Kasetsart J. (Nat. Sci.) 40 : 306 - 313 (2006)

Chloroplast Diversity and Phylogeny in Wild and Cultivated Rice (*Oryza* spp.)

Siriporn Chuayjaeng[1]* and Hugo Volkaert[1, 2]

ABSTRACT

Genetic diversity was studied in the chloroplast genome of wild and cultivated rice, especially in species belonging to the A genome group, to clarify the phylogenetic relationships in *Oryza*. Specific DNA fragments were PCR-amplified from 140 accessions of cultivated and wild rice and investigated by restriction analysis and SSCP (PCR-RF-SSCP or PRS). After the SSCP patterns were scored, amplified fragments from representative accessions of each observed SSCP pattern were sequenced. Haplotypes of cultivated rice and closely related wild relatives (A genome) were more similar than those of distantly related species (B, C/CD, E and G genomes). The polymorphisms were due to insertion/deletion (indels) and nucleotide substitutions. Substitutions and indels at ORF100, and rpoA locus could separate the Asian cultivated rice *indica* type from *japonica* type. Additional substitutions and indels at all four cpDNA loci could separate the rice species with the AA genome from those belonging to other genomes (but these mutations were not congruent with *indica/japonica* division). Within the AA genome four groups could be separated, Asian wild rice with Asian cultivated rice, *O. glumaepatula* group, African wild rice with African cultivated rice (*O. glaberrima*) and *O. meridionalis* group.

Key words: chloroplast, diversity, phylogeny and *Oryza*

INTRODUCTION

Rice is one of the most widely cultivated cereal crops in tropical and subtropical regions of the world. The rice genus *Oryza* consists of about 23 species distributed in Asia, Africa, Australia and the Americas (Vaughan *et al*., 2003). All species have been classified into six diploid genome types (AA, BB, CC, EE, FF and GG) and four tetraploid genome types (BBCC, CCDD, HHJJ and HHKK) (Ge *et al*., 1999; Vaughan *et al*., 2003) based on chromosome pairing in F1 hybrids between different taxa. The AA genome group includes the cultivated rice species, Asian (*O. sativa*) and African cultivated rice (*O. glaberrima*), while the other species are wild rice (Ge *et al*., 1999). The diversity and phylogeny of rice have been investigated using the nuclear, chloroplast, and mitochondrial genomes but uncertainties concerning the origin of cultivars still remain. Several techniques have been used such as Restriction Fragment Length Polymorphism (RFLP) (Fukuoka *et al*., 2003), Amplification Fragment Length Polymorphism (AFLP) (Prashanth *et al*., 2002), Random Amplified Polymorphic DNA (RAPD) (Buso *et al*., 2001), microsatellites (Nishikawa *et al*., 2005) and single strand conformation polymorphism (SSCP) (Sato

[1] Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Sean Campus, Nakhon Pathom 73140, Thailand.
[2] NSTDA/BIOTEC, 131 Thailand Science Park, Phaholyothin Road, Klong 1, KhlongLuang, Pathumthani 12120, Thailand.
* Corresponding, e-mail: silk1071@yahoo.com

Received date : 30/09/05 Accepted date : 01/03/06

and Nishio, 2003; Shirasawa *et al*., 2004). The chloroplast genome is well suited for evolutionary and phylogenetic study. It is considered to be conservative, fairly slowly changing in nucleotide sequence without recombination and it behaves as a haploid, maternally inherited molecule, present in multiple copies per cell (Fujii *et al*., 2001).

This research was aimed to study the genetic diversity at specific cpDNA loci in wild and cultivated rice, especially in the AA genome group, by analyzing nucleotide sequence evolution and inferring the cpDNA phylogeny of *Oryza* species. Restriction enzyme digestion of PCR amplified fragments followed by SSCP electrophoresis, a simple and rapid method that could detect DNA sequence variation in a large number of individuals, and DNA sequencing were used.

MATERIALS AND METHODS

Plant materials

Total of 140 accessions representing different *Oryza* species (Table 1) were obtained from the germplasm collection maintained by the International Rice Research Institute (IRRI, Los Banos, Philippines). One accession of *O. granulata* was collected in Kamphaeng Phet Province,Thailand. Seeds were germinated on filter paper and the seedlings were used for DNA extraction.

DNA extraction.

Genomic DNA was extracted from fresh young leaves of 7-14 days old rice seedlings using the Qiagen DNeasy plant DNA extraction kit (Qiagen, Germany) according to the manufacturer's instructions. The quantity and quality of the extracted DNA were checked by electrophoresis in 0.8% agarose gel.

