Asian Journal of Conservation Biology, December 2013. Vol. 2 No. 2, pp. 144–151 AJCB: FP0029
ISSN 2278-7666 ©*TCRP* 2013

Genetic relatedness of genus *Oryza* from Eastern Himalayan region as revealed by chloroplast *matK* gene

Doris Zodinpuii, Souvik Ghatak, Subhajit Mukherjee & Nachimuthu Senthil Kumar*
Department of Biotechnology, Mizoram University (A Central University), Aizawl - 796004, Mizoram, India

(Accepted November 11, 2013)

ABSTRACT

Phylogenetic relationship was studied in wild and cultivated rice using the chloroplast *matK* gene. The aligned sequence fragments were 826bp in length with 7.02% variable and 4.47% phylogenetically informative sites and the estimated Transition/Transversion bias (R) was 1.97. Seven hundred and two characters were constant, 74 variable characters were parsimony-uninformative and 50 were parsimony–informative. Haplotypes of Mizoram rice and wild relatives (A genome) were more similar than those of distantly related species (B, C/CD, E and G genomes). It further revealed that the EE genome species is most closely related to the CC genome and CCDD genomes. The BBCC genome species had different origins, and their maternal parents had either the BB or CC genome. An additional genome type, HHKK was recognized in *O. coarctata* and *O. schlechteri*. Within the AA genome the African, *O. glaberrima* and *O. longistaminatea* and American, *O. glumipatula* and *O. barthii* were closer to the Indian *Oryza* species, *O. nivara* and *O. rufipogon*. The unknown genome *O. malampuzhaensis* from India is closer to BB and BBCC genome containing respectively *O. punctata* from Cameroon and *O. minuta* from Philippines. CpG rich *matK* sequences were rich in GG and FF genotypes, whereas CpA rich sequences belonged to BB and BBCC related genomes variety.

Key words: *Oryza* species, *matK* gene, phylogeny, genotype, divergence, eastern himalayas

INTRODUCTION

Rice is one of the most ancient and widely cultivated cereal crops from tropical to temperate regions and has evolved into a standard food for more than half of the world's population. The rice genus, *Oryza*, consists of approximately 26 species distributed in Asia, Africa, Australia and the Americas. 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) based on chromosome pairing in F1 hybrids between different taxa (Ge *et al*., 1999; Vaughan *et al*., 2003*, www.gramene.org*). The most common AA genome group includes the cultivated rice species of Asian (*O. sativa*), African (*O. glaberrima*), Australian (*O. meridionalis*) origins, while the other types belong to wild species (Ge *et al*., 1999). Rice plants are a suitable model for understanding the genetic mechanism of crop domestication and improvement because of its diploid nature and small genome. The diversity and phylogeny of rice species have been investigated using the nuclear, chloroplast and mitochondrial genomes using several techniques (Fukuoka *et al*., 2003; Prashanth *et al*., 2002; Buso *et al*., 2001; Nishikawa *et al*., 2005; Shirasawa *et al*., 2004), but the origin of many cultivars still remains unsolved.

As in other parts of world, rice is the staple food crop of hilly North Eastern India (NEI). It is estimated that the total rice production of NEI region is approximately 5.50 million tonnes with average productivity of 1.57 tonnes/ha, which is much lower than the national average of 2.08 tonnes/ha (Pattanayak *et al*., 2006). Shifting or *jhum* cultivation is followed by the farmers in the biodiversity rich hilly state of Mizoram under the Eastern Himalayan range (WWF-US 2005), Moreover, the folded structure of the Mizoram ranges are at the junction of the moving Indian and Burmese tectonic plates (DesiKachar, 1974).Traditional rice varieties and the wild rice species are being lost through genetic erosion. Modern and hybrid rice varieties are more grown and no longer are the traditional varieties being given importance. In this region, few morphologically different varieties of rice have been cultivated for a few hundred years and, surprisingly, not much information is available on their ecotype genetic diversity. It is very important to understand the evolution of the rice varieties that are planted and grown, in order to have a true appreciation for this dietary staple. An understanding of the genotype and the genetic variations will be useful for the conservation and future crop improvement needs of these traditional varieties and related wild genera to cope with the many biotic and abiotic stresses that are challenging rice production for both current and future generations.

The chloroplast genome is a useful subject for evolutionary and phylogenetic study as it is mostly conserved, has without recombination, haploid, maternally inherited, and present in multiple copies per cell (Fujii *et al*., 2001). The *matK* gene has been used effectively for phylogenetic studies at different taxonomic levels in different plant groups (Hilu and Liang, 1997; Olmstead and Palmer 1994) including *Oryza* (Ge *et al*., 1999; Hilu *et al*., 1999). In this paper, we are particularly interested in revealing the phylogenetic relationship between *Oryza* from the Eastern Himalayan range (Mizoram) and *Oryza* species from elsewhere in the world.

