR using Ex-*Taq* DNA polymerase (TaKaRa, Shiga, Japan). The PCR conditions consisted of pre-denaturation for 3 min at 94 ℃ followed by 30 cycles of polymerization

reaction, each consisting of a denaturation step for 10 s at 98 ℃, an annealing step for 45 s at 60 ℃, and an extension step for 45 s at 72 ℃, with a final extension step for 7 min at 72 ℃. PCR products were visualized under UV light following electrophoresis on 1.5 % agarose gel in 0.5x TBE buffer.

4.2.4 DNA blot analysis

Five micro gram of genomic DNA of each sample was digested with restriction endonuclease, *Eco*RI for 12 h at 37 ℃. Digested DNA was separated by 1% agarose gel in 1X TAE buffer and then transferred onto Hybond N$^+$ nylon membrane (GE Healthcare, Buckinghamshire, UK) by alkaline transfer method. Probes for *mPing*, *Ping*, and *Pong* were labeled with digoxygenin (DIG) by PCR DIG Probe Synthesis Kit (Roche Diagnostics, Tokyo, Japan) using specific primers (5′-TGGAGGGGTTTCTCTTTGACG-3′ and 5′-CAGTGTATGATTTTCCCTTTCC-3′ for *mPing*, 5′-GAGAGATGGGTGAAAAGGCACCC-3′ and 5′-GTCAATTTACACATATTGCATAGC-3′ for *Ping*, 5′-GTCAGTACTTACTATGTCAGGATC-3′ and 5′-TAGCTGCAGCTTGGGGTGTTGG-3′ for *Pong*). Hybridization signals were detected by CDP-Star detection module (Roche Diagnostics).

4.2.5 Transposon display

The copy numbers of *mPing* were determined by transposon display (TD) according to protocol of Casa et al. (2000). The genomic DNA of each plant material were digested

with *Mse*I and ligated to an adapter. The primary PCR was carried out using an adaptor-specific primer (MseI+0: 5'-GACGATGAGTCCTGAGTAA-3') and an *mPing*-specific primer (*mPing*-P1: 5′-AATGTGCATGACACACCAG-3′). The secondary PCR was carried out using an adaptor primer and another *mPing*-specific primer (*mPing*-P2: 5′-CAGTGAAACCCCCATTGTGAC-3′) labeled with TAMRA fluorescence reagent. The primary PCR conditions were as follows: pre-denaturation for 2 min at 72 °C and subsequent pre-denaturation for 3 min at 94 °C followed by 30 cycles of polymerization reaction, each consisting of a denaturation step for 30 s at 95 °C, an annealing step for 45 s at 58 °C and an extension step for 45 s at 72 °C, with a final extension step for 7 min at 72 °C. PCR products were diluted between 20- and 50-fold with 1/10 TE buffer and used as templates for secondary PCR (selective amplification). Selective amplification cycling started from pre-denaturation for 3 min at 94 °C followed by 7 cycles of touch down PCR. The first touch down cycle consisted of a denaturation step for 30 s at 95 °C, an annealing step for 45 s at 64 °C, and an extension step for 45 s at 72 °C, and from the second cycle, temperature at the annealing step was lowered by 1.0 °C per cycle until it reached to 58 °C. The touch down PCR was followed by 30 cycles of polymerization reaction, each consisting of a denaturation step for 30 s at 95 °C, an annealing step for 45 s at 58 °C and an extension step for 45 s at 72 °C, with a final extension step for 7 min at 72 °C. The copy number of *mPing* was estimated by counting all the bands amplified when using combinations of the *mPing*- and adapter-specific primers.

4.2.6 Investigation of *de novo* insertions of *mPing*

Putative *de novo* bands were excised from TD gel, and reamplified by PCR using an adaptor primer and *mPing*-P2 primers. Reamplified DNA fragments were sequenced using a CEQ8000 automated DNA sequencer (Beckman-Coulter, Fullerton, CA, USA), and *mPing* sequence was eliminated from the resulting sequences to obtain only *mPing*-flanking sequence. The chromosomal position of *de novo* insertion was determined by BLAST search in RAP-DB (Rice Annotation Project Database, http://rapdb.dna.affrc.go.jp/) using the *mPing*-flanking sequences as queries. Primers were designed at both of the *mPing*-flanking sequences, and PCR was carried to confirm the presence of *mPing*.

4.3 Results

4.3.1 Existence of *mPing* family transposons in NERICA lines and their parents

As a first step to apply *mPing* for efficient gene tagging in NERICA lines, the existence of *mPing*, *Ping*, and *Pong* in NERICA lines and their parent strains was investigated by PCR analysis. In PCR analysis using *mPing*-specific primer pair, the recurrent parent strain and all NERICA lines excepting WAB450-I-B-P-24-HB yielded PCR products of appropriate size (approximately 430 bp) (Fig. 21). However, no PCR product of expected size was obtained from the donor parent. When PCR was carried out using *Ping*- and *Pong*-specific primer pairs, no strains and parents yielded PCR products of *Ping*-ORF1 (264 bp) and *Ping*-ORF2 (369 bp) (Fig. 22). Although three other primer pairs designed at different position of *Ping* were used for PCR analysis, same results

were obtained from all the primer pairs (data not shown). *Ping* fragments of the expected size were successfully obtained from Nipponbare that is known to have at least one copy of the element (Jiang et al. 2003), indicating that *Ping* is absent in NERICA lines and their parents. On the other hand, the recurrent parent and all NERICA lines yielded PCR products of *Pong*-ORF1 (500 bp) and *Pong*-ORF2 (586 bp) (Fig. 23). This, together with the result that PCR product of *mPing* was obtained from NERICA lines and the recurrent parent, indicates that NERICA lines have *mPing* and *Pong* derived from the recurrent parent strain.
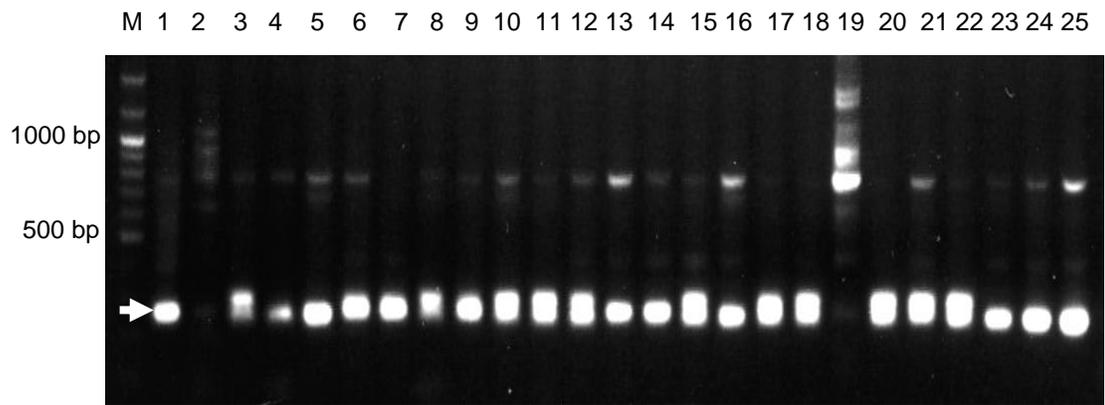
Fig. 20. Determination of presence or absence of *mPing* in NERICA lines. M: DNA size marker (100-bp ladder, Nacalai, Japan). Lane numbers represent the serial number of the plant materials as indicated in Table 7. Arrowhead shows the PCR amplicon of appropriate sizes (ca. 200-bp).
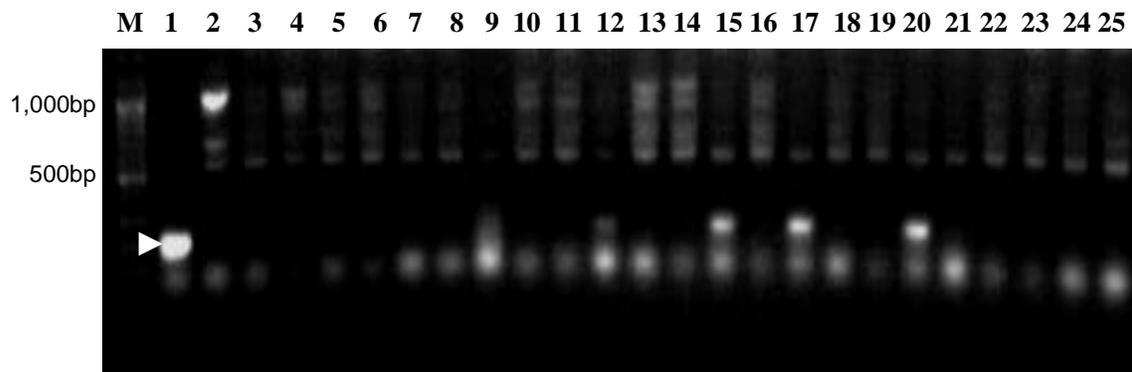
Fig. 21. Determination of the presence or absence of *Ping* in NERICA lines using *Ping* specific primer. M: DNA size marker (100-bp ladder, Nacalai, Japan). Lane numbers represent the serial number of the plant materials indicated in table 7. Arrowhead shows the expected sized PCR amplicon (ca. 205-bp).
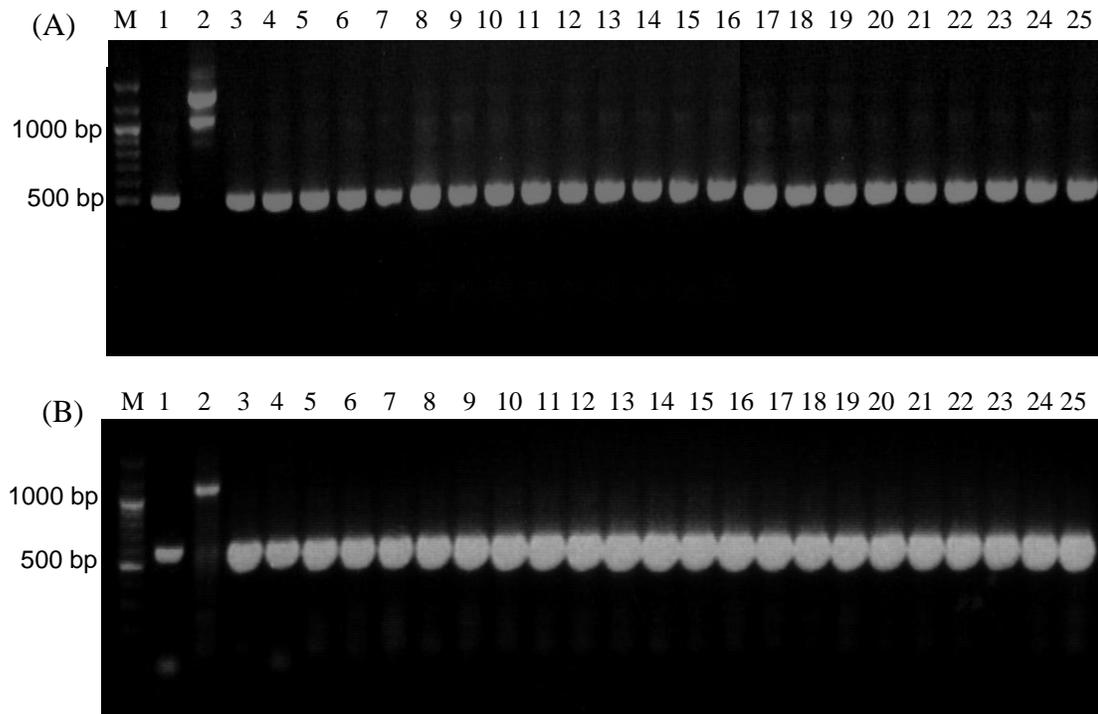
Fig. 22. PCR amplifications of *Pong*-ORF1 (A, expected size = 500-bp) and *Pong*-ORF2. (B, expected product size = 586-bp). M: DNA size marker (100-bp ladder, Nacalai, Japan). Lane numbers represent the serial numbers of the plant materials as listed in Table 7. All the NERICA lines and *O. sativa* parent produced only the specific PCR products, but *O. glaberrima* produced bands longer than the appropriate size and were considered to be PCR artifacts.