Selection of primers.

Two chloroplast loci, rpl16 intron and rpoA intergenic region, have been previously developed (Volkaert and Vanavichit, 2001). Primer

Table 1 List of *Oryza* species used for studies and chloroplast haplotypes.

Oryza species	Genome	Number of samples	Number of chloroplast haplotypes
O. sativa	AA	101	10
O. rufipogon	AA	6	3
O. nivara	AA	5	3
O. glaberrima	AA	2	1
O. barthii	AA	2	1
O. glumaepatula	AA	5	5
O. meridionalis	AA	2	2
O. spontanea	AA	2	2
O. punctata	BB	1	1
O. eichingeri	CC/CCCC	2	1
O. officinalis	CC	2	2
O. rhizomatis	CC	2	2
O. alta	CCDD	2	1
O. grandiglumis	CCDD	2	1
O. latifolia	CCDD	2	2
O. australiensis	EE	1	1
O. granulata	*GG*	1	1

sequences for the ORF100 locus, which identified two major types (*indica* and *japonica*) of cpDNA in Asian cultivated rice, were taken from Sun *et al*. (2002). In addition, the ndhA intron primers for PCR and sequencing were designed by comparing three available complete rice cpDNA sequences: AP006728 (*O. nivara*), AY522329 (*O. sativa* indica cultivars group) and NC_001320 (*O. sativa* japonica cultivar group). Details of all primers are presented in Table 2.

PCR-RF-SSCP analysis.

Samples were amplified by using specific primer sets for chloroplast loci. Each reaction was contained in 15 µl volume consisting 20 ng of template DNA, 10 pmol of each primer, 200 µM of each dNTPs (Promega), 1X PCR buffer, 2mM $MgCl_2$ and 0.5U of Taq DNA polymerase (Fermentas). Amplification was performed on a T1 Thermocycler (Biometra) or on a PTC-100 (MJ Research). The PCR cycling was initiated by a 3 min at 94°C denaturation followed by 35 cycles of 45 sec at 94°C, 45 sec at 48-53°C, 1-1.30 min at 72°C and a final extension at 72°C for 5 min. The PCR products were electrophoresed on agarose gel to assess the amplification and cpDNA size.

After PCR amplification, products were digested with a restriction enzyme (Table 2) under conditions recommended by the supplier, except for the ORF 100 locus, whose size was small enough without digestion.

The DNA fragments were separated by electrophoresis in non-denaturing polyacrylamide gels (Sequagel MD, National Diagnostics, U.S.A.) using modification the Single Strand DNA Polymorphism approach (Sato and Nishio, 2003; Shirasawa, *et al*., 2004). Before sample loading, the products were mixed with loading buffer and heat denatured at 95°C for 10 min, followed by quick cooling on ice. The fragments were run in 0.6X TBE buffer with constant wattage of 8 watt in a 4°C refrigerator for 14-16 hrs. After the electrophoresis the gels were stained with silver.

DNA cloning and sequencing.

The PCR fragments from selected accessions of rpl16 and ORF100 were purified with MinElute PCR Purification Kit (Qiagen) using conditions recommended by the supplier and then ligated into pGem-T vector (Promega) and

Table 2 Primer sets for PCR and sequencing, restriction enzyme and fragment of PCR amplified.

Locus		Sequence	Restriction enzyme	Fragment length (bp)	# SSCP pattern	# Nucleotide haplotypes
PCR primers						
ORF100	F	5' taaagtaagtggacctgactc 3'	-	394-465	9	9
	R	5' cagccgaggtcgtggtaaatc 3'				
rpl16	F	5' gctatgcttagtgtgtgact 3'	*Eco*RI	1099-1138	12	12
	R	5' cattcttcctctatgttgttt 3'				
ndhA	F	5' cattcttcctctatgttgttt 3'	*Nla*III	1063-1092	10	10
	R	5' tgtgcttcaactatatcaactgt 3'				
rpoA	F	5' ccatttcgtcgtccagtagc 3'	*Hin*dIII	1306-1313	10	11
	R	5' cagtggaagtgtgttgaatc 3'				
Sequencing Primer						
rpoA	F	5' tttcttggaaacgagtacc 3'				
	R	5' gctgtatccatgcctgttc 3'				
ndhA	F	5' ccatcaaaagttagtgagg 3				
	R	5' gagttgggtattagttcatg 3'				

cloned into *Escherichia coli* according to the suppliers' instructions. Plasmids were extracted from transformed cells with QIAprep Spin Miniprep kit (Qiagen) and sent for sequencing at Macrogen, Inc. (Seoul, Korea). Alternatively, ndhA and rpoA PCR amplified DNA fragments were sent for direct sequencing using specific internal sequencing primers (Table 2). As for PCR amplification for direct sequencing, the same condition were used except that primer and dNTP concentration were reduced by 10% and the volume was increased to 50 µl.

