ORIGINAL ARTICLE

Use of the *p-SINE1-r2* in inferring evolutionary relationships of Thai rice varieties with AA genome

Preecha Prathepha[1] and Supachai Samappito[2]

Abstract

Prathepha, P. and Samappito, S.
Use of the *p-SINE1-r2* in inferring evolutionary relationships of Thai rice varieties with AA genome
Songklanakarin J. Sci. Technol., 2006, 28(1) : 39-48

In a previous study we described the prevalence and distribution in Thailand of the retroposon *p-SINE1-r2*, in the intron 10 of the *waxy* gene in cultivated and wild rice with the AA genome. In this study, additional varieties of rice were collected and sequencing was used to further characterize *p-SINE1-r2*. It was found that the length of the *p-SINE1-r2* nucleotide sequences was about 125 bp, flanked by identical direct repeats of a 14 bp sequence. These sequences were compared and found to be similar to the sequences of *p-SINE1-r2* found in Nipponbare, a rice strain discussed in a separate study. However, when compared the 48 DNA sequences identified in this study, much dissimilarity was found within the nucleotide sequences of *p-SINE1-r2*, in the form of base substitution mutations. Phylogenetic relationships inferred from the nucleotide sequences of these elements in cultivated rice (*O. sativa*) and wild rice (*O. nivara*). It was found that rice accessions collected from the same geographical distribution have been placed in the same clade. The phylogenetic tree supports the origin and distribution of these rice strains.

Key words : retroposon, *p-SINE1-r2*, *Oryza sativa*, AA genome, waxy gene

[1]Ph.D.(Biology), Assoc. Prof., Walairukhavej Botanical Research Institute [2]Ph.D.(Pharmaceutical Chemistry and Natural Product), Department of Biotechnology, Faculty of Technology, Mahasarakham University, Maha Sarakham, 44000 Thailand.
Corresponding e-mail: preecha.p@msu.ac.th
Received, 14 February 2005 Accepted, 3 June 2005

บทคัดย่อ

ปรีชา ประเทพา[1] และ ศุภชัย สมัปปิโต[2]

การอนุมานสายสัมพันธ์ทางวิวัฒนาการของสายพันธุ์ข้าวไทยจากลำดับเบสของ *p-SINE1-r2*

ว. สงขลานครินทร์ วทท. 2549 28(1) : 39-48

ข้อมูลการปรากฏและการกระจายของดีเอ็นเอชนิด *p-SINE1-r2* ที่มีตำแหน่งบนอินทรอน 10 ของยีนแวคซีในประชากรข้าวไทยได้รายงานโดยผู้เขียนและการศึกษานี้ได้ดำเนินการต่อเนื่องโดยมีวัตถุประสงค์เพื่อการอนุมานสายสัมพันธ์ทางวิวัฒนาการของสายพันธุ์ข้าวโดยการวิเคราะห์ลำดับเบสของดีเอ็นเอชนิดนี้ในสายพันธุ์ข้าวที่ใช้ศึกษาจำนวน 48 ตัวอย่าง และนำข้อมูลไปสร้างเป็นแผนภูมิทางวิวัฒนาการ

ผลการศึกษา พบว่า *p-SINE1-r2* มีขนาดโมเลกุล 125 คู่เบสที่ล้อมรอบด้านปลาย 5' และปลาย 3' ด้วยดีเอ็นเอที่มีลักษณะที่เรียกว่า direct repeat มีขนาดความยาว 14 คู่เบส ซึ่งลักษณะเช่นนี้เหมือนกันกับที่พบในข้าวญี่ปุ่นพันธุ์นิปปอนบาเร ผลจากการเปรียบเทียบลำดับเบสของดีเอ็นเอชิ้นนี้ของข้าวทั้ง 48 สายพันธุ์ พบว่ามีการเปลี่ยนแปลงของเบสบนสายดีเอ็นเอ และเมื่อนำข้อมูลไปสร้างแผนภูมิความสัมพันธ์ทางพันธุกรรม พบว่าสายพันธุ์ข้าวที่มีถิ่นกำเนิดและการกระจายทางภูมิศาสตร์ในบริเวณใกล้เคียงกันจะมีความใกล้ชิดกันทางพันธุกรรม

[1]สถาบันวิจัยวลัยรุกขเวช [2]ภาควิชาเทคโนโลยีชีวภาพ คณะเทคโนโลยี มหาวิทยาลัยมหาสารคาม อำเภอเมือง จังหวัดมหาสารคาม 44000

Evolutionary studies require the use of the segment of DNA that can freely accommodate all types of mutations (point, translocation, inversion, etc.). The evolutionary study of distance measurement is used to accurately identify the true relationships among rice strains. The accuracy of these measurements is enhanced by using DNA sequences of intron because these sequences are not needed for function, i.e., mutations can accumulate without affecting function.