4.3.2 Copy number and possible mobilization of *mPing* and *Pong* in NERICA lines

TD and DNA blot analyses were employed to investigate the copy numbers and possible mobilization of *mPing* and *Pong*, respectively. TD analysis revealed that the recurrent parent had eight copies of *mPing* in the genome, while the copy numbers of *mPing* in NERICA lines ranged from two to 12 (Fig. 24, Fig. 25, Table 8). WAB450-24-3-4-P18-3-1 and WAB450-24-3-P3-1-HB, WAB450-I-B-P-157-1-1 and WAB450-I-B-P-157-2-1, and WAB450-I-B-P-163-2-1 and WAB450-I-B-P-163-4-1 showed same banding patterns between each other, but all lines showed marked difference in banding patterns from each other and that of their parents. The donor parent produced one band, but PCR analysis using *mPing*-specific primer pair showed the absence of *mPing* in the donor parent. Thus, we concluded that the band shared with the donor parent was a non-specific product, and the changed banding patterns of *mPing* in NERICA lines relative to the recurrent parent was not due to transfer from the donor parent. Although no band common to all lines was detected, all lines excepting WAB450-I-B-P-24-HB produced at least one band with the same migration as that produced by the recurrent parent. Moreover, all lines excepting WAB450-I-B-P-82-2-1 produced *de novo* bands that are not in the recurrent parent. These indicate that at least one of *mPing* elements is inherited from the recurrent parent to NERICA lines, but some of *mPing* elements may transpose from the original position to another position in these lines.
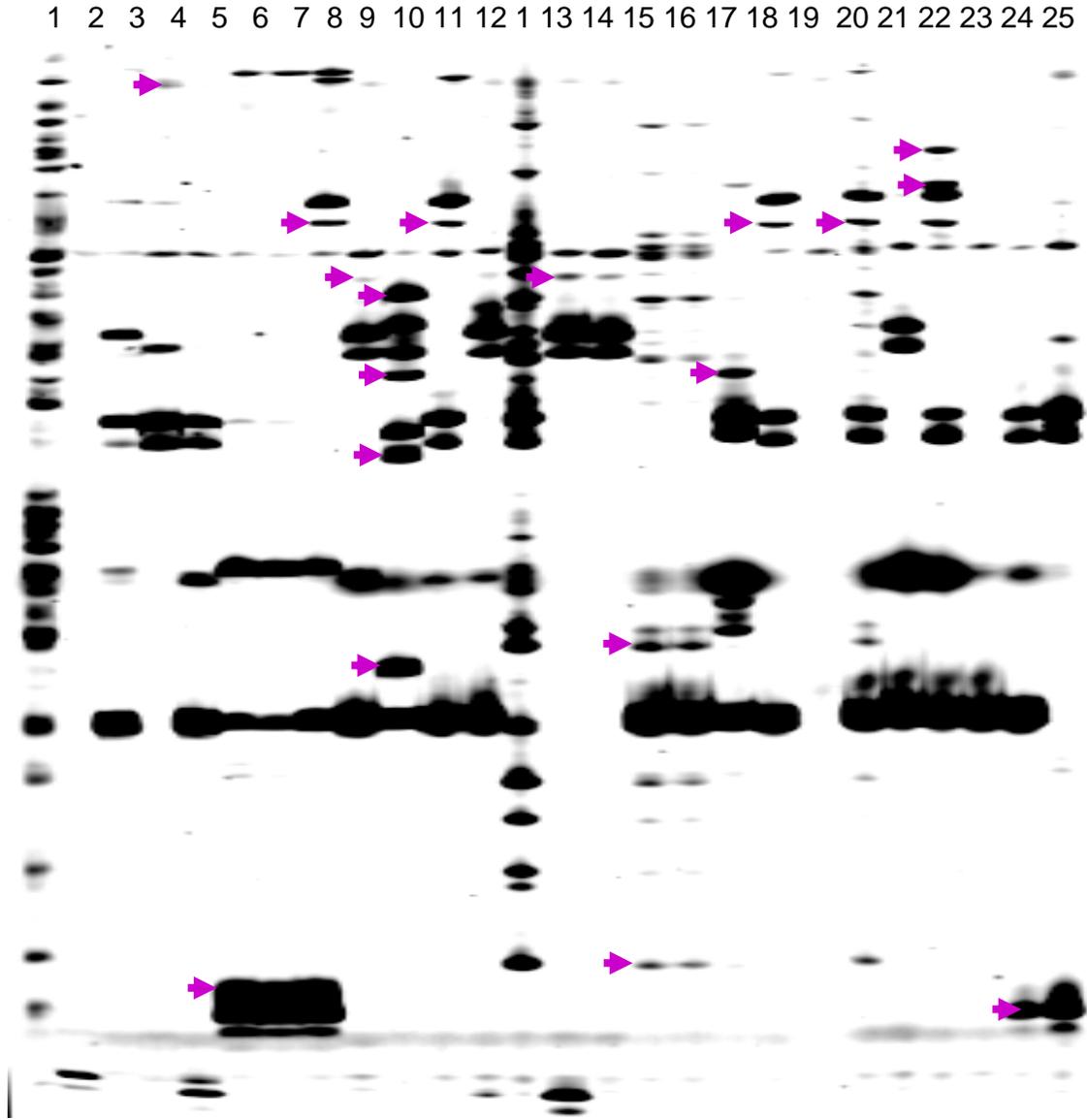
Fig. 23. Transposon display for *mPing* in NERICA lines. *Mse*I-digested DNA samples were ligated to an adapter and used as template for PCR with an *mPing* specific primer and an adapter primer. Lanes are- 1: *Oryza sativa*; 2: *O.glaberrima*; 3: O. sativa parent strain, 4-25: NERICA lines. The lane numbers represent the serial numbers of the accessions as listed in Table 7. Arrows indicate the new *mPing* bands.
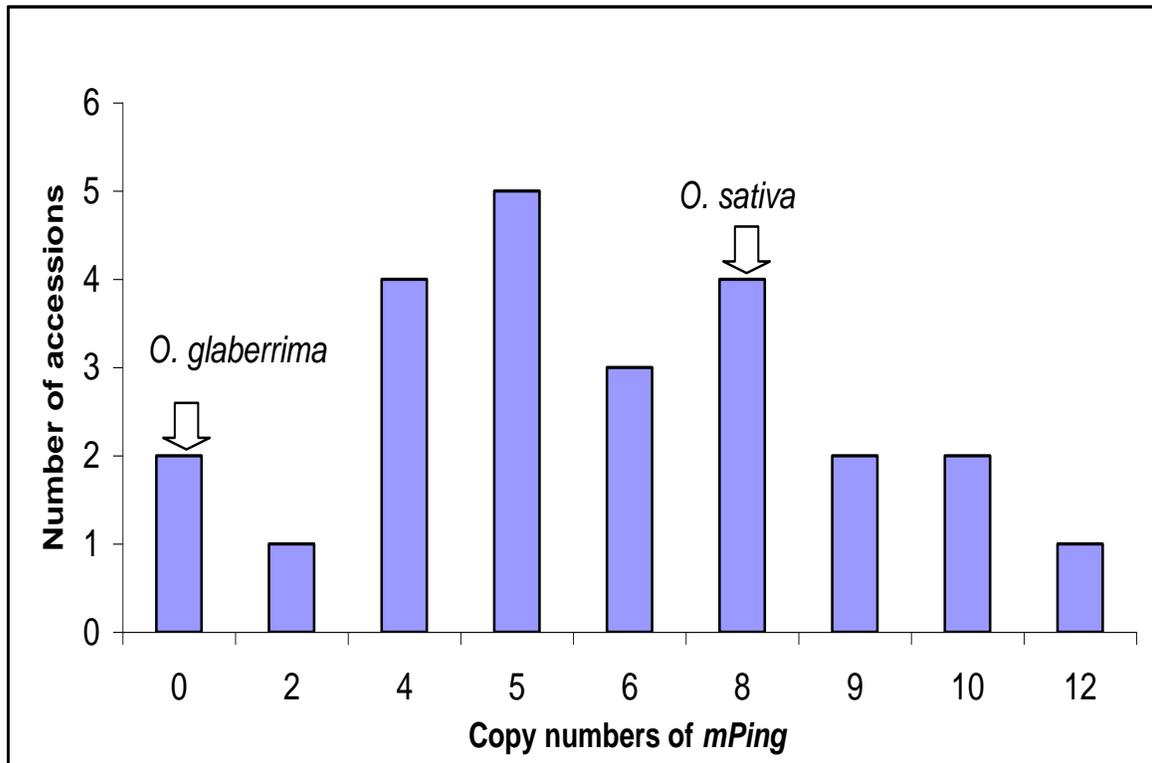
Fig. 24. **A g**raphical representation of variations in *mPing* copy numbers among the NERICA lines. Five of the NERICA lines have *mPing* copy numbers more than that of their parent strain, *O. sativa*.