Phylogenetic analysis

For each of the amplified fragments, the nucleotide sequences were analyzed. Sequences were aligned and manually edited using Genedoc software. Insertion – deletions longer than a single nucleotide were recorded such that each event was counted as a single mutation. Phylogenetic analysis was done using programs in the PHYLIP3.6 software package (Felsenstein, 1989). To evaluate the support for branches, 100 bootstrap replications were constructed using the SEQBOOT program. Maximum parsimony (MP) analysis was done using DNAPARS, and maximum likelihood trees were calculated using the DNAML program using the Kimura 2-parameter model for base substitutions. Genetic distances were calculated using the program DNADIST with Kimura 2-parameter model. The neighbour joining procedure in the NEIGHBOR program was used to obtain phylogenetic trees from the genetic distance matrices. Consensus trees for each procedure were generated using the CONSENSE program. The phylogenies were viewed using TREEVIEW (Page, 1996). The resulting trees were rooted using *O. granulata* (GG genome) from Thailand as the outgroup.

RESULTS AND DISCUSSION

Genetic diversity

140 *Oryza* accessions were assayed using PCR-RF-SSCP for the alleles at four chloroplast loci (ORF100, rpl16, ndhA and rpoA). The primer sets successfully amplified fragments from all *Oryza* samples. On agarose gel, different alleles were not visible except for ORF100 where 2 alleles could be distinguished (Figure1). The amplified DNA fragments were digested with restriction enzyme (Table 2) except ORF100, then denatured and electrophoresed on non-denaturing polyacrylamide gel (SSCP). The obtained bands corresponding to the single strand DNAs were scored and accessions representing each observed haplotype were analysed further by sequencing. The indels (deletion/insertion) and base substitution were found from sequencing.

The ORF100 primers amplified a PCR fragment of 394-465 bp. A deletion of 69 bp in ORF100 region divided the cultivated rice into 2 groups. Most *Indica* type cultivars possessed a deletion of 69 bp, while most *japonica* type did not have the deletion as previously reported (Kanno *et al.*, 1993; Sun *et al.*, 2002). However, a few of *indica* cultivars showed the *japonica* type and some *japonica* showed the *indica* type. In wild rice the deletion was found only in some Asian accessions belonging to *O. rufipogon* and *O. nivara*. Other species in AA genome as well as the wild rice in other genomes (BB, CC, EE, GG, BBCC, CCDD,) did not have the deletion. Thus the deletion was a derived character state unique to some accession within *O. sativa, O. nivara* and *O. rufipogon*. Three single nucleotide polymorphisms (SNP), A/G at positions 117, 360 and T/G at position 146, could separate the rice AA genome from other genomes. SSCP assay revealed that four patterns and four sequences haplotype were found at the ORF100 locus in AA genome. The primers of ndhA locus amplified the intron in the *ndhA* gene. The PCR fragment ranged in sizes between 1063-1092 bp. Both the SSCP pattern and sequencing revealed 10 haplotypes. Nucleotide polymorphisms among alleles were found at 40 positions, 12 indels and 28 SNPs. Whithin the AA genome, 4 haplotypes were found

due to 2 SNPs and 3 indels. The A/T polymorphism at position 809 separated the African species (*O. glaberrima* and *O. barthii*) from other species. Three SNPs (T/C at positions 121, 230 and G/A of 438 respectively) could separate *Oryza* AA genome from other genomes. The primers for the rpoA locus amplified a PCR fragment of 1306-1313 bp containing almost complete *rpoA* gene and the intergenic region between *rpoA* and *petD*. SSCP analysis revealed 10 patterns while sequencing indicated 11 nucleotide haplotypes. Nucleotide polymorphisms were found at 35 positions consisting of 1 indel and 34 SNPs. Within the AA genome only 5 SSCP patterns and 6 nucleotide haplotypes were found. The nucleotide polymorphisms were due to 1 insertion and 5 SNPs. The SNPs G/T, C/A and C/T at positions 609, 791 and 1140 could separate rice AA genome from other genomes.