*Corresponding Author's E-mail: nskmzu@gmail.com

MATERIALS AND METHODS

Collection of samples

Six seed samples (MZURV1 – MZURV6) were collected from Indian Council of Agricultural Research, Kolasib, Mizoram, India. Two samples (MZURV 7–MZURV8) were collected from local farmers of Kawlkulh village, Aizawl District, Mizoram (Figure 1). The samples were stored in a dry moist free container at room temperature.

Isolation of Genomic DNA

200 mg of de-husked rice seed was placed in a 1.5 ml centrifuge tube, with 700 µl of extraction buffer (15% sucrose, 50 mM Tris-HCl-pH 8.0, 10% SDS, 50 mM EDTA-pH8.0, and 250mM NaCl) added, homogenized using a micro-pestle and incubated in water bath at 37˚C for 1 hour. After incubation, 400 µl of CTAB buffer (20 mM Tris-HCl -pH 8.0, 10 mM EDTA-pH 8.0, 10% CTAB, 5% PVP) and 10 µl of proteinase k (10 mg/ml) were added, incubated at 65˚C for 1 h and centrifuged at 12,000Xg for 10 minutes. The supernatant was transferred into a fresh tube and treated with equal volume Phenol: Chloroform: Isoamyl (PCI), mixed gently by inverting the tube for 2 to 3 minutes, centrifuged at 10,000Xg for 10 minutes, and the supernatant transferred into a fresh tube. PCI treatment was repeated, 20 mg/ml (5 µl) of RNase was added, and the tube incubated at 37˚C for 30 min. PCI treatment was done for a third time, the tube centrifuged at 10,000Xg for 10 min and the supernatant transferred into a fresh tube. 400 µl of ice cold isopropanol with 1/10 volume 3 M Sodium Acetate (pH 5.2) was added and the tube stored at -20˚C overnight for DNA precipitation followed by Centrifugation at 12,000Xg for 10 min and the supernatant decanted. The pellet was washed with 70% ethanol and centrifuged at 12,000Xg for 10 min. The supernatant was poured off without dislodging the pellet and was air dried properly. The pellet was dissolved with 40 µl of TE buffer (10 mM Tris-pH 8.0, 1 mM EDTA-pH 8.0).

PCR amplification of matK gene

Gradient PCR was performed using eight samples in order to determine the optimal annealing temperatures of the primers *matK*22F (5´CGATCTATTCATTCAATA TTC3´) and *matK*22R (5´- TCTAGCACACGAAAG-TCGAAGT -3´) (Yu *et al.,* 2011). The 25 µl of PCR reaction mixture consisted of 1X Taq buffer, 250 µmM dNTPs, and 4.5 pmol each primer, 1U Taq DNA polymerase, and 50-100 ng template DNA. Thermal cycling conditions were as follows: 94˚C for 3 min, followed by 35 cycles of 94˚C for 40 sec, 49.5˚C for 40 sec, 72˚C for 1 min, and a final extension at 72˚C for 7 min. The PCR products were verified by electrophoresis in 1.5% agarose gels stained with ethidium bromide. PCR was repeated under the same conditions for those samples that had no band or weak bands. PCR products were sequenced by Sanger method at GCC Biotech, Kolkata, India.

Sequence retrieval and local sequence alignment

The partial sequences of *matK* of Mizoram rice varieties have been submitted to NCBI GenBank (Table 1). 27 *Oryza matK* sequences from different geographical regions, belonging to different genotypes, were retrieved from GenBank (Table 1). Multiple sequence alignment of all these sequences (826 bp DNA; 275 aminoacid) with our MZURV sequences were first aligned using ClustalX 2.0.11 (Larkin *et al*, 2007). The variability of the aligned sequences was evaluated using the sliding window method in DNAsp ver. 4.5 (Rozas &Sanchez-DelBarrio 2003). The window length was set to 600 bp, the typical length of DNA barcodes. The step size was set to 50 for relatively accurate positioning of variable regions. We only considered regions with a number of polymorphic sites. The regions were identified according to the original annotations, then extracted and compared among the genera after precise alignments. Regions were excluded from further consideration if they were present in fewer than three genera. Nucleotide and Amino acid divergence (O'Donnell, 1992) and genetic distances (Kimura, 1993) were calculated by MEGA 5.05. The variable substitution rate over site was calculated by the formula:

$$H_i = \left(\sum_{i=1}^{4} P_j log_2 P_j\right)$$

The variability at site 'i' is measured by the entropy (or information), where 'j' = 1, 2, 3, 4 corresponding to nucleotide A, C, G and T, and 'pj' is the proportion of nucleotide 'j' at site 'i'. If all nucleotides at site 'i' are identical, then 'Hi' = 0. Substitutions will lead to polymorphic sites at which the 'H' value will be larger than 0. It is 'H_i' that is plotted over sites (Xia, 2000, DAMBE 5.327).