Studies of the rice *Waxy* gene have been energized by the dramatic finding that this gene encodes the granule-bound starch synthase (GBSS). GBSS is a key enzyme for the synthesis of amylose in rice endosperm and pollen (Okagaki and Wessler, 1988). There is evidence that a short interspersed element (SINE), named *p-SINE1*, is present in the rice *Waxy* gene (Umeda *et al*., 1991). This short interspersed element (SINE), which is defined as a nucleotide sequence, present initially as a cellular RNA transcript, has been reincorporated into the genome via a cDNA intermediate in the retroposition process (Maraia, 1995). The first plant SINE, *p-SINE1*, has been identified in *O. sativa* and its relatives (Umeda *et al*., 1991; Mochizuki *et al*., 1992). Shedlock and Okada (2000) suggested that SINE insertion analysis could be used as a tool to infer a phylogenetic relationship based on DNA sequence variation. Recently, it has been used to study the relationship of rice species with the AA genome (Cheng *et al*., 2002; Cheng *et al*., 2003).

The nucleotide sequences of the rice *Waxy* gene in two species, *O. sativa* and *O. glaberrima*, were first studied by Umeda *et al*. (1991). The results showed that the *Wx* sequence of *O. sativa* is different from that of *O. glaberrima* by base substitutions and insertions and/or deletions. However, base substitutions occurring in several exons do not severely alter the amino acid sequence of the waxy protein. Interestingly, these striking differences have been observed in the intron 10 of these two species. Specifically, *O. sativa* exhibits 139-bp, whereas there is a 139-bp deletion in *O. glaberrima*. In *O. sativa*, 125-bp of the 153-bp sequence is flanked by direct repeats of a 14-bp sequence, and this same pattern has been found in several loci in rice genomes (Mochizuki *et al*., 1992). This finding indicates that these elements have diversified the rice *Waxy* gene by insertion into introns in the rice genome. Furthermore, Mochizuki *et al*. (1993) reported that *p-SINE1-r2*,

a member of the SINE family, was found only in two closely related species with the AA genome, *O. sativa* and *O. rufipogon*. Hirano *et al.* (1994) suggested that the insertion of *p-SINE1-r2* at this locus occurred after divergence of the ancestral species of *O. sativa* and *O. rufipogon* from other species of *Oryza*.

The distribution of *p-SINE1-r2* at the *Waxy* locus among populations with the AA genome indicates that a wild ancestor of *O. sativa* has differentiated into two groups at the *Waxy* locus. This occurrence strongly corresponds to an annual-perennial differentiation in the primary gene pool of the ancestral species (Yamanaka *et al.*, 2004). Recently, Prathepha (2003), studying the Thai rice population, tried to determine if this inserted DNA was a member of the *p-SINE1* family. The study included the use of two sets of primer pairs designed by Yamakata *et al.* (2004). The result indicates that the presence or absence of the inserted DNA segment varies within and among the populations of *O. sativa, O. nivara* and *O. rufipogon*. The evidence provides the example of the most plausible explanation for these intraspecific and inter-specific variations, based on the distribution of the inserted DNA: that they are the result of being ancestrally polymorphic for insertion of the DNA segment among populations during speciation.

DNA sequence variations of microsatellites have been assumed to be selectively neutral because, in part, microsatellites are often located in a noncoding region (Ford, 2002). Similarly, *p-SINE1-r2* is located in the intron 10 of the rice *Waxy* gene. The *p-SINE1* elements have been recognized as good candidates for studying the phylogenetic relationship among *Oryza* species. This study is therefore initiated to use the *p-SINE1-r2* sequences of the representative rice accessions of *Oryza* (*O. sativa* and *O. nivara*) in deciphering the evolutionary relationships among haplotypes of these rice accessions.

Materials and Methods

Plant material

Seeds of the *Oryza* species, represented by 48 accessions (varieties, landraces, wild rice), were collected from local farmers in three regions of the country and the Phatthalung rice research center in southern Thailand (Table 1). In addition to *O. sativa* samples, six glutinous wild rice accessions and one nonglutinous wild rice accession of a closely related species, *O. nivara*, were sampled for use in this study (Table 1). Grains of each accession were germinated on moist filter paper in Petri plates, and their seedlings were transferred to individual pots. The samples were classified into indica or japonica type based on the presence or absence of 69 bp within the ORF100 region of plastid DNA by using the PCR method as described elsewhere (Kanno *et al.*, 1993).

Genomic DNA extraction

DNA was isolated from the fresh young leaves of a single 15-day seedling per accession using the CTAB method (Doyle and Doyle, 1987).