DNA blot analysis using specific probe for *Pong* showed that the recurrent parent had nine copies of *Pong*, and the copy number of *Pong* in NERICA lines ranged from three to 10 (Fig. 25). On the other hand, no hybridization signal was detected from the donor parent. WAB450-24-3-1-P37-HB and WAB450-24-3-4-P18-3-1, WAB450-I-B-P-163-2-1 and WAB450-I-B-P-163-4-1 showed same banding patterns between each other, but no NERICA lines showed same banding pattern as the recurrent parent. No band common to all lines was detected, but all lines excepting WAB450-I-B-P-157-2-1 had parental bands derived from the recurrent parent. In addition to parental bands, *de novo* bands that are not in the recurrent parent were detected in 11 NERICA lines. These indicate that not only *mPing* but also *Pong* has mobilized in these lines.
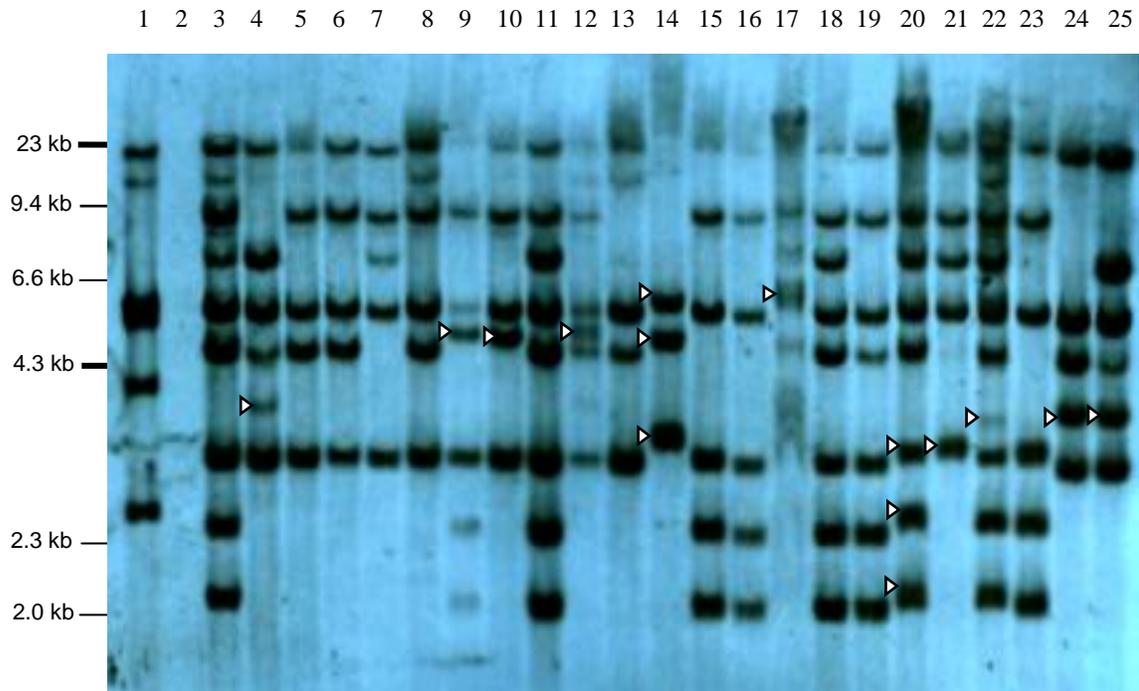
Fig. 25. Transposition of *Pong* in NERICA lines as evidenced by DNA blot analysis. Genomic DNA was digested with *Eco R*I and probed with DIG labeled *Pong* probe. The migration of DNA marker is indicated on the left in units of Kilo-base pairs (Kb). Lane numbers represent the serial numbers of the accessions as listed in Table 7. Arrowheads indicate the new bands.

Table 8. Copy number of *mPing* and *Pong* in NERICA lines

	mPing		Pong	
	Total	*de novo* *	Total	*de novo* *
CG14 (Donor parent)	0	-	0	-
WAB450-104 (Recurrrent parent)	8	-	9	-
WAB450-11-1-P26-2-HB	5	2	6	1
WAB450-24-3-1-P37-HB	4	1	5	0
WAB450-24-3-4-P18-3-1	6	4	5	0
WAB450-24-3-P3-1-HB	6	4	5	0
WAB450-25-2-9-1-1	9	6	6	0
WAB450-5-1-BL1-DV6	4	3	7	1
WAB450-I-B-P-129-HB	8	7	5	1
WAB450-I-B-P-133-HB	7	3	8	0
WAB450-I-B-P-135-HB	5	3	7	1
WAB450-I-B-P-157-1-1	3	2	5	0
WAB450-I-B-P-157-2-1	3	2	3	3
WAB450-I-B-P-163-2-1	11	10	6	0
WAB450-I-B-P-163-4-1	11	10	6	0
WAB450-I-B-P-20-HB	9	6	5	1
WAB450-I-B-P-23-HB	6	2	8	0
WAB450-I-B-P-24-HB	0	0	7	0
WAB450-I-B-P-28-HB	12	8	7	3
WAB450-I-B-P-33-HB	4	1	5	1
WAB450-I-B-P-1-1	8	3	10	1
WAB450-I-B-P-82-2-1	1	0	6	0
WAB450-I-B-P-38-HB	5	1	5	1
WAB450-11-1-P31-1-HB	6	3	6	1

*, Copy number of *de novo* bands that are not in the recurrent parent.