The transition and transversion ratio was calculated by MEGA 5.05 (Kimura, 1993) and the number of transitions and transversions versus divergence and display of substitution saturation and CpG, TpG, CpA value was respectively plotted and calculated by DAMBE 5.327 (Xia, 2000).

Phylogenetic analysis

Maximum-parsimony analysis of this coding data set was performed using PAUP 4.01 (Swofford,1993) with *Ehrharta longifolia* and *Phyllostachys aurea* as outgroups. All characters were equally weighted, and heuristic search with 50 repetitions and TBR, stepwise addition, and MULTIPARS options was used in the search for most parsimonious trees. A strict consensus tree was generated from the equally most-parsimonious trees, and decay analysis (Bremer support) up to four steps longer was conducted to evaluate the support for individual clades (Bremer, 1988). The sequence data were also analyzed with a neighbor-joining (NJ) method using the Juke-Cantor and Kimura two parameter distance estimates (Kimura, 1993; Saitou and Nei, 1987). Topological robustness was assessed by bootstrap analysis with 1000 replicates using simple taxon addition (Felsenstein, 1985). Gaps were treated as missing data.

RESULT

Sequence characteristics

The mean guanine-cytosine (GC) content of the rice *matK* partial sequence was 34.84% and adenine-thymine (AT)

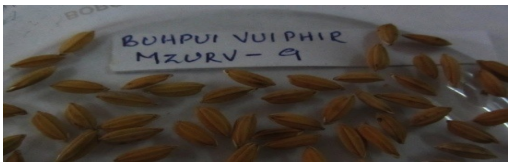
S. No	Sample Accession number, local name and *matK* gene GenBank Accession No.	
1	MZURV-2 Kungrei JX192665	
2	MAURV-3 Mangbuh JX192666	
3	MZURV-4 Shaan JX192667	
4	MZURV-5 Halflong buh JX192668	
5	MZURV-6 Roenga JX192669	
6	MZURV-7 Project Buh JX192670	
7	MZURV-8 Buhsakei JX192671	
8	MZURV-9 Vuiphir JX192672	

Figure 1. Rice samples used in the present study from Mizoram, NEI.

content is 65%. The aligned sequences resulted in a final data matrix with 826 bp with no alignment gaps. 58 of the characters (7.02%) were variable in the data set (Figure 2), out of which 37 (63.9%) were potentially phylogenetic informative sites. The alignment of these partial *matK* sequences showed 702 characters were constant, 74 variable characters are parsimony-uninformative and 50 numbers of parsimony–informative characters are present for the maximum parsimonious tree including outgroup. The Mizoram Oryza Species are most closely related with *O. barthi* (Cameroon), *O. ridleyi* (Thailand), *O. rufipogon* (India),*O. rufipogan* (Thailand), *O. sativa* (China), *O. meridionalis* (Australia). Pairwise divergence of sequences ranges from 0.008 to 0.074% between the outgroups and *Oryza specie*s and from 0 to 0.014% within *Oryza* species.

The mean guanine-cytosine (GC) content of the rice *matK* partial sequence was 34.84% and adenine-thymine (AT) content is 65%. The aligned sequences resulted in a final data matrix with 826 bp with no alignment gaps. 58 of the characters (7.02%) were variable in the data set (Figure 2), out of which 37 (63.9%) were potentially phylogenetic informative sites. The alignment of these partial *matK* sequences showed 702 characters were constant, 74 variable characters are parsimony-uninformative and 50 numbers of parsimony–informative characters are present for the maximum