The primers for rpl16 locus amplified the intron in the *rpl16* gene. The PCR fragments ranged in size between 1099-1138 bp. The SSCP analysis and sequencing showed 12 haplotypes. Nucleotide polymorphisms were found at 57

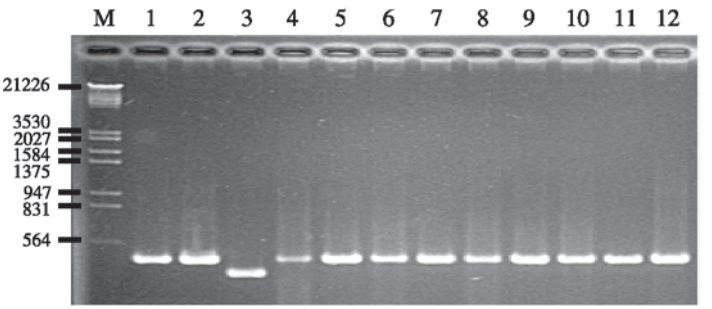

Figure 1 ORF100 PCR fragments separated on agarose gel revealed the unique deletion mutation in some rice cultivars. Lane 1 to 9 are *Oryza* species with AA genome. Lane 1and 2: *O. sativa japonica*; lane 3 and 4: *O. sativa indica*; lane 5 and 6:*O. nivara*; lane 7and 8:*O. rufipogon*; lane 9:*O. glaberrima*; lane 10*: O. punctata* (BBCC genome); lane11: *O. rhizomatis* (CC genome) and lane12: *O. australiensis* (EE genome). Marker is Lamda/*Hin*dIII+*Eco*RI.

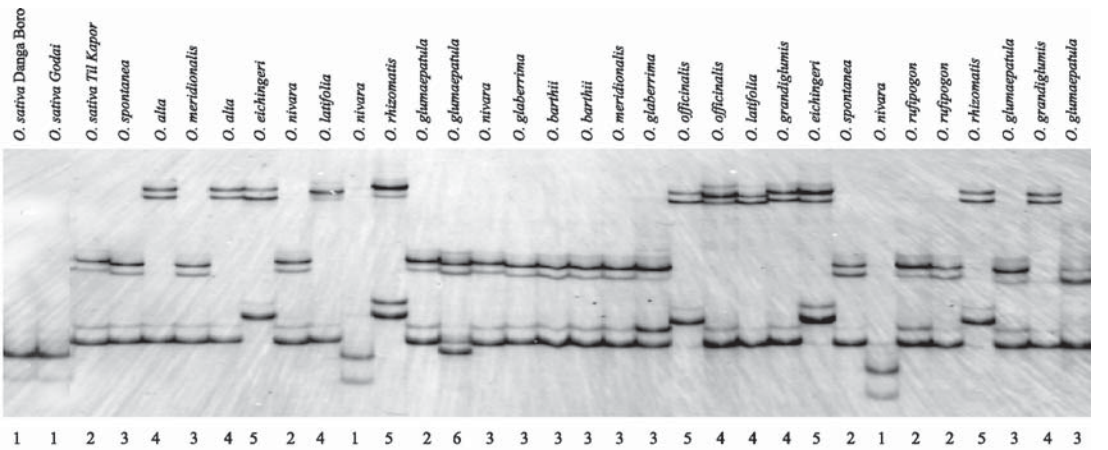

Figure 2 Analysis of DNA fragments corresponding to the ORF100 locus separated by SSCP polyacrylamide gel. Different banding patterns were identified (pattern 1 to 6). Three additional patterns (*O. granulata, O. australiensis and O. punctata*) are not shown.

positions, 28 indels and 29 SNPs. Within the AA genome, 6 haplotypes were found due to 4 SNPs and 8 indels. The indels at position 33, 280 and SNPs, A/G at position 285 and T/C at positions 660 and 954, could separate rice species with the AA genome from other genomes.