PCR amplification

To detect the insertion in the 10th intron of the rice *Waxy* gene, DNA of the second half region of this gene was amplified using a set of primers previously reported (Prathepha, 2003). The primers were Wax-2F (5'-ACGCCGGTGGAGG GCAGG AAGATCAACT-3') and Wax-4R (5'-TCAAGGA GCAGCCACGTTCTCC TTGGCG-3'). Amplifications were performed in a volume of 20 µl containing 10 mM Tris-HCl (pH 9.0), 50 mM KCl and 0.1% Triton® x-100, 1.5 mM $MgCl_2$, 0.1 mM each of dNTPs, 10 pM of each primer, and approximately 150 ng genomic DNA, and 1 U of *Taq* polymerase (Promega). Amplification was performed in a Hybaid thermal cycler, using the following parameters: 94ºC for 2 min, followed by 35 cycles of 94ºC for 1 min, 65ºC for 1 min, and 72ºC for 2 min. To prove whether the reaction was successful, PCR products were electrophoresed in a 1.4% agarose gel. To detect the presence or absence of *p-SINE1-r2*, the amplified fragments were employed as template DNAs by using a set of primers, M6 (5'-GGAGGACGTGC-AGATCG TTC-3') and R7 (5'-ACGAGTCCACC-GGTGG ACGC-3'). The primer pairs used in this case can

Table 1. The rice accessions with AA genome used in this study.

Code	Accession	Species/subspeciesa	Origin	Endosperm type[b]
SQ10016	Glutinous Japonica rice	*O. sativa*/ J	Japan	W
SQ10017	Tamasakae	*O. sativa*/ J	Japan	W
SQ10018	Yoneshiro	*O. sativa*/ J	Japan	NW
SQ10019	Gohyakumangoku	*O. sativa*/ J	Japan	NW
SQ10020	Bael Leu	*O. sativa*/ J	Northern, Thailand	W
SQ10021	Bael Jao Blao	*O. sativa*/ J	Northern, Thailand	W
SQ10022	Bael Plao Chong	*O. sativa*/ I	Northern, Thailand	W
SQ10023	Biawku	*O. sativa*/ J	Northern, Thailand	W
SQ10024	Biaw Mai Yan Rai	*O. sativa*/ J	Northern, Thailand	NW
SQ10025	Bao Bud	*O. sativa*/ J	Northern, Thailand	W
SQ10026	Lai Noi	*O. sativa*/ J	Northern, Thailand	W
SQ10027	Yim	*O. sativa*/ J	Northern, Thailand	W
SQ10028	Pirnneonyim	*O. sativa*/ J	Northern, Thailand	W
SQ10029	Ja Ngai	*O. sativa*/ J	Northern, Thailand	W
SQ10030	Sew	*O. sativa*/ I	Northern, Thailand	W
SQ10031	Ramtang	*O. sativa*/ J	Northern, Thailand	W
SQ10032	Bae Dao Derk	*O. sativa*/ J	Northern, Thailand	W
SQ10033	Be Jah	*O. sativa*/ J	Northern, Thailand	W
SQ10034	Be Lia Tia Tao	*O. sativa*/ J	Northern, Thailand	NW
SQ10035	Pheuaknam	*O. sativa*/ I	Southern, Thailand	NW
SQ10036	Paung Tawng	*O. sativa*/ I	Southern, Thailand	NW
SQ10037	Nangphaya132	*O. sativa*/ I	Southern, Thailand	NW
SQ10038	Chiangpatthalung	*O. sativa*/ J	Southern, Thailand	NW
SQ10039	Kaen Jan	*O. sativa*/ I	Southern, Thailand	NW
SQ10040	Lebnokpatthani	*O. sativa*/ I	Southern, Thailand	NW
SQ10041	Jam Pah	*O. sativa*/ I	Southern, Thailand	NW
SQ10043	Thadokkam	*O. sativa*/ I	Laos	W
SQ10044	Nam Pa	*O. sativa*/ I	Laos	W
SQ10045	Phatang	*O. sativa*/ I	Laos	W
SQ10046	Accession10046	*O. sativa*/ I	Laos	W
SQ10047	Hom Mali	*O. sativa*/ I	Laos	NW
SQ10048	Accession10048	*O. sativa*/ J	Kunming, China	NW
SQ10049	Accession10049	*O. sativa*/ J	Kunming, China	NW
SQ5085	KDML105	*O. sativa*/ I	Northeastern, Thailand	NW
SQ5086	KDML105	*O. sativa*/ I	Northeastern, Thailand	NW
SQ8320	Hao Kaen Du	*O. sativa*/ J	Northeastern, Thailand	W
SQ8321	U Kham	*O. sativa*/ I	Northeastern, Thailand	W
SQ8322	Pong Aew	*O. sativa*/ I	Northeastern, Thailand	W
SQ8323	Khitom	*O. sativa*/ I	Northeastern, Thailand	W
SQ7689	RD6	*O. sativa*/ I	Northeastern, Thailand	W
SQ10042	Accession10042	*O. nivara*/I	Northeastern, Thailand	NW
SQ5087	Accession5087	*O. nivara*/I	Northeastern, Thailand	W
SQ8215	Accession8215	*O. nivara*/I	Northeastern, Thailand	W
SQ8314	Accession8314	*O. nivara*/I	Northeastern, Thailand	W
SQ8316	Accession8316	*O. nivara*/I	Northeastern, Thailand	W
SQ8317	Accession8317	*O. nivara*/I	Northeastern, Thailand	W
SQ8318	Accession8318	*O. nivara*/I	Northeastern, Thailand	W

[a]**I =** ***indica*****; J =** ***japonica*****; assessed by deletion or addition of ORF100 in cp-DNA .**
[b]**W= waxy; NW = nonwaxy endosperm.**

amplify the target DNA and its flanking region. The PCR reaction mixture contained approximately 250 ng of amplified fragments, 10 pM primer pairs M6/R7, 1.5 mM $MgCl_2$, 0.1 mM dNTPs, and 1 Units *Taq* polymerase (Promega). The amplification conditions in this case were 30 cycles of : 1 min, 94ºC; 1 min, 55ºC; and 1.5 min, 72ºC. PCR products were electrophoresed in 2% agrose gel.