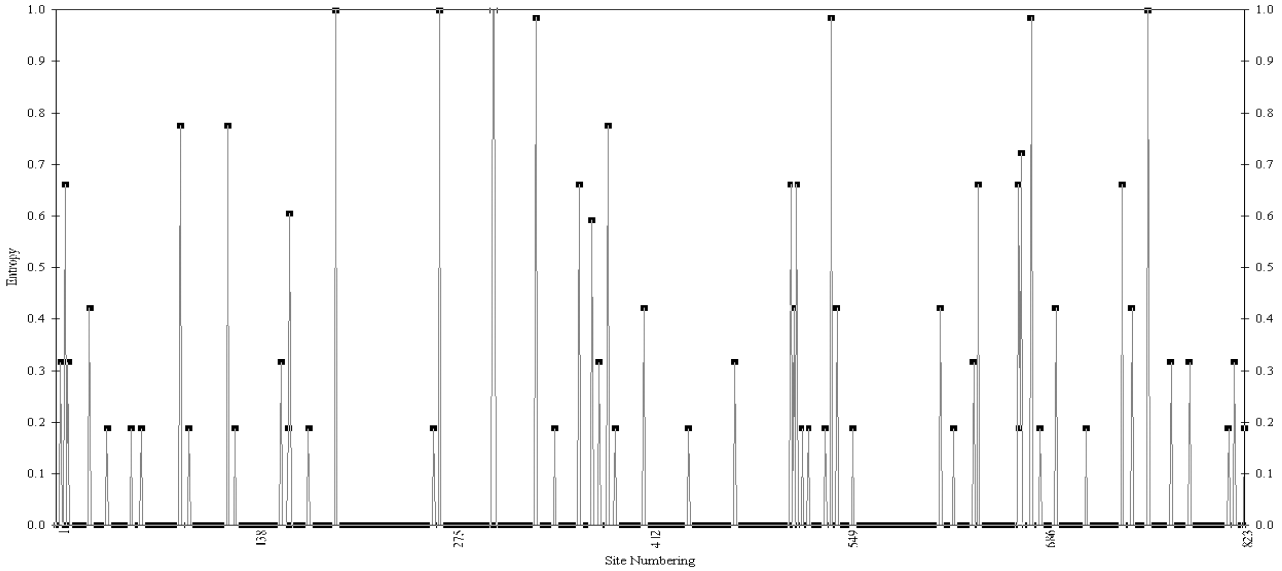


Figure 2. Variable substitution rate over sites, this plot show visually which sequence segments are conservative and which are variable, the sharp pick followed by black dot represents particularly variable sites. (Xia, X. 2000, DAMBE 5.327).

parsimonious tree including outgroup. The Mizoram Oryza Species are most closely related with *O. barthi* (Cameroon), *O. ridleyi* (Thailand), *O. rufipogon* (India), *O. rufipogan* (Thailand), *O. sativa* (China), *O. meridionalis* (Australia). Pairwise divergence of sequences ranges from 0.008 to 0.074% between the outgroups and *Oryza speci*es and from 0 to 0.014% within *Oryza* species.

The transition/transversion rate ratios are k1 = 5.314 (purines) and k2 = 3.21 (pyrimidines). The estimated Transition/Transversion bias (R) is 1.97 (Table 2). Substitution pattern and rates were estimated under the Tamura (1992) model, where R = [A×G×k1 + T×C×k2] / [(A+G) × (T+C)] and the maximum Log likelihood for this computation was -1576.086.

Figure 3 shows that substitutions of all the three codons can be easily fitted to a two different straight lines, indicating that they are unsaturated. To understand the necessity and usefulness of the weighting, we estimated the relative frequencies of transitions and transversions using ML composite Tamura-Nei parameter distance method (ML-CompositeTN93). Here, we observed the number of observed transitions relative to that of

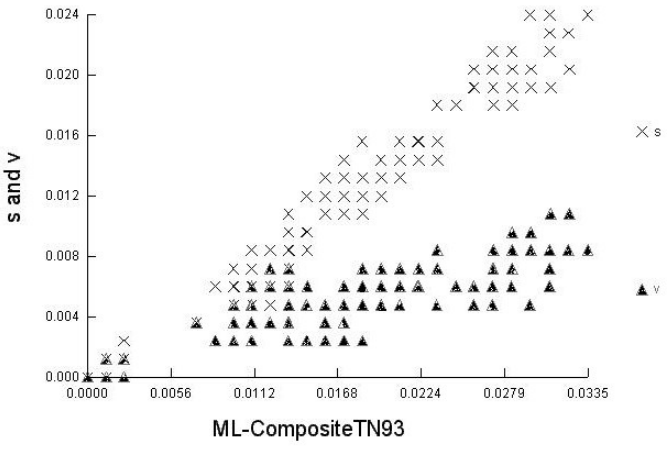


Figure 3. Transition and transversion versus divergence. Here '×S' represent Transition rate and '▲V' represent Transversion rate. The number of observed transitions relative to that of transversions gradually decreases with increasing divergence. A plot of the number of transitions and transversions versus divergence and display of substitution saturation (Xia 2000, DAMBE 5.327).