4.3.3 Molecular characterization of *de novo mPing* insertion sites in NERICA lines

Change in banding pattern and increase in copy number of *mPing* in NERICA lines strongly point out the mobilization of *mPing* in their genomes. However, there is a possibility that recombination of chromosomal regions containing *mPing* is occurred by interspecific cross. To confirm the mobilization of *mPing*, we carried out PCR analysis using locus specific primers designed on the rice genomic sequences adjacent to *mPing*. To this end, we first randomly chose putative *de novo* bands from three different NERICA lines, and reamplified PCR products using *mPing*- and adaptor-specific primers. Sequence analysis revealed that the reamplified products consisted of 5′-flanking sequence and the expected partial *mPing* sequence. This implies that *de novo* bands in TD gel represent not non-specific amplifications but *mPing de novo* insertions in the NERICA lines. BLAST search in RAP-DB showed that the putative new insertions were into single copy regions or near transcribed DNA (Table 9). Based on database search results, locus specific primers were designed on both 5′ and 3′-flanking sequences of *mPing* insertion sites (sequence available upon request). When PCR analysis was carried out using locus specific primer pair, only NERICA line that produced *de novo* band yielded 430 bp longer fragment than other lines and the parents (Fig. 27 A; Fig. 28). To further verify the presence of *mPing* at these loci, PCR analysis was carried out at all seven loci using *mPing*-internal primer in conjugation with each of the locus-specific 5′-flanking primer. In all primer combinations, obvious band was detected in NERICA line that produced 430 bp longer band when using locus-specific 5′- and 3′-flanking primers (Fig. 27 B). Both parents yielded only shorter fragment at all seven loci using locus-specific primer pairs. This observation was also completely consisted with the PCR result

using *mPing*-internal and locus-specific 5′-flanking primer pairs. These results confirmed that *de novo* bands in TD gel are the consequence of *mPing de novo* insertions, and *mPing* preferentially transposes to single copy regions or near transcribed DNA in NERICA lines.
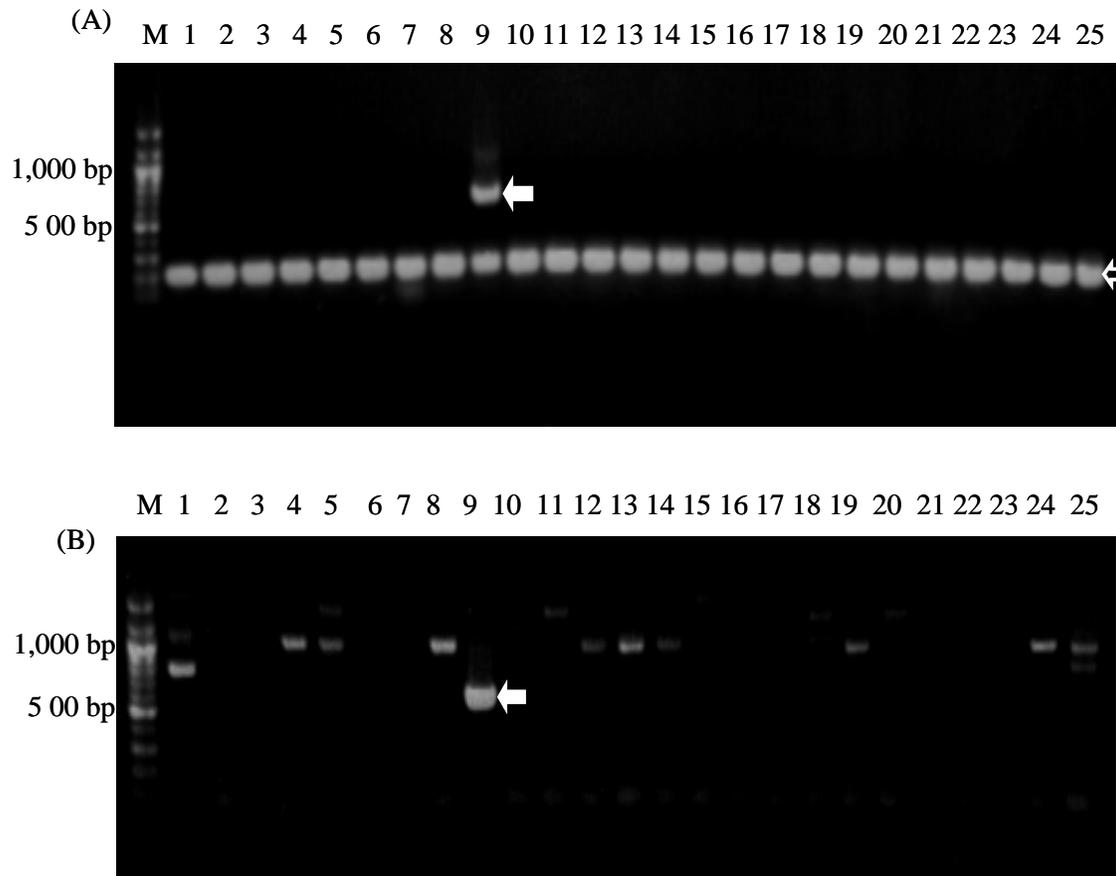
Fig. 26. Confirmation of *de novo mPing* insertion site by locus specific PCR. PCR done with the flanking primers bracketing the *mPing* element (A) and with a flanking primer and an *mPing* specific primer (B) showed the amplification of the appropriate PCR amplicon only in the respective line. The filled arrows indicate the amplicons containing *mPing* insertions (in A: 653-bp and in B: 560-bp) and the unfilled arrow indicates the loci with out *mPing* (223-bp).
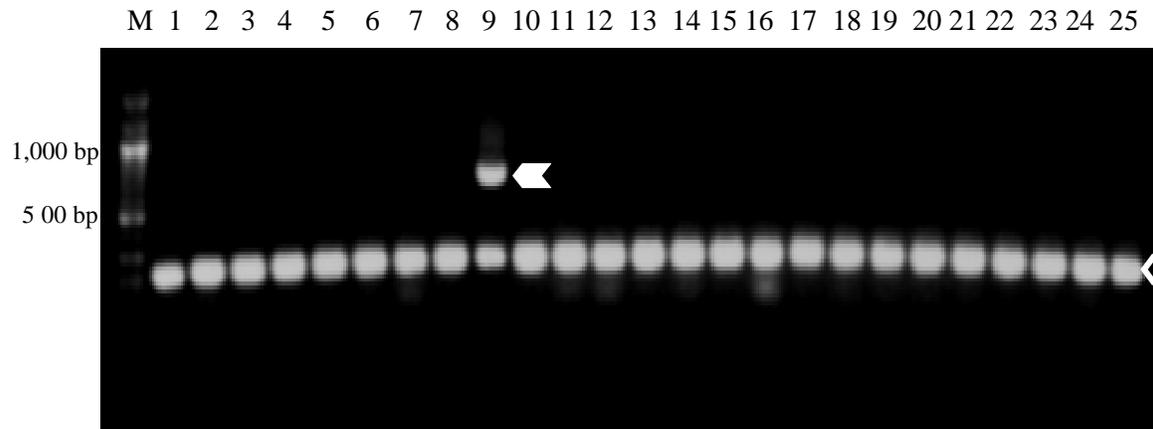
Fig. 27. Confirmation of another *de novo mPing* insertion site by locus specific PCR. PCR done with the flanking primers bracketing the *mPing* element showed the amplification of the appropriate PCR amplicon only in the respective line. The filled arrow indicates the amplicon containing *mPing* insertion (643-bp) and the unfilled arrow indicates the amplicons with out *mPing* (213-bp).