Sequencing protocol

The PCR products (upper and lower bands) were purified prior to sequencing using a QIA quick Gel extraction kit (Qiagen). Direct sequencing was performed by using the amplification primers. The Taq Dye Terminator Cycle Sequencing kit (Applied Biosystems) was amplified as recommended by the manufacturer. The PCR products were analyzed on an ABI 737A automated sequencer. For each strain, forward and reverse sequencing reactions were performed for sequence confirmation.

Sequence analysis and phylogenetic tree

Nucleotide sequences of *p-SINE1-r2* were aligned with each other using the CLUSTAL W computer program(CLUSTAL W WWW Service at the European Bioinformatics Institute,(http://www.ebi.ac.uk/clustalw). They were also aligned with the DNA sequence of *p-SINE1-r2* in the rice strain Nipponbare, which was previously reported by Mochizuki *et al*. (1992) for the beginning and the end of the sequences. The DNA sequences have been deposited in the GenBank (accession nos. AY222747-222794). These DNA sequences were also aligned by using the CLUSTAL W program and exported as a FASTA alignment for calculating the nucleotide diversity (π) (Nei, 1987), and mutation rate (θ) (Watterson, 1975), using DnaSP 3.51 software (Rozas and Rozas, 1999). Phylogenetic trees were generated by using Neighbor-Joining (NJ) as distance method (Saitou and Nei, 1987) with the TREECON 1.3b computer program (Van de Peer and De Wachter (1997). Bootstrap values, taken as an index of support for strain clades, were calculated from 100 replicates. The phylogentic tree was outgroup rooted using the DNA sequence of *p-SINE1-r2* of *O. rufipogon*, because its represented a progenitor of cultivated rice.

Results and Discussion

Genome polymorphism caused by *p-SINE1-r2* insertion in *Waxy* locus

As shown in Figure 1, PCR-amplified fragments of the second half of the *Waxy* gene were monomorphic in *O. sativa*. They exhibited only the upper amplified fragments. In *O. nivara*, the PCR results showed polymorphism among individuals. Most strains (or accessions) of *O. nivara* possessed the upper DNA fragments (carried the inserted DNA), while other strains showed both upper and the lower DNA fragments (without the inserted DNA) suggesting that they are heterozygotes. Subsequent PCR amplifications

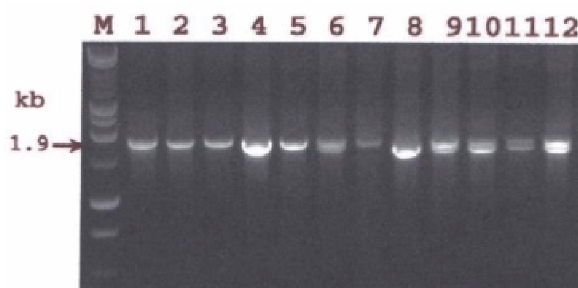

Figure 1. A garose gel showing the amplified bands. Lanes 1,2,3,5 and 7 showed one band (U = upper band) consisted of *p-SINE1-r2*. Lanes 4 and 8 showed one band (L = lower band) without insertion of *p-SINE1-r2*. Lanes 6, 9, 10, 11 and 12 showed two bands (U/L). M = 1kb ladder used as MW marker.

with nested primers supported this evidence. The polymorphism occurring among them was created by the carrying of inserted DNA in the intron 10 of the *Waxy* gene.