Phylogenetic analysis

Sequences from the four chloroplast loci were combined and analyzed as a single block of sizes 3900-3994 bp. Total of 154 nucleotide polymorphisms were found consisting of 48 indels and 106 SNPs. Total of 28 haplotypes were obtained from the 140 accessions. The results were in close agreement with previous studies of *Oryza* taxonomy (Ge *et al*., 1999; Vaughan *et al*, 2003). The analysis using genetic distances, maximum likelihood and maximum parsimony all resulted in very similar trees. Therefore, only the maximum parsimony results are presented. The most parsimonious tree of *Oryza* species based on the 4 cpDNA loci combined is shown in Figure 3. All *Oryza* species were clearly distinguished. The species belonging to the *Oryza* AA genome were separated form those belonging to the other genomes (BB, CC/CCDD, EE and GG genome). All rice species within CC/DD genome formed a monophyletic group. The BB genome (*O. punctata*) was closely related to the AA genome (Ge *et al*., 1999; Nishikawa *et al*., 2005). Within the AA genome group, phylogenetic analysis could separate 4 groups, though bootstrap values were generally low. Asian wild rice *O. rufipogon* / *O. nivara* clustered in the same group with almost all Asian cultivated rice (*O. sativa*). Most of the *O. glumaepatula* from South and Central America clustered together. The African rice group (*O. barthii* and *O. glaberrima*) was sister to three

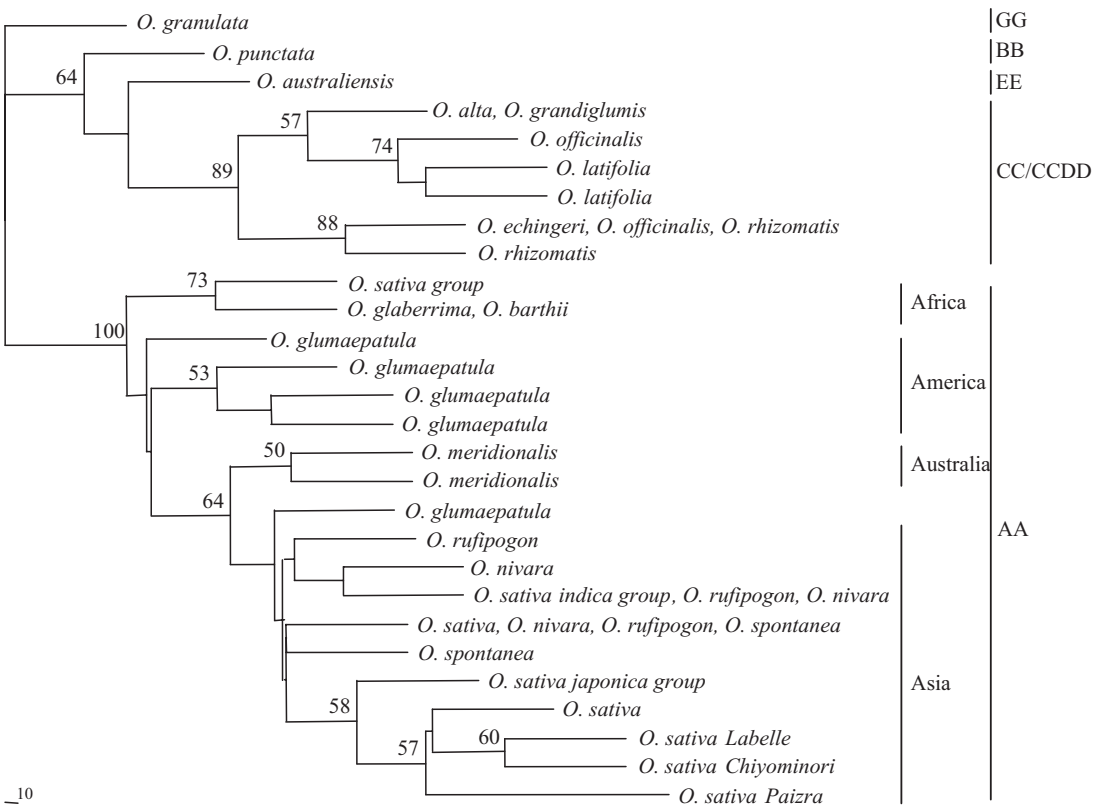

Figure 3 Phylogenetic parsimonious tree of *Oryza* species based on chloroplast loci. Bootstrap values are shown above branches.

accessions of Asian cultivars. The Australian *O. meridionalis* accessions formed a monophyletic group. One *O. glumaepatula* accession was sister to the Asian group, indicating possibly more recent admixture of Asian rice into the Americas as also speculated by Vaughan *et al*. (2003) and Nishikawa *et al.* (2005). Each *indica* and *japonica* cultivars grouped with some O. *nivara* /*O. rufipogon* accessions. Thus the chloroplast diversity existed in the wild ancestors before domestication of the two cultivar groups. The result supported the diphyletic hypothesis for the origin of *indica* and *japonica*, which proposed that *japonica* and *indica* originated from different areas (Sun *et al*., 1978; Lu *et al*., 2002; Chen *et al*., 2003).