Table 9. Molecular characterization of seven *de novo mPing* insertion sites in NERICA lines

NERICA lines	BAC/PAC	Chr. No.	Position (bp)	Adjacent gene		
				Acc. No.	Description	Distance
WAB450-24-3-P3-1-HB	AP 008214	8	21,842,346	AK120424	Similar to receptor-like protein kinase precursor (EC 2.7.1.37)	2186-bp upstream
WAB450-I-B-P-163-2-1	AL607101	3	21,786,723	AK062576	Conserved hypothetical protein	3.8-kb upstream
WAB450-5-1-BL1-DV6	AC 105363	3	1,152,434	Os03g0121200	Similar to Peroxidase 1	Intron
WAB450-5-1-BL1-DV6	AC 137547	3	12,118,984	AK102918	Zinc finger, CCCH-type domain containing protein	5' UTR
WAB450-5-1-BL1-DV6	AC 079935	10	12,307,223	Os10g0377400	Similar to Ras-related protein Rab11D	Intron
WAB450-I-B-P-135-HB	AC 134885	3	12,575,991	Os03g0337500	Similar to Potassium transporter 2 (AtPOT2)	5' UTR
WAB450-I-B-P-135-HB	AC 084884	10	2,973,149	Os10go150200	Sterile alpha motif homology domain containing protein	1256-bp upstream

4.4 Discussion

Until now, it has been reported that *mPing* family transposon, which consists of *mPing*, *Ping*, and *Pong*, is present in only *O. sativa* and its direct ancestor *O. rufipogon* (Hu et al. 2006). Moreover, *mPing* and *Pong* are distrbuted in all *O. sativa* and *O. rufipogon*, but the distribution of *Ping* is limited in *O. sativa* ssp. *japonica* and some of *O. rufipogon* accessions (Jiang et al. 2003; Kikuchi et al. 2003; Hu et al. 2006). In the present study, *mPing* family was found to be abcent in *O. gleberrima*, the donor parent of NERICA lines. *O. barthii* is known to be a direct ancestor of *O. glaberrima*, and it also lacks *mPing* family in the genome. Thus, the result in the present study strongly supports the evolutionary history of *O. glaberrima*. It was found that WAB450-104, which is *O. sativa* ssp. tropical *japonica* and the recurrent parent of NERICA lines, had *mPing* and *Pong* but not *Ping* in the genome. This result also confirms that *Ping* is absent in *O. sativa* ssp. tropical *japonica*.

Like other MITEs, *mPing* is small (430 bp) and incapable of catalysing its own transposition. Thus, the mobilization of *mPing* must be catalysed by a transposase encoded *in trans* by the autonomous elements *Ping* and/or *Pong*. It was found that not only *Ping* but also *Pong* could mobilize *mPing* (Jiang et al. 2003; Shan et al. 2005; Lin et al. 2006; Yang et al. 2007). In the present study, both the recurrent and donor parents lack *Ping*, but the transpositions of *mPing* were detected in all NERICA lines excepting WAB450-I-B-P-82-2-1. Moreover, *Pong* has transposed in NERICA lines in itself. Thus, *Pong* must be responsible for the mobilization of *mPing* in NERICA lines.

The endogenous factors responsible for transpositions of *mPing* family transposons are not clear even though various types of exogenous stresses such as tissue

culture, irradiation, and hydrostatic pressurization, have successfully mobilized them (Kikuchi et al. 2003; Nakazaki et al. 2003; Lin et al. 2006). The copy number of *mPing* dramatically varies among *japonica-indica* subspecies that are domesticated independently from a common wild species, *Oryza rufipogon* (Jiang et al. 2003). Moreover, we found that the insertion sites and copy numbers of *mPing* were considerably different among *O. rufipogon* accessions (Shanta et al. in submission). These suggest that *mPing* might be mobilized by not only exogenous stresses but also endogenous stresses. Recently, it was found that *mPing* and *Pong* were mobilized in homologous recombinant inbred lines, derived from hybridization between cultivated rice and wild rice (*Zizania latifolia* Griseb.), harboring introgressed genomic DNA from wild rice (Shan et al. 2005). All NERICA lines used in the present study are interspecific lines developed by crossing a tropical *japonica* variety (WAB 56-104) as the recurrent parent to an *O. glaberrima* variety (CG14) as the donor parent followed by eight generations of inbreeding to fix the lines (BC$_2$F$_8$ lines), even though there are other NERICA lines which passed anther culture to derive doubled haploids after the interspecific cross (Semagn et al. 2007). Moreover, it was known that, on the average, 6.3 % of donor genome was introgressed to NERICA lines (Semagn et al. 2007). Thus, the mobilizations of *mPing* and *Pong* in NERICA lines were most probably activated by introgression of exogenic genome from the donor parent, *O. glaberrima*.

Nowadays, active transposons, such as *Ac/Ds* and *Tos17*, are used for transposon tagging to isolate genes in rice (An et al. 2005). Mutant lines harboring *Ac/Ds* system are transgenic rice plants, hence environmental concern must be mattered. However, rice plants are larger than other model plants, such as *Arabidopsis* and tomato, indicating that

it is difficult to handle thousands of transgenic rice plants in a greenhouse. Moreover, it has been known that inactivation of *Ds* occur in later generations (Kim et al. 2002). Although *Tos17* tagging system is easy to generate thousands of mutant lines, tagging efficiency of Tos17 is only 5-10 % (Hirochika et al. 1996). In the present study, it was found that interspecific cross induced the mobilization of *mPing* in NERICA lines, and *mPing* preferentially insert within or near genes. Ability of *mPing* family transposons to be mobilized by interspecific cross makes them suitable elements for transposon tagging system in rice, because no somaclonal variations will occur in the mutant lines induced by *mPing* mobilization. Moreover, *de novo mPing* insertions within or near to genes make it more suitable for identifying the target genes. Until now, there are no reports of *mPing* mobilization in both *O. sativa* ssp tropical *japonica* and *O. glaberrima*. Thus, the results of the present study provide sufficient evidences for the development of *mPing* into an efficient gene tagging tool for functional genomics and molecular characterizations of various rice species, varieties, and cultivars.

Chapter 5

General discussion

5.1 Origin of MITEs

DNA transposons are typically grouped into nine known superfamilies (Wicker et al. 2007) and, in principle, members of the same family are derived from a common ancestral element. Although MITEs are a particular type of defective DNA transposons, the circumstances in which they are produced are still unknown. Some of them have been shown to be originated by internal deletion from autonomous DNA transposons encoding the transposase (Guermonprez et al. 2008). *Emigrant* MITE in *Arabidopsis* is believed to have been originated by an internal deletion of a single-copy *pogo*-like element called *Lemi1* because they share extensive sequence similarities (Loot et al. 2006). Similarly, *P Instability Factor* (*PIF*) and *Tourist*-like MITE family called *miniature PIF* (*mPIF*) share identical TIRs and subterminal sequences, indicating the possibility that *mPIF* is derived from the autonomous element *PIF* (Zhang et al. 2004). Fig. 29 is a schematic representation of an evolutionary model for origin of MITEs proposed by Feschotte et al. (2002). The majority of the MITEs may have originated from *TC1/mariner* and *PFI/IS5* superfamilies of DNA transposons (Jiang et al. 2004a). However, the origin of other MITEs like *Ditto*, *Kiddo*, *Mashu*, and *Wanderer* can not be explained because no autonomous elements related to these MITEs have been identified. Furthermore, not all the autonomous elements have the ability to produce MITEs. Recently, it has been

reported that *MtLemi1* (*Lemi1* related element of *Medicago truncatula*) had attained more than 30 copies, whereas no *MtLemi1-* related MITEs have been generated yet (Guermonprez et al. 2008). In the present study, it was found that MITEs identified in *O. sativa* were widely present in genus *Oryza*. Bioinformatic analysis using database has displayed its great force in the discovery of hidden transposons. The completion of wild rice genome sequence will allow the researchers not only to identify autonomous elements but also to elucidate the molecular evolution of each MITEs in genus *Oryza*.

Autonomous
transposons-
few copies

Transposase

Internal deletion
derivatives -few
copies, variable sizes

Amplification of MITEs-
many copies, conserved
size
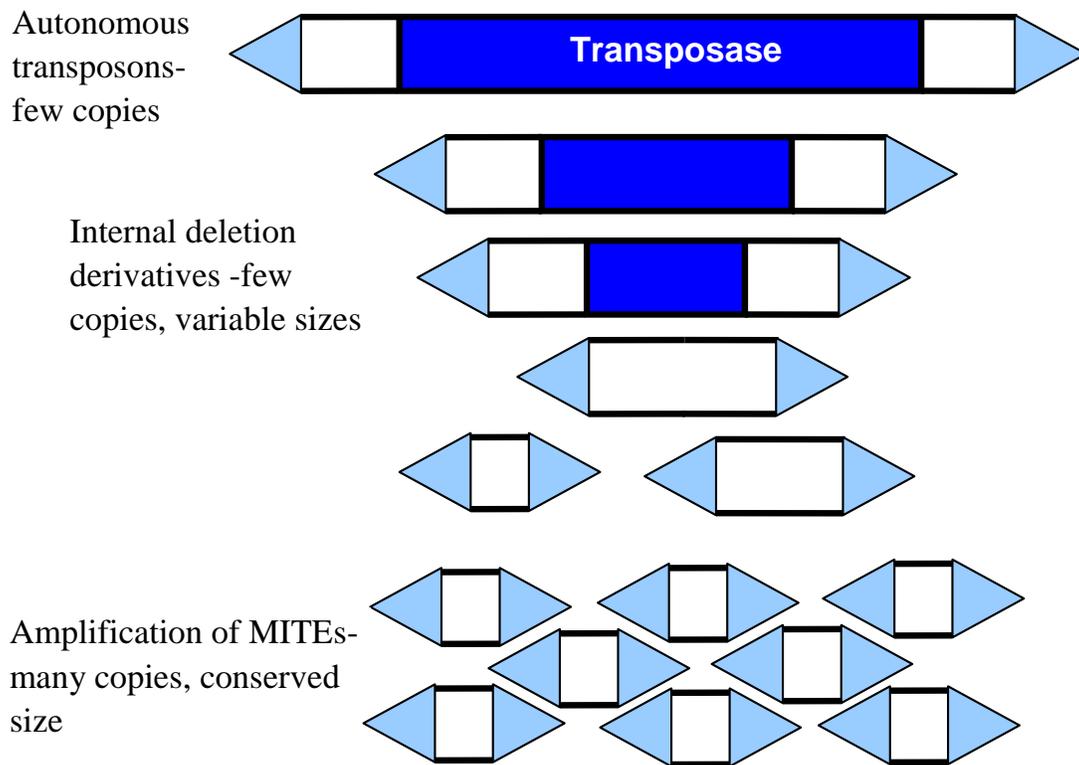
Fig. 28. Model for origin and amplification of MITEs