Nucleotide sequence of *p-SINE1-r2*

The previous report indicated that each molecule of *p-SINE1-r2* consists of 120-125 bp. The upstream and downstream of both ends were flanked by direct repeat of a 14-bp sequence (Umeda *et al*., 1991). In this study, the length of this inserted DNA element, in all the rice accessions examined was 125 bp. The element *p-SINE1-r2* showed length polymorphism (150-153 bp), resulting from the insertions and/or deletions of mutations. In addition, these sequences had a T-rich pyrimidine tract at their defined 3' end and were flanked by direct repeats of a target sequence of mostly 14-bp long (Figure 2). The features of these elements in the rice samples agree with the previous studies (Motohashi *et al*., 1997). This suggests that genomes of rice used in this study contained *p-SINE1-r2* and are consistent with their nucleotide sequences and direct repeats when compared with the *p-SINE1-r2* sequences previously reported by Mochizuki *et al*. (1992) and Motohashi *et al*. (1997). Furthermore, the present result showed that *p-SINE1-r2* is present at the corresponding locus in both *O. sativa indica* and *japonica* types, suggesting that the *indic*a and *japonica* type strains cannot be distinguished from each other using this element. In contrast, *p-SINE1-r4* inserted in a corresponding locus can be used as DNA marker for classifying *indica* and *japonica* rice (Mochizuki, 1992). However, a variation occurred by the mutation in the *p-SINE1-r2* DNA sequences of wild rice strain with waxy endosperm that had been collected from northeastern Thailand. The variation was the deletion and/or insertion of the three bases (TAT). Based on this mutation, rice strains from southern Thailand were also classified into two groups (presence and absence) using the deletion and/or insertion of the three base, whereas, rice cultivars from northern and northeastern Thailand, and from Japan and Laos were present the three nucleotides.

Table 2 summarizes the results of nucleotide diversity for *p-SINE1-r2* sequence of cultivated rice (*O. sativa*) and wild rice (*O. nivara*). The DNA sequence data of these rice accessions were also classified as endosperm type data and geographical origin data for nucleotide diversity analysis. Within cultivated rice, nucleotide diversity was π = 0.05 and θ = 0.07 among 41 samples. The nucleotide diversity (per site) for the entire wild rice (*O. nivara*) sample was π = 0.07 and θ = 0.06. Based on the endosperm type data, *p-SINE1-r2* sequences of non-glutinous rice showed higher levels of

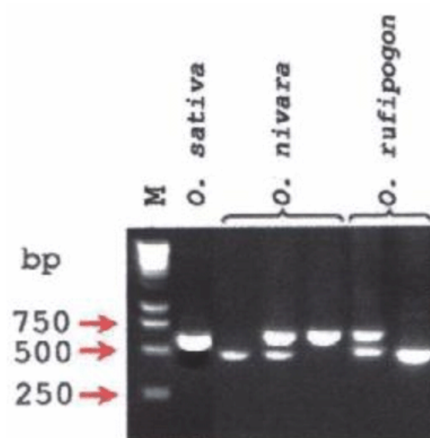

Figure 2. An example of agarose gel electrophoresis of amplified band by PCR using primer pairs (M6/R7). Species indicated in this panel represent the number of band patterns found in population of each species.

Table 2. Population variability parameters for nucleotide sequences of the *p-SINE-r2* in the 10th intron of the rice *Waxy* gene in cultivated and wild rice with AA genome.

	Sample size	Length(bp)	H	*S*	θ *	π **
Species (n=48)						
O. sativa	41	151-154	20	43	0.067 (0.02)	0.047 (0.011)
O. nivara	7	151-154	7	23	0.06 (0.03)	0.067 (0.011)
Endosperm type (n=48)						
Glutinous rice	31	151-154	20	29	0.05 (0.02)	0.03408 (0.007)
Nonglutinous rice	17	150-153	11	40	0.08 (0.03)	0.07205 (0.019)
Geographical origin (n=48)						
Japan	4	153	2	1	0.004 (0.004)	0.00327 (0.002)
China	2	153	2	36	0.24 (0.12)	0.24 (0.17)
Laos	5	153	5	16	0.05 (0.023)	0.044 (0.016)
Northern, Thailand	15	153-154	3	2	0.004 (0.003)	0.004 (0.001)
Northeastern, Thailand	15	153-154	14	23	0.048 (0.019)	0.043 (0.01)
Southern, Thailand	7	150	5	19	0.052 (0.025)	0.0415 (0.015)

Length, (excluding sites with alignment gaps); H, number of haplotypes; S, number of segregating sites. Standard deviation (S.D.) are shown in parentheses.
* $\theta = S / \sum_{i=1}^{n-1}(1/i)$, **where S is the total number of segregating sites, n is the number of nucleotide sequences.**
** **π = (Pi), nucleotide diversity.**

nucleotide diversity than that of glutinous rice. With respect to geographical origin, rice accessions from the northern areas have the lowest nucleotide diversity when compared to that of other regions of Thailand.

Phylogenetic tree

The analysis of data from the *p-SINE1-r2* sequences of the representative rice accessions for phylogenetic construction is shown in Figure 4. Based on the *p-SINE1-r2*, it is evident that rice samples used in this study agree with respect to their biogeographic pattern. Most rice accessions from each region of Thailand formed clusters, while rice accessions from other regions (Japan, China and Laos) are already members of clusters. Moreover, rice samples from northern Thailand, Japan and some from Laos, were closely related to cultivars from Japan. Indeed, *japonica* rice is found mostly in the northern region of Thailand. As such, this area could be recognized as the center of diversity of *japonica* rice (Prathepha and Baimai, 2004). In addition, most of the upland rice cultivars grown in the highlands and mountainous regions of northern Laos have been classified as *japonica* rice (Yamanaka *et al*., 2001). Taken together, those reports, including this study, demonstrate that, although the evolutionary event is still not known, there is a close relationship between the *japonica* rice of northern Thailand, Laos and China. Bautista *et al*. (2001) suggested that the geographic origin of *O. sativa japonica* appears to be in China. This consideration is also supported by the oldest archaeological evidence (ca. 7000-8000 years ago) of *japonica* rice cultivation found along the Yangtze River in China, pointing to China as the birthplace of *japonica* cultivars (Sato, 1996). In the case of *indica* rice, Sato (1996) suggested that the place *indica* rice cultivation began was in the flooded plains of large rivers, such as Mekong in Laos-Cambodia-Vietnam, Chaopraya in Thailand. Irawadi in Mynmar, and Ganga-Bramaptra River in India. However, future studies should be undertaken to determine the origin of *indica* rice, analyzing more accessions of rice cultivars and its progenitor (*O. rufipogon*).