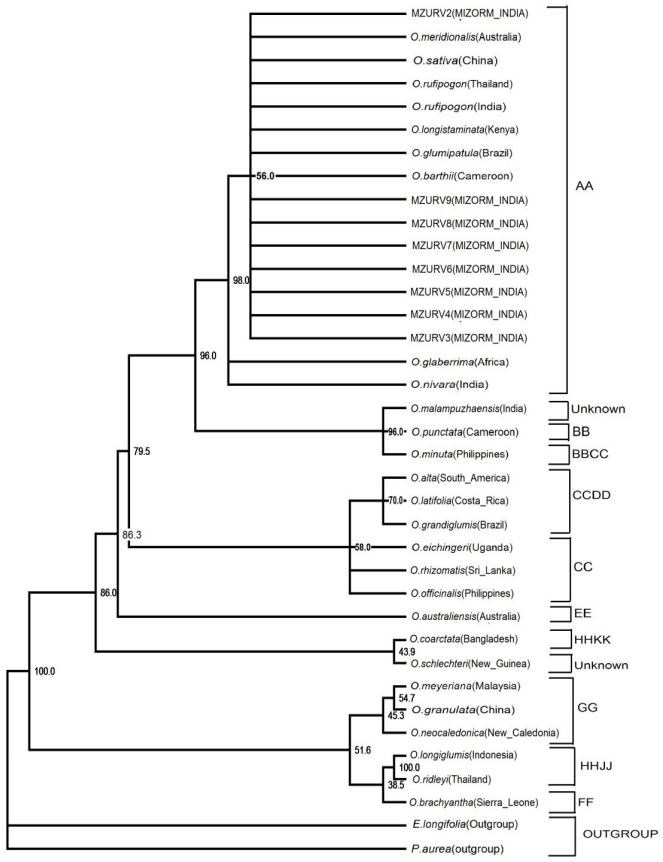


Figure 4. Strict consensus of three equally parsimonious trees of 8 *Oryza species* from Mizoram, India, 27 *Oryza species* from other countries and two outgroups (*Ehrharta longifolia*, and *Phyllostachys aurea*) generated from sequences of *matK*. Tree length = 383, consistency index = 0.904, retention index = 0.944. Numbers on the branches are bootstrap values from 1000 replicates. Capital letters following a bracket indicate the previously recognized genome type of the species. Major genome types are represented as a monophyletic group.

Table 1. List of *Oryza* species used in the present study with their chloroplast haplotype and *matK* characters (Ge *et al*., 1999; Vaughan *et al*., 2003).