5. 2 Transposition and multiplication mechanisms of MITEs

MITEs are nonautonomous elements, hence they cannot catalyze their own movements. The classification of MITEs suggests that they are mobilized by a transposase provided from an autonomous member of their own or a related family. In several plant and animal species, a few autonomous *Tc1–Mariner* TEs are found to be responsible for the origin and activation of large populations of non-autonomous elements, such as tens of thousands of *Stowaway* MITEs in rice (Feschotte et al. 2003). Recently, it has been reported that a fungal MITE *mimp1* could transpose by utilizing the transposase coded by *impala*, an autonomous element of *Tc1* family (Dufresne et al. 2007). Likewise, *PIF–Harbinger* TEs control the activation of non-autonomous *Tourist* MITEs in plants, nematodes, insects, and fish (Jiang et al. 2004a). Many of the autonomous elements are not residing in the existing genome database. This leads to an another transposition mechanism called "cross-mobilization" for MITE transposition and amplification. Cross-mobilization refers to the situation where the MITEs are mobilized by elements from which they had not descended by deletion (Jaing et al. 2004a). In the present study, *mPing* in *O. rufipogon* strains showed high levels of interspecific polymorphisms both for banding patterns and copy numbers even in the accessions without *Ping*. Furthermore, the mobilization of *mPing* was indeed detected in NERICA lines without *Ping*. On the basis of them, *Pong* must be the source of transposase responsible for the mobilization of *mPing* in these plant materials, even though *Pong* is considered not to be ancestral element of *mPing*, i.e. the mobilization of *mPing* induced by *Pong* is cross-mobilization.

The copy number of MITEs hardly increases due to the cut-and-paste mechanism. However, in fact, MITEs reach copy number in thousands and tens of thousands. Now, two possible mechanisms are proposed to explain this paradox (Feschotte and Pritham 2007). The first mechanism is the repair of double-strand break left by excision of MITE. When MITE is present on the homologous chromosome, gap repair homologous recombination results in the reintroduction of the transposon at the donor site. The second mechanism is the transposition of the element during DNA replication from a newly replicated chromatid to an unreplicated site. MITE is thereby is effectively replicated twice. Successful and rapid proliferation, however, depends on multitude of factors, such as transposition rates, number of autonomous and non-autonomous copies, environmental factors, and timing of invasion, host defense mechanisms, and selection influences. Several studies have shown that the environmental and genomic stresses increase transposition activity of MITEs. The results from the present study will give cues to evaluate the factors responsible for MITE mobilization in the nature.

Summary

The belief that TEs are solely selfish and parasitic DNA sequences is beginning to change and the prospect of TEs as important genetic entity have been given scientific recognition. The importance of TEs is accounted for creation of genetic variations and for adaptive evolution of genomes. Among different types of TEs, MITEs are likely to contribute significantly in gene regulation and host genome evolution, because MITEs are usually present as high copy number elements in genomes and prefer to insert within or near genes. In rice (*Oryza sativa* L.), MITE is the numerically predominant type of TE. This indicates that MITEs have significantly contributed to rice genome evolution. However, the research resolving how MITEs and rice genome coevolved could not be carried out, because no MITE family had been shown to be actively transposing until *mPing* was identified. In order to elucidate the impact of MITE on rice genome evolution, it is necessary to analyze distribution, proliferation, and transposition of *mPing* family transposons in genus *Oryza*.

In this study, presence and distribution of eight kinds of MITEs identified in *O. sativa* (*Castaway_Os*1, *Ditto-Os*1, *Kiddo*, *mPing*, *Mashu*, *Stowaway2_Os* like, *Stowaway17_Os* like, and *Wanderer*) were investigated in 19 *Oryza* species. PCR amplifications revealed that *Mashu*, *Wanderer*, and *Stowaway17_Os* like MITE are present in all species of AA genome. *Castaway_Os1* is present in all the AA genome species excluding *O. meridionalis*. *Stowaway2_Os* like MITE is present in *O. punctata* (BB, BBCC) and *O. minuta* (BBCC) besides every AA genome species. *Kiddo* is present in AA, BB, BBCC, and CC genome species. *Ditto_Os1* has the widest distribution among

the eight MITEs analyzed because of its presence in species with AA, BB, BBCC, CC, CCDD, EE, and FF genomes, whereas *mPing* has the narrowest distribution. *mPing* is present only in two species, *O. rufipogon* and *O. sativa*, among the 19 species assessed. These results indicate that *Ditto_Os1* is the oldest among all the MITEs examined in this study and *mPing* had originated later than the other MITEs during the evolutionary process of genus *Oryza*: in other words, *mPing* is a comparatively young MITE. Usually, young TE remains actively mobile in the genome. Thus, *mPing* may be still active not only in *O. sativa* but also in *O. rufipogon*.

In order to verify the possible mobilization of *mPing* in *O. rufipogon*, distribution of autonomous elements, *Ping* and *Pong*, was investigated in 37 *O. rufipogon* accessions. As a result, *Pong* was detected in all accessions, whereas *Ping* was present in 17 out of 37 accessions. Transposon display (TD) analysis revealed that the insertion sites and copy numbers of *mPing* varied considerably among the *O. rufipogon* accessions. A phylogenetic tree constructed by using haplotypes around *qSH1* which is a major QTL responsible for the reduction of seed shattering showed that *O. rufipogon* accessions used in this study could be divided into three major clusters independent of the copy number of *mPing* and the presence of *Ping*. These results propose that the *mPing* family had already evolved and was present in the common ancestor of the *O. rufipogon* accessions, and that, later on, *Ping* was eliminated from the genomes of some of the accessions by random drift and/or natural selection along with the advancement of diversification in *O. rufipogon*. Moreover, *O. rufipogon* accessions were found to have several *Kiddo* elements at the same chromosomal positions. Thus, insertion polymorphisms and

different copy number of *mPing* in *O. rufipogon* accessions indicate that *mPing* may be currently active even in *O. rufipogon*.

Possible mobilization of *mPing* in *O. rufipogon* accessions suggests the availability of *mPing* in transposon tagging system for *Oryza* species other than *O. sativa*. The distribution of *mPing* family transposons was investigated in NERICA lines that are the interspecific lines developed by crossing a tropical *japonica* variety as the recurrent parent to an *O. glaberrima* variety as the donor parent followed by eight generations of inbreeding to fix the lines. NERICA lines and the recurrent parent had *mPing* and *Pong* but no *Ping*, whereas the donor parent had no *mPing* family transposons. In TD analysis, NERICA lines produced *de novo* bands that are not in the recurrent parent, indicating that *mPing* transpose from the original position to another position in these lines. Locus specific PCR confirmed that *de novo* bands in TD gel were the consequence of *mPing de novo* insertions, and *mPing* preferentially transposed to single copy regions or near transcribed DNA in NERICA lines. Moreover, *de novo* bands of *Pong* were detected in 11 NERICA lines. From these results, it was found that the mobilization of *mPing* and *Pong* in NERICA lines was most probably activated by introgression of exogenic genome from the donor parent, *O. glaberrima*. Furthermore, the results of the present study provide sufficient evidences for the development of *mPing* into an efficient gene tagging tool for functional genomics and molecular characterizations of various rice species, varieties, and cultivars.

Acknowledgements

The accomplishment of this doctoral dissertation marks the conclusion of three years of hard work and eventful journey for which I would like to express my thankfulness to numerous people for their care, hospitality, help, and support throughout my stay. First and foremost, my utmost gratitude goes to my supervisor, Professor Takatoshi Tanisaka for his guidance, scientific advice, and invaluable support during my study. I am greatly indebted to him for his constant assistance, parental care, and concerns to make my life and study in Japan, a comfortable, fruitful, and incredibly memorable one.

My special thanks go to Assistant Professor, Takuji Tsukiyama for his precious time, help, critical discussions, and suggestions. Moreover, I appreciate his great efforts and constructive comments during the preparation and revision of the manuscripts.