CONCLUSIONS

The newly designed specific primer sets (rpl16, ndhA, and rpoA) were used successfully to amplify cpDNA fragments. Base substitutions and indels at ORF100 and rpoA locus could separate the Asian cultivated rice of *indica* type from *japonica* type. In all four cpDNA loci, there were base substitutions and indels that could separate the rice species with the AA genome from those with other genomes. The phylogenetic tree clearly separated the rice species with the AA genome from those with other genomes. The AA genome could be divided into four groups, Asian wild rice with Asian cultivated rice, *O. glumaepatula* group, African wild rice with African cultivated rice (*O. glaberrima*) and *O. meridionalis* group. The origin of cultivars still remained uncertainties concerning. This result supported to the diphyletic hypothesis for the origin of *indica* and *japonica* and proposed that *japonica* and *indica* originated in different places.

ACKNOWLEDGEMENTS

The research is supported by the Center for Agricultural Biotechnology through the fund from Subproject Graduate Study and Research in Agricultural Biotechnology under Higher Education Development Project, the Ministry of Education and the Graduate School, Kasetsart University. Thanks are also due to International Rice Research Institute (IRRI, Los Banos, Philippines) for supplying seeds.

LITERATURE CITED

Buso, G.C., P.N. Rangel and M.E. Ferreira. 2001. Analysis of random and specific sequences of nuclear and cytoplasmic DNA in diploid and tetraploid American wild rice species (*Oryza* spp.). **Genome** 44: 476-494.

Chen, C., R. Motohashi, S. Tsuchimoto, Y. Fukuta, H. Ohtsubo and E. Ohtsubo. 2003. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. **Mol. Biol. Evol**. 20: 67-75.

Felsenstein, J. 1989. PHYLIP – Phylogeny Inference Package (version 3.2). **Cladistics** 5: 164-166

Fujii, N., N. Tomaru, K. Okuyama, T. Koike, T. Mikami and K. Ueda. 2001. Chloroplast DNA phylogeography of *Fagus crenata* (Fagaceae) in Japan. **Plant Syst. Evol**. 232: 21-33.

Fukuoka, S., N.V. Alpatyeva, K. Ebana, N.T. Luu and T. Nagamine. 2003. Analysis of Vietnamese rice germplasm provides an insight into *japonica* rice differentiation. **Plant Breeding** 122: 495-502.

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

Kanno. A. and A. Hirai. 1993. A transcription map of the chloroplast genome from rice (*Oryza sativa*). **Curr. Genet**. 23: 166-174.

Lu, B.R, K.L. Zheng, H.R. Qian and J.Y. Zhuang. 2002. Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. **Theor. Appl. Genet**. 106: 101-106.

Nishikawa, T., Vaughan, D.A. and Kadowaki, K.

2005. Phylogenetic analysis of *Oryza* species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. **Theor. Appl. Genet.** 110: 696-705.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. **Computer Applications in the BioSciences** 12: 357-358

Prashanth, S.R., M. Parani, B.P. Mohanty, V. Talame, R. Tuberosa anf A. Parida. 2002. Genetic diversity in cultivars and landraces of *Oryza sativa* subsp. *indica* as revealed by AFLP markers. **Genome** 45:451-459.

Sato, Y. and T. Nishio. 2003. Mutation detection in rice waxy mutants by PCR-RF-SSCP. **Theor. Appl. Genet**. 107: 560-567.

Shirasawa, K., L. Nonna, S. Kishitani and T. Nishio. 2004. Single nucleotide polymorphisms in randomly selected genes among japonica rice (*Oryza sativa* L.) varieties identified by PCR-RF-SSCP. **DNA Res**. 11: 275-83.

Sun, C.Q., X.K. Wang and K. Doi. 2002. Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). **Theor. Appl. Genet**. 104: 1335-1345.

Vaughan, D.A., H. Morishima and K. Kadowaki. 2003. Diversity in the *Oryza* genus. **Current Opinion in Plant Biology** 6: 1-8.

Volkaert, H. and A. Vanavichit. 2001. Genetic diversity in wild and cultivated rice (*Oryza* spp.) at specific DNA loci. **Final Report: Project code BT-B-07-RM-B8-4302**. National Sceince and Technology Development Agency (NSTDA).