Sl. No	Name of the Species	Genome	Gene Bank Accession No.	Origin	No. of variable characters against MZURV2		RA (CpG)	RA (TpG)	RA (CpA)
					Nucleotide	Amino acid			
1	*O. glaberrima*	AA	AF148654	Africa	1.0	1	0.7962	0.9904	0.9199
2	*O. glumipatula*	AA	AF148653	Brazil	1.0	Non	0.836	0.973	0.8936
3	*O. grandiglumis*	CCDD	AF148666	Brazil	10.0	4	0.7631	0.9911	0.8805
4	*O. granulata*	GG	AF148674	China	24.0	11	0.9258	0.9587	0.8331
5	*O. latifolia*	CCCD	FJ908264	Costa Rica	10.0	4	0.7631	0.9911	0.8805
6	*O. longiglumis*	HHJJ	AF148672	Indonesia	25.0	14	0.8502	0.914	0.8998
7	*O. longistaminata*	AA	AF148656	Kenya	Non	Non	0.8014	0.9904	0.8994
8	*O. malampuzhaensis*	unknown	AF489915	India	10.0	3	0.7388	0.9785	0.9487
9	*O. minuta*	BBCC	AF148663	Philippines	10.0	3	0.7388	0.9785	0.9487
10	*O. neocaledonica*	GG	FJ908266	New Caledonia	23.0	12	0.8769	0.95	0.8319
11	*O. nivara*	AA	AF148652	India	1.0	1	0.7962	0.9904	0.9199
12	*O. officinalis*	CC	AF148658	Philippines	9.0	3	0.7582	0.9911	0.9006
13	*O. punctata*	BB	AF148662	Cameroon	11.0	3	0.7331	0.9531	0.9487
14	*O. rhizomatis*	CC	FJ908263	Sri Lanka	9.0	3	0.7582	0.9911	0.9006
15	*O. ridleyi*	HHJJ	FJ908265	Thailand	25.0	14	0.8502	0.914	0.8998
16	*O. rufipogon*	AA	FJ908261	India	Non	Non	0.8014	0.9904	0.8994
17	*O. rufipogon*	AA	EU434286	Thailand	Non	Non	0.8014	0.9904	0.8994
18	*O. sativa indica*	AA	AF148650	China	Non	Non	0.8014	0.9904	0.8994
19	O. meridionalis	AA	AF148657	Australia	Non	Non	0.8014	0.9904	0.8994
20	*O. schlechteri*	unknown	AF148668	New Guinea	14.0	6	0.8023	0.9771	0.916
21	*O. alta*	CCDD	AF148664	South America	10.0	4	0.7631	0.9911	0.8805
22	*O. barthii*	AA	AF148655	Cameroon	Non	Non	0.8805	0.9904	0.8994
23	*O. brachyantha*	FF	AF148670	Sierra Leone	26.0	14	0.9167	0.9117	0.8507
24	*O. coarctata*	HHKK	HE586098	Bangladesh	15.0	7	0.7971	0.9804	0.9102
25	*O. eichingeri*	CC	AY318858	Uganda	9.0	3	0.7582	0.9911	0.9006
26	*O. australiensis*	EE	AF148667	Australia	15.0	6	0.8105	0.9669	0.9111
27	*O. meyeriana*	GG	AF148673	Malaysia	23.0	12	0.9088	0.9625	0.8562
28	MZURV2	AA	JX192665	Mizoram, NEI	Non	Non	0.8014	0.9904	0.8994
29	MZURV3	AA	JX192666	Mizoram, NEI	Non	Non	0.8014	0.9904	0.8994
30	MZURV4	AA	JX192667	Mizoram, NEI	Non	Non	0.8014	0.9904	0.8994
31	MZURV5	AA	JX192668	Mizoram, NEI	Non	Non	0.8014	0.9904	0.8994
32	MZURV6	AA	JX192669	Mizoram, NEI	Non	Non	0.8014	0.9904	0.8994
33	MZURV7	AA	JX192670	Mizoram, NEI	Non	Non	0.8014	0.9904	0.8994
34	MZURV8	AA	JX192671	Mizoram, NEI	1.0	Non	0.7962	0.9937	0.8936
35	MZURV9	AA	JX192672	Mizoram, NEI	Non	Non	0.8014	0.9904	0.8994
36	*E.longifolia* (Outgroup)		AF164392	South Africa	76.0	32	0.6733	0.9499	0.9416
37	*P. aurea* (Outgroup)		AF164390	USA	46.0	25	0.7484	0.9818	0.905

RA - Relative abundance, CpG - — Cytosine —phosphate— Guanine —, TpG - — Thymine —phosphate— Guanine —, CpA - — Cytosine —phosphate— Adenine ----; Non - — No Variation from MZURV2; MZURV2 have more conserved sequence compare to the other Accession, so we took MZURV2 as a standard from Eastern Himalayan region *Oryza* species.

Table 2. Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution.

	A	T	C	G
A	-	5.92	3.07	***13.79***
T	4.67	-	***9.86***	2.59
C	4.67	***19***	-	2.59
G	***24.84***	5.92	3.07	-

*** Rates of different transitional substitutions are shown in **bold** and those of transversionsal substitutions are shown in *italics*. The nucleotide frequencies are 28.75% (A), 36.41% (T/U), 15.96% (C) and 18.89% (G) (Kimura, 1980; Tamura 1992).

transversions gradually decreased with increasing divergence.

matK gene based signature sequence

We propose a signature for characterizing Oryza chloroplast matK gene using several distinctive oligonucleotide relative abundance values. These include measurements of G+C content, the dinucleotide relative abundance values of CpG, CpA, and TpG (Table 1). CpG rich matK sequences were O. granulata (China, GG), O. brachyantha (Sierra Leone, FF), O. meyeriana (Malaysia, GG) and CpA rich sequences belonged to BB related genome varieties. All the genotypes were rich in TpG sequences.

Phylogenetic analysis

Parsimony analysis yielded three equally parsimonious trees, each 383 steps long with a consistency index (CI) of 0.904 and a retention index (RI) of 0.944. The topology of the ingroup (all *Oryzeae species*) was exactly same when *Ehrharta longifolia* or one species of subfamily *Bambusoideae* (*Phyllostachys aurea*) were specified as outgroups. The analysis using genetic distances, maximum likelihood and maximum parsimony all resulted in very similar trees. The most parsimonious tree of *Oryza* species based on *matK* sequence is shown in Figure 4.

All *Oryza* species groups were clearly distinguished. The species belonging to the *Oryza* AA genome group clustered together and were separated from those belonging to the other genome groups (BB, BBCC, CC/CCDD, EE, GG, HHJJ and FF). All rice species within CC/DD genome formed a monophyletic group with 70% bootstrap.