I extend my sincere gratefulness to Dr. Yutaka Okumoto and Dr. Tetsuya Nakazaki for their brilliant ideas, continuous encouragements, and suggestions from which this research work has benefited tremendously. I am also thankful to Dr. Masayoshi Teraishi, whose previous works on transposable elements led to the identification *mPing* and its effect on gene expression.

I am extremely obliged to all my friends, seniors, juniors, and other members of the Laboratory of Plant Breeding, Kyoto University for their warm co-operation and immense help without which this thesis would not have been completed on time.

References

Aggarwal, R.K., Brar, D.S., and Khush, G.S. (1997) Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. *Mol. Gen. Genet.* **254 (1)**: 1-12.

Aggarwal, R.K., Brar, D.S., Nandi, S., Huang, N., and Khush, G.S. (1999) Phylogenetic relationships among *Oryza* species revealed by AFLP markers. *Theor. Appl. Genet*. **98**: 1320-1328.

Ammiraju, J.S., Zuccolo, A., Yu, Y., Song, X., Piegu, B., Chevalier, F., et al. (2007) Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in genus *Oryza*. *Plant J.* **52**: 342-351.

An, G., Jeong, D.H., Jung, K.H., and Lee, S. (2005) Reverse genetic approaches for functional genomics of rice. *Plant Mol. Biol.* **59**: 111-123.

Benjak, A., Forneck, A., and Casacuberta, J. M. (2008) Genome-wide analysis of the "cut- and-paste" transposons of grapevine. *PLoS ONE* **3(9)**: e3107. doi: 10.1371/journal.pone.0003107.

Bennetzen, J.L. (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**: 251-269.

Bennetzen, J.L. (2007) Patterns in grass genome evolution. *Curr. Opin. Plant Biol.* **10**: 176-181.

Bergthorsson, U., Adams, K.L., Thomason, B., and Palmer, J.D. (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**: 197-201.

Biemont, C. and Vieira, C. (2006) Genetics: Junk DNA as an evolutionary force. *Nature* **443**: 521-524.

Bundock, P. and Hooykaas, P. (2005) An *Arabidopsis hAT*-like transposase is essential for plant development. *Nature* **436**: 282-284.

Bureau, T.E., Ronald, P.C., and Wessler, S.R. (1996) A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 8524-8529. doi:10.1073/pnas.93.16.8524. PMID:8710903.

Casa, A. M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., and Wessler, S. R. (2000) The MITE family Heartbreaker (Hbr): molecular markers in maize. *Proc. Natl. Acad. Sci. USA* **97**: 10083-10089.

Cheng, C., Tsuchimoto, S., Ohtsubo, H., and Ohtsubo, E. (2002) Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. *Genes and genetic systems* **77 (5):** 323-334.

Cowan, R.K., Hoen, D.R., Schoen, D.J., and Bureau, T.E. (2005) MUSTANG is a novel family of domesticated transposases genes found in diverse Angiosperms. *Mol. Biol. Evol.* **22 (10)**: 2084-2089.

Diao, X., Freeling, M., and Lisch, D. (2006) Horizontal transfer of a plant transposon. *PLoS Biol.* **4 (1)**: 119-128.

Dufresne, M., Hua-Van, A., Wahab, H.A., Barek, S.B.M., Vasnier, C., Teysset, L., Kema, G.H.J., and Daboussi M.J. (2007) Transposition of a Fungal miniature inverted-repeat transposable element through the action of a *Tc1*-like transposase. Genetics **175**: 441-452.

Feng, Q., Zhang, Y., Hao, P. *et al*. (71 co-authors) (2002) Sequence and analysis of rice chromosome 4. *Nature* **420**: 316-320.

Fedoroff, N.V., Furtek, D.B., and Nelson, O.E. (1984) Cloning of the *bronze* locus in maize by a simple and generalizable procedure using the transposable controlling element Activator (Ac). *Proc. Natl. Acad. Sci. USA* **81**: 3825-3829.

Fedoroff, N. (2000) Transposons and genome evolution in plants. *Proc. Natl. Acad. Sci. USA* **97 (13)**: 7002-7007.

Feschotte, C., Jiang, N., and Wessler, S.R. (2002) Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* **3**: 329-341.

Feschotte, C., Swamy, L., and Wessler, S.R. (2003) Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with Stowaway Miniature Inverted Repeat Transposable Elements (MITEs). *Genetics* **163**: 747-758.

Feschotte, C., Osterlund, M.T., Peeler, R., and Wessler, S.R. (2005) DNA-binding specificity of rice *mariner*-like transposases and interactions with *Stowaway* MITEs. *Nucleic Acids Res.* **33 (7)**: 2153-2165.

Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331-68.

Fujino, K., Sekiguchi, H., and Kiguchi, T. (2005) Identification of an active transposon in intact rice plants. *Mol. Gen. Genomics* **273**: 150-157.

Ge, S., Sang, T., Lu, B.R., and Hong, D.Y. (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl. Acad. Sci. USA* **96 (25)**: 14400-14405.

Guermonprez, H., Loot, C., and Casacuberta, M.J. (2008) Different strategies to persist: The *pogo*-like *Lemi1* transposon produces miniature inverted-repeat transposable elements or typical defective elements in different plant genomes. *Genetics* **180**: 83-92.

Guo, Y.L. and Ge, S. (2005) Molecular phylogeny of *Oryzeae* (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *American Journal of Botany* **92 (9)**: 1548-1558.

Han, B. and Xue, Y. (2003) Genome-wide intraspecific DNA-sequence variations in rice. *Curr. Opin. Plant Biol.* **6**: 134-138.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M. (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA* 93: 7783-7788.

Holmes, I. (2002) Transcendent elements: Whole –genome transposon screens and open evolutionary questions. *Genome Res.* **12**: 1152-1155.

Hu, H., Mu, J., Zhang, H.J., Tao, Y.Z., and Han, B. (2006) Differentiation of a Miniature inverted transposable element (MITE) system in Asian rice cultivars and its inference for a diphyletic origin of two subspecies of Asian cultivated rice. *Journal of Integrative Plant   Biology* **48 (3)**: 260-267.

Huang, J.T. and Dooner, H.K. (**2008**) Macrotransposition and other complex chromosomal restructuring in maize by closely linked transposons in direct orientation. *Plant Cell* **20:** 2019-2032.

Hurst, G.D.D. and Werren, J.H. (2001) The role of selfish genetic elements in eukaryotic evolution. *Nat. Rev. Genet.* **2**: 597-606.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* **436**: 793-799.

Iwamoto, M., Nagashima, H., Nagamine, T., Higo, H., and Higo, K. (1999) A *Tourist* element in the 5'-flanking region of the catalase gene *Cat*A reveals evolutionary relationships among *Oryza* species with various genome types. *Mol. Gen. Genet.* **262**: 493-500.

Jeon, J.S. and An, G. (2001) Gene tagging in rice: a high throughput system for functional genomics. *Plant Science* **161**: 211-219.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., and Wessler, S.R. (2003) An active DNA transposon family in rice. *Nature* **421**: 163-167.

Jiang, N., Feschotte, C., Zhang, X., and Wessler, S.R. (2004 a) Using rice to understand the origin and amplification of MITEs. *Curr. Opin. Plant Biol.* **7**: 115-119.

Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., and Wessler, S.R. (2004 b) Pack-MULE transposable elements mediate gene evolution in plants. *Nature,* **431**: 569-573.

Jones, M.P., Dingkuhn, M., Aluko, G.K., and Semon, M. (1997) Interspecific *Oryza sativa* L. x *O. glaberrima* Steud. Progenies in upland rice improvement. *Euphytica* **92**: 237-246.

Juretic, N., Hoen, D.R., Huynh, M.L., Harrison, P.M., and Bureau, T.E. (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res.* **15**: 1292-1297.

Kanazawa, A., Akimoto, M., Morishima, H., and Shimamoto, Y. (2000) Inter- and intra-specific distribution of *Stowaway* transposable elements in AA-genome species in rice. *Theor. Appl. Genet.* **101**: 327-335.

Kapitonov, V.V. and Jurka, J. (2006) Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **103 (12)**: 4540-4545.

Kashkush, K., Feldman, M., and Levy, A.A. (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* **33**: 102-106.

Kikuchi, K., Terauchi, K., Wada, M., and Hirano, H.Y. (2003) The plant MITE *mPing* is mobilized in anther culture. *Nature* **421**: 167-170.

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

Konishi, S., Izawa, T., Lin, S., Ebana, K., Fukuta, Y., Sasaki, T., and Yano, M. (2006) An SNP caused loss of seed shattering during rice domestication. *Science* **312 (5778)**: 1392-1396. doi: 10.1126/science.1126410.

Kovach, M.J., Sweeney, M.T., and McCouch, S.R. (2007) New insights into the history of rice domestication. *TRENDS in Genetics* **23 (11)**: 577-587.