The evolutionary and geographical origins of rice in the mainland of Southeast Asia have

```
         1111111111222222222233333333334444444444555555555566666666667777777777888888888899999999991111111111111111111111111111111111111111111111111111111111
         1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678900000000001111111111222222222233333333334444444444555555555
                                                                                                  012345678901234567890123456789012345678901234567890123456789
SQ8321   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10044  TTTGCCCC-TGAATTGAAAAATGCTCAAAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ5086   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCCCCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ7689   TTTGCCCCCTGAATTGAGAAATGCTCAGAGGTCTTCCGGCT-- AACACCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGAAATTATTAGGT-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ5085   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCCCCCCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10025  TTTGCCCC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10048  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10047  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGG-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10046  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10040  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGG-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10045  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGAAATTATTAGGG-CCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10034  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10022  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10043  TTTGCCCC-TGAATTGAGAAAAGATCAGAGGTCTTCCGGCG--AACACCACAAGGTGGTGGACTAGACGATCTGGGGTGAAAGCCTCCCCCCCTCTAATTATTTGATAATATTAGGG-CCTTCCCTAAAATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10029  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10019  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10026  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10023  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10030  TTTGCCCC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10027  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10033  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10028  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10017  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10018  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10016  TTTGCCGC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8318   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8322   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCCCCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8316   TTTGCCCC-TGAATTGAAAAATGCTCAAAGGTCTTCCGGCT---AACCCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCCCCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8323   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCCCCCCTTCTAATTATTTGAGATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8215   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACCCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCCCCCCCTCTAATTATTTGATATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8320   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACCCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCCCCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAAAATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8314   TTTGCCCC-TGAATTGAAAAATGCTCAGAGGTCTTCCGGCT---AACCCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCCCCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ8319   TTTGCCCC-TGAATTGAAAAATGCTCATAGGTCTTCCGGCT --AACCCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCTCCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10032  TTTGCCCC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10024  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10020  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10031  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAGACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGT-CCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ10021  TTTGCCAC-TGAATTGAGAAATGCTCAGAGGTCTTCCGGCT---AACTCCACAAGGTGGTGGACTAAACGATCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGATATTATTAGGTGCCTTCCCTAATATTCGAGTTTTTTTTTT-GCCACTGAATT
SQ5087   TTTGCCCC-TGAATTGAAAAATGCTCAGGGCTCTTTCGGCT--- AGCTCCACAAGGTGGTGGGCTAAATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGT-TCTTCCCTAAAATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10037  TTTGCCAC-TGAATCGAGAAATGCTCAGAGCTCTTCCGGCT---AGCTCTACAAGGTGGTGGGCTAGATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGT-TCTTCCCTAATATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10038  TTTGCCAC-TGAATCGAGAAATGCTCAGAGCTCTTCCGGCT---AGCTCTACAAGGTGGTGGGCTAGATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGT-TCTTCCCTAATATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10042  TTTGCCAC-TGAATCGAGAAATGCTCAGAGCTCTTCCGGCT---AGCTCTACAAGGTGGTGGGCTAGATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGT-TCTTCCCTAAAATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10036  TTTGCCAC-TGAATCGAGAAATGCTCAGAGCTCTTCCGGCT---AGCTCTACAAGGTGGTGGGCTAGATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGT-TCTTCCCTAATATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10039  TTTGCCAC-TGAATCGAGAAATGCTCAGAGCTCTTCCGGCT---AGCTCTACAAGGTGGTGGGCTAAATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGT-TCTTCCCTAATATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10035  TTTGCCCC-TGAATCGAGAAATGCTCAGAGCTCTTCCGGCT---AGCTCTACAAGGTGGTGGGCTAAATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGT-TCTTCCCTAATATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ10041  TTTGGCCC-TTAATCCAAAAATGGTCAAAGCTCTTCCGGCT--- AGCTCTACAAGGTGGTGGGCTAGATGACCTGGGTTCAAAGCCTCACCCCTTCTAATTATTTGA---TATTAGGG-TCTTCCCTAAAATTTGAGTTTTTTTTTT-GCCACTGAATT
SQ8317   TTTGCCCC-TGAATCGAAAAATGCTCAGAGCTCTTCCGGCT---AGCTCTACAAGGGGGGGGGCTAAATGACCTGGGTTCAAAGCCCCCCCCCTTCTAATTATTTGA---AATTAGGT-TCTTCCCTAAAATTTGAGTTTTTTTTTTTGCCACTGAATT
SQ10049  TTTTTCGC-CGAATTAACGAAAGATGATAGGTGCTCCGGCGGCCAGCTCAAGGAGGTGGTGGGATAGATCTGGGGGGTTAAAAGCCCCCCCCTTATAATTTATAA---TATTAGGG-CCTTCCCTAAAATTTGAGTTTTTTTTTT-GCCACTGAATT
            **  * ***     **  **   *  * *  *   * **         ****      *   **     *    *     *    *   *****        *         * *        *       *  *  * ****      * **                        *
```