DISCUSSION

Some mutationally (or substitutionally) active regions are present in the genomes, and the *matK* region of 12 genera indicated that such regions exist in chloroplast genomes. The number of polymorphic sites for nucleotide sequence varied from 0 to 26 (Table 1) with an arithmetic mean value of 8.1 in 826 bp and for amino acids sequence varied from 0 to 14 with an arithmetic mean value 3.65 in 275 bp, indicating great potential for finding variable regions carrying phylogenetic information. Mutationally active regions in chloroplast genomes are frequently shown as problematic for phylogenetic analyses at higher taxonomic levels because of recombination and sequence convergence (Muller *et al*., 2006).

Monophyletic groups revealed by the phylogenetic reconstruction (Figure 4.) are either concordant or discordant with taxonomic sections recognized in the most recent classification of the genus (Vaughan, 1994). The AA, BB, and CC genomes are most closely related and together form a sister group with the DD genome. This monophyletic group, containing the AA through EE genomes, corresponds to section *Oryza*. The GG genome, which occupies the most basal position of the genus, constitutes *O. granulata* from China and O. m*eyeriana* from Malaysia. The HHJJ genome types that are included in *O. ridleyi*, however, form a monophyletic group in the phylogenetic hypothesis (Figure 4) (Song *et al*., 1999). According to the crossability between *O. sativa* and other rice species from other genome types, the wild species have been categorized as the primary, secondary, and tertiary gene pools for the cultivars (Khush, 1997). Species of the AA genome are easily crossed with *O. sativa* and are regarded as the primary gene pool. BB to EE genomes represent the secondary gene pool, and the remaining genomes represent the tertiary gene pool. The phylogenetic relationships between *Oryza* species correlate with the crossability of rice genomes (Figure 4). The tertiary gene pool contains less than one-third of the species diversity, but nearly half of the genomic diversity (HHKK, HHJJ and FF). These genomes have great potential to provide novel, beneficial genes. The AA genome, which contains cultivated rice, is one of the most recently diverged lineages within the rice genus (Figure 4). The AA genome contains the maximum number of diploid species and is geographically the most cultivated rice genome. The relationships within the AA genome show that the widely cultivated species *O. sativa* is most closely related to two wild species distributed in Asia, *O. nivara* and *O. rufipogon*, supporting the previous hypothesis of an Asian origin of *O. sativa* (Vaughan, 1994; Khush, 1997). The BB [*O. punctata* (Cameroon)] and BBCC genome [*O. minuta* (Philippines)] are closely related to the AA genome (Ge *et al*., 1999; Nishikawa *et al*., 2005) with 96% bootstrap. Within the AA genome group, phylogenetic analysis could separate 2 groups, through bootstrap values were barely more than 50%. Asian wild rice *O. rufipogon* / *O. nivara* clustered in the same group with almost all Asian cultivated *O. sativa* with high bootstrap. The African cultivated species, *O. glaberrima*, is most closely related to the African wild species, *O. longistaminata* and Asian *O. nivara*, as well as *O. glumaepatula* and *O. barthii* which occur in Central and South America. The African rice group (*O. longistaminata* and *O. glaberrima*) were sister to all accessions of Asian cultivars. The Australian, *O. meridionalis* clustered with Asian *Oryza* species because of the presence of the AA genotype. The Australian, *O. australiensis* (EE genome), accessions formed a separate monophyletic group with 83% bootstrap. The GG genome representing *O. meyeriana* (Malaysia), *O. granulata* (China) and *O. neocaledonica* (New Caledonia) formed a single clade. The two unknown genome varieties clustered in different positions on the tree with *Oryza* species from different locations. The *O. malampuzhaensis* (India, Unknown genome) clustered with

O. punctata (Cameroon BB genome) and *O. minuta* (Philippines, BBCC genome) with 96% bootstrap and *O. schlechteri* (New Guinea, Unknown) clustered with O. *coarctata* (Bangladesh, HHKK).

Our phylogenetic analysis of the *matK* sequences of *Oryza* leads to the following conclusions: *Oryza* species form a monophyletic group and all genome varieties are monophyletic, the hypothesis of close affinities between monoecious genera in *Oryza* was not supported by the *matK* sequence data, suggesting the possibility of multiple origins of the floral structures between the *Oryza* species. Based on the phylogenetic reconstruction, we were able to recognize the unknown genome variety *O. malampuzhaensis* from India to be closer to BB and AA genomes, whereas *O. schlechteri* from New Guinea is closer to HHKK genome variety. The phylogenetic tree clearly separated the rice species with the AA genome from those with other genomes. This phylogenetic study revealed that the EE genome is most closely related to the CC genome progenitor that gave rise to the CCDD genome. The clade having AA genotype did not resolve properly indicating that there are missing species and there is a need for further research on *Oryza* species of this Eastern Himalayas region.