Kumar, S. and Fladung, M. (2003) Somatic mobility of the maize element *Ac* and its utility for gene tagging in aspen. *Plant Mol. Biol.* **51**: 643-650.

Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**: 150-163.

Le, Q.H., Wright, S., Yu, Z., and Bureau, T. (2000) Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **97 (13)**: 7376-7381.

Linares, O.F. (2002) African rice (*Oryza glaberrima*): History and future potential. *Proc. Natl. Acad. Sci. USA* **99**: 16360-16365.

Lin, X., Long, L., Shan, X., Zhang, S., Shen, S., and Liu, B. (2006) *In planta* mobilization of *mPing* and its putative autonomous element *Ping* in rice by hydrostatic pressurization. *Journal of Experimental Botany* **57 (10)**: 2313-2323.

Liu, H. and Visner, G.A. (2007) Applications of *Sleeping Beauty* transposons fro non-viral gene therapy. *IUBMB Life* **59 (6)**: 374-379.

Londo, J.P, Chiang, Y.C., Hung, K.H., Chiang, T.Y., and Schaal, B.A. (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci. USA* **103 (25)**: 9578-9583.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S.S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, R.A., and Wing, R.A. (2000) Rice transposable elements: A survey of 73,000 sequence-tagged-connectors. *Genome Res.* **10**: 982-990.

McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proc. Nat. Acad. Aci. U. S. A.* **36**: 344-355.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., and Rafalski, A. (2005) Gene duplication and exon shuffling by *Helitron*-like transposons generate intraspecies diversity in maize. *Nature genetics* **37**: 997-1002.

Morishima, H. and Barbier, P. (1990) Mating system and genetic structure of natural populations in wild rice, *Oryza rufipogon*. *Plant Species Biol*. **5**: 31-39.

Murray, M. and Thompson, W.F. (1980) Rapid isolation of high-molecular weight plant DNA. *Nucleic Acids Res*. **8**: 4321-4325.

Naito, K., Cho, E., Yang, G., Campbell, M.A., Yano, K., Okumoto, Y., Tanisaka, T., and Wessler, S.R. (2006) Dramatic amplification of a rice transposable element during recent domestication., *Proc. Natl. Acad. Sci. USA* **103 (47)**: 17620-17625.

Nakazaki, T., Okumoto, Horibata, Y.A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H., and Tanisaka, T. (2003) Mobilization of a transposon in the rice genome. *Nature* **421**: 170-172.

Ni, J., Colowit, P.M., and Mackill, D.J. (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Science* **42**: 601-607.

Onishi, K., Takagi, K., Kontani, M., Tanaka, T., and Sano, Y. (2007) Different patterns of genealogical relationships found in the two major QTLs causing reduction of seed shattering during rice domestication. *Genome* **50**: 757-766.

Opperman, R., Emmanuel, E., and Levy, A.A. (2004) The effect of sequence divergence on recombination between direct repeats in Arabidopsis. *Genetics*, **168**: 2207-2215.

Park, K.C., Lee, J.K., Kim, N.H., Shin, Y.B., Lee, J.H., and Kim, N.S. (2003) Genetic variation in *Oryza* species detected by MITE-AFLP. *Genes and Genetic Systems* **78**: 235-243.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D.S., Jackson, S., Wing, R.A., and Panaud, O. (2006) Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic

expansions in *Oryza   australiensis*, a wild relative of rice. *Genome Res.* **16**: 1262-1269.

Ramirez, L.M., Lewis, K.C., Landsman, D., and Jordan, I.K. (2005) Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet. Genome Res.* **110**: 333-341 (DOI: 10.1159/000084965).

Sarla, N. and Swamy, B.P.M. (2005) *Oryza glaberrima*: A source for the improvement of *Oryza sativa*. *Current Science* **89**: 955-963.

Sang, T. and Ge, S. (2007) The puzzle of rice domestication. *Journal of Integrative Plant Biology* **49 (6)**: 760-768.

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

Semagn, K., Ndjiondjop, M.N., Lorieux, M., Jones, M., and McCouch, S. (2007) Molecular profiling of an interspecific rice population derived from a cross between WAB 56-104 (*Oryza sativa*) and CG 14 (*Oryza glaberrima*). *African Journal of Biotechnology* **6**: 2014-2022.

Shan, X., Liu, Z., Dong, Z., Wang, Y., Chen, Y., Lin, X., Long, L., Han, F., Dong, Y., and Liu, B. (2005) Mobilization of the Active MITE Transposons *mPing* and

Pong in Rice by Introgression from Wild Rice (*Zizania latifolia* Griseb.). *Mol. Biol. Evol.* **22(4)**: 976-990.

Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**: 272-285.

Takagi, K., Nagano, H., Kishima, Y., and Sano, Y. (2003) MITE-transposon display efficiently detects polymorphisms among the *Oryza* AA-genome species. *Breed. Sci.* **53**: 125-132.

Takata, M., Kishima, Y., and Sano, Y. (2005) DNA methylation polymorphisms in rice and wild rice strains: Detection of epigenetic markers. *Breeding Science* **55(1)**: 57-63.

Takata, M., Kiyohara, A., Takasu, A., Kishima, Y., Ohtsubo, H., and Sano, Y. (2007) Rice transposable elements are characterized by various methylation environments in the genome. *BMC Genomics* **8**: 469. doi: 10.1186/1471-2164-8-469.

Teraishi, M., Okumoto, Y., Hirochika, A., Yamagata, H., and Tanisaka, T. (1999) Identification of mutable slender glume gene in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* **261**(3): 487- 494.

Tu, Z. (1997) Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **94**: 7475-7480.

Turcotte, K., Srinivasan, S., and Bureau, T. (2001) Survey of transposable elements from rice genomic sequences. *Plant Journal* **25**: 169-180.

Uozu, S., Ikehashi, H., Ohmido, N., Ohtsubo, H., Ohtsubo, E., and Fukui, K. (1997) Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol. Biol.* **35**: 791-799.

Vaughan, A. D., Kadowaki, K., Kaga, A., and Tamooka, N. (2005) On the phylogeny and biogeography of the genus *Oryza. Breeding science* **55**: 113-122.

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

Wicker, T., Sabot, F., Huavan, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., Sanmiguel, P., and Schulman, A.H. (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**: 973-982.

Xu, J.H., Cheng, C., Tsuchimoto, S., Ohtsubo, H., and Ohtsubo, E. (2007) Phylogenetic analysis of *Oryza rufipogon* strains and their relations to *Oryza sativa* strains by insertion polymorphism of rice SINEs. *Genes Genet. Syst*. **82**: 217-229.

Yang, G., Dong, J., Chandrasekharan, M.B., and Hall, T.C. (2001) Kiddo, a new transposable element family closely associated with rice genes. *Mol. Genet. Genomics* **266(3)**: 417-424. PMID: 11713671.

Yang, G., Lee, Y.H., Shi, X., Kertbundit, S., and Hall, T.C. (2005) A two-edged role for the transposable element *Kiddo* in the *rice ubiquitin 2* promoter. *Plant cell* **17**: 1559-1568.

Yang, G., Zhang, F., Hancock, C.N., and Wessler, S.R. (2007) Transposition of the rice miniature inverted repeat transposable element *mPing* in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **104**: 10962-10967.

Yang, H.P. and Barbash, D.A. (2008) Abundant and species-specific *DINE-1* transposable elements in 12 *Drosophila* genomes. *Genome Biology* **9:** R39 doi:10.1186/gb-2008-9-2-r39.

Zhang, X., Jiang, N., Feschotte, C., and Wessler, S.R. (2004) PIF- and *Pong*-like transposable elements: distribution, evolution and relationships with *Tourist*-like miniature inverted-repeat transposable elements. *Genetics* **166**: 971-986.

Zhu, Q.H. and Ge, S. (2005). Phylogenetic relationships among A-genome species of the   genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytol.***167**:249-265.

List of publications

Shanta Karki, Takuji Tsukiyama, Yutaka Okumoto, Govinda Rizal, and Takatoshi Tanisaka. (2008) Differential distribution of miniature inverted-repeat transposable elements in wild *Oryza* species. Journal of Crop Research

Shanta Karki, Takuji Tsukiyama, Yutaka Okumoto, Govinda Rizal, Ken Naito, Masayoshi Teraishi, Tetsuya Nakazaki, and Takatoshi Tanisaka. (2008) Analysis of distribution and proliferation of *mPing* family transposons in a wild rice (*Oryza rufipogon* Griff.). Submitted to Genome

Shanta Karki, Takuji Tsukiyama, Yutaka Okumoto, Govinda Rizal, and Takatoshi Tanisaka. (2009) Mobilization of *mPing* and *Pong* transposons by interspecific cross between *Oryza sativa* and *O. glaberrima*. Submitted to Molecular Genetics and Genomics