Figure 3. Nucleotide sequences of *p-SINE1-r2* of the 48 rice accessions used in this study. Deletions or insertions are indicated by dashes. Polymorphic sites are indicated by asterisks. Sequence numbers refer to codes of rice accessions listed in Table 1.

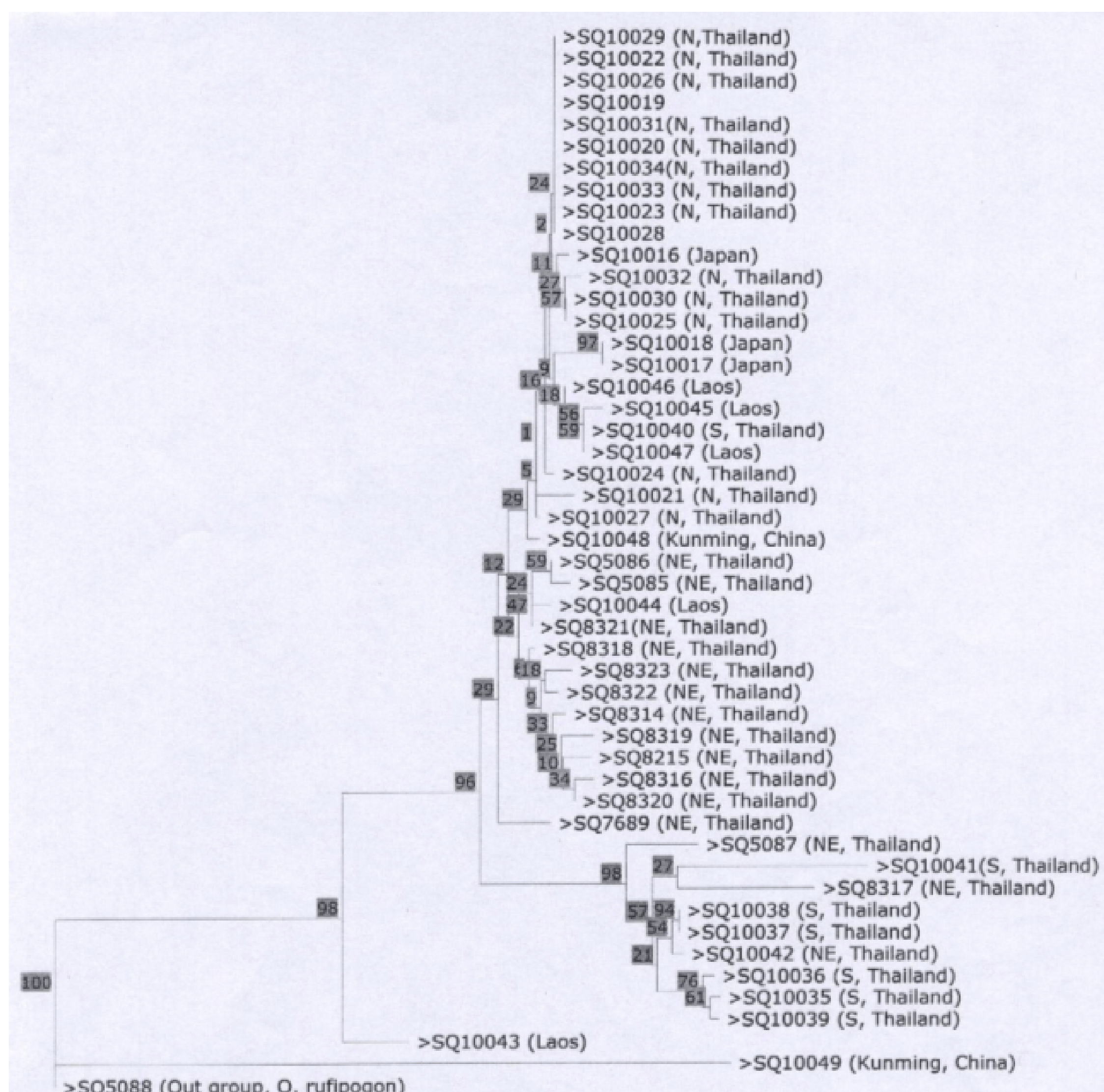

Figure 4. A phylogenetic tree of the 48 rice strains conducted by using NJ method.

remained obscure, the phylogenetic tree resulting from this study might provide a means of inferring the origin and dispersal of *p-SINE1-r2* that have evolved over course of domestication. It could also provide growing evidence on the molecular evolution of the rice *Waxy* gene.

Acknowledgment

The author would like to thank an anonymous reader for critical reading of the manuscript. This work was partly supported by the TRF/BIOTEC Special Program for Biodiversity Research and Training grant BRT R-245010.

References

Bautista, N.S., Solis, R., Kamijima, O. and Ishii, T. 2001. RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice. Genes Genet Syst 76: 71-79.

Cheng, C., Tsuchimoto, S., Ohtsubo, H. and Ohtsubo, E. 2002. Evolutionary relationships among rice

species with AA genome based on SINE insertion analysis. Genes Genet Syst 77: 323-334.

Cheng, C., Motohashi, S., Tsuchimoto, S., Fugita, Y., Ohtsubo, H. and Ohtsubo E. 2003. Phylogenetic origin of cultivated rice: based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20: 67-75.

Doyle, J.J. and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bull. 19: 11-15.

Ford, M.J. 2002. Applications of selective neutrality tests to molecular ecology. Molecular Ecology 11: 1245-12162.

Hirano, H.Y., Mochizuki, K., Umeda, M., Ohtsubo, H., Ohtsubo, E. and Sano, Y. 1994. Retrotransposition o a plant SINE into the wx locus during evolution of rice. J Mol Evol 38: 132-137.

Kanno, A., Watanabe, N., Nakamura, I. and Hirai, A. 1993. Variations in chloroplast DNA from rice (*Oryza sativa*): differences between deletions mediated by short direct-repeat sequences within a single species. Theor Appl Genet 86: 579-584.

Maraia, R.J. 1995. The impact of short interspersed elements (SINEs) on the host genome. Springer, New York, NY.

Mochizuki, K., Umeda, M., Ohtsubo, H. and Ohtsubo, E. 1992. Characterization of a plant SINE, *p-SINE1*, in rice genomes. Jpn J Genet 67: 155-166.

Mochizuki, K., Ohtsubo, H., Hirano, H., Sano, Y. and Ohtsubo, E. 1993. Classification and relationships of rice strains with AA genome by identification of transposable elements at nine loci. Jpn J Genet 68: 205-217.

Motohashi, R., Mochizuki, K., Ohtsubo, H. and Ohtsubo, E. 1997. Structures and distribution of *p-SINE1* members in rice genome. Theor Appl Genet 95: 359-368.

Nei, M. 1987. Molecular evolution genetics. Columbia University Press, New York.

Okagaki, R.J. and Wessler, S.R. 1988. Comparison of non-mutant and mutant waxy gene in rice and maize. Genetics 120: 1137-1143.

Prathepha, P. 2003. Distribution of *p-SINE1-r2* at the *Wx* locus in cultivated and wild rice in Thailand. International Rice Research Notes 28.1: 26-27.

Prathepha, P. and Baimai, V. 2004. Variation of *Wx* microsatellite allele, waxy allele distribution and differentiation of chloroplast DNA in a collection of Thai rice (*Oryza sativa* L.). Euphytica 140: 231-237.

Rozas, J. and Rozas, R. 1999. DnaSP: an integrated program for molecular population genetics and molecular evolution analysis, version 3.5, Bioinformatics 15: 174-175.

Sato, Y.I. 1996. Origin and dissemination of cultivated rice in the eastern Asia. **In**: Interdisciplinary perspectives on the origins of the Japanese, pp. 143-153. International Research Center for Japanese Studies, Tokyo.

Saitou, N. and Nei, M. 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.

Shedlock, A.M. and Okada, N. 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22: 148-160.

Umeda, M., Ohtsubo, H. and Ohtsubo, E. 1991. Diversification of the rice *Waxy* gene by insertion of mobile DNA elements into introns. Jpn J Genet 66: 569-586.

Van de Peer, Y. and De Wachter, R. 1997. Construction of evolutionary distance trees with TREECON for Windows: accounting for variation in nucleotide substitution rate among sites. Comput Applic Biosci 13: 227-230.

Watterson, C.A. 1975. On the number of segregating sites in genetical models without recombination. Theor Pop Biol 7: 256-276.

Yamanaka, S., Nakamura, I., Nakai, H. and Sato, Y.-I. (No date). Annual-perennial and *indica-japonica* differentiations related to retrotransposon polymorphism at *Waxy* locus in wild relatives of rice. Online]. Available: http://www.carleton.ca/~bgordon/Rice/papers/yama20.htm [2004, May 23].

Yamanaka, S., Ishikawa, R., Fukuta, Y., Nakamura, I., Sato, T. and Sato, Y.I. 2001. Evaluation of upland rice genetic resources in northern Laos. Rice Genetic Newsletter 18: 9-10.