ACKNOWLEDGEMENTS

This work was supported by research grants from the Bioinformatics Infrastructure Facility and State Biotech Hub sponsored by Department of Biotechnology (DBT), New Delhi, India.

REFERENCES

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42(4): 795-803.

Buso, G.C., Rangel, P.N. and Ferreira, M.E. 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.

Desikachar, O. S. V. 1974. A review of the tectonic and geological history of eastern India in terms of 'plate tectonics' theory. Journal of the Geological Society of India 33: 137-149.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39(4): 783-791.

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

Hilu, K.W. and Liang, H. 1997. The matK gene: sequence variation and application in plant systematics. American Journal of Botany 84: 830-839.

Hilu, K.W., Alice L,.A. and Liang, H. 1999. Phylogeny of Poaceae inferred from matK sequences. Annals of Missouri Botanical Garden 86:835-851.

Jaiswal, P., Ni, J., Yap, I. and Ware, D. 2005. Gramene: A genomics and genetics resource for rice. Rice Genetics Newsletter 22(1): 9-16.

Johnson, L. A. & Soltis, D. E. 1994. matK DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Systematic botany 19(1): 143-156.

Khush, G.S. 1997. Origin, dispersal, Cultivation & variation of rice. Plant molecular biology 35: 25-34.

Kojima, Y.,, K., Fukuoka, S., Nagamine, T. and Kawase, M. 2002. Development of an RFLP-based Rice Diversity Research Set of Germplasm. Genome(3): 451-459.

Kojima, Y., Ebana, K., Fukuoka, S., Nagamine, T. and Kawase, M. 2005. Development of an RFLP-based Rice diversity research set of germ plasm. Breeding Science 55: 431-440.

Kumar, S., Tamura, K. and Nei, M. 1993. MEGA: Molecular evolutionary genetics analysis software for microcomputers. Oxford Journals 10:189-191.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R. and McGettigan, P.A. 2007. Clustal W and clustal X version 2.0. Bioinformatics 23: 2947-2948.

Muller, K. 2006. Incorporating information from length -mutational events into phylogenetic analysis. Molecular Phylogenetics and Evolution 38: 667-676.

Nishikawa, T., Duncan, A.,Vaughan, 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.

Olmstead, R.G. and Palmer, J.D. 1994. Chloroplast DNA systematics: a review of methods and data analysis. American Journal of Botany 81: 1205-1224.

Pattanayak, A., Ngachan, S.V. and Mohanty, A.K. 2006. Status paper on rice in North East India. Rice Knowledge Management Portal (RKMP), India pp:1-84.

Prashanth, S.R., Parani, M., Mohanty, B.P., Talame, V., Tuberosa, R. and Parida, A. 2001. Genetic diversity in cultivars and landraces of sativa subsp.revealed by AFLP markers. Genome 44 (3): 476-494.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

Saitoun, M. and Nei,. 1987. The neighbor-joining method: a new method for reconstruction phylogenetic trees. Molecular Biology and Evolution 4: 406- 425.

Salles, G., Buso, C., Rangel, P.H.N. and Ferreira, M.E. 1996. Analysis of random and specific sequences of nuclear and cytoplasmic DNA in diploid and tetraploid American wild rice species (Oryza spp.). Canadian Journal of Botany 74(1): 125-134.

Sang, T., Crawford, D.J. and Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution,

and biogeography of *Paeonia* (Paeoniaceae). American Journal of *Botany* 84: 1120-1136.

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

Shirasawa, K., Monna, L., Kishitani, S. and Nishio, T. 2004. Single nucleotide polymorphisms in randomly selected genes among japonica Rice (Oryza sativa L.) varieties identified by PCR-RF-SSCP. DNA Research 11: 275-283.

Song, G.E., Tao S., Bao-Rong, L.U. and De-Yuan, H. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. PNAS 96: 14400-14405.

Swofford, D. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.01. Sinauer Associates: Sunderland, Massachusetts).

Vaughan, D.A. 1994. The wild relatives of rice. A ge netic resources handbook (International Rice Research Institute press: Philippines). P. 157.

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

Ware, D. and Jaiswal, P. 2001. Gramene home page, Cold Spring Harbor Laboratory and Cornell University, USA. *http://www.gramene.org./*

WWF-US, Asia Program. 2005. Ecosystem Profile, Eastern Himalayas Region. Pp. 1-100 *Asia* Program (The Critical Ecosystem Partnership Fund: India)

Xia, X. and Xie, Z. 2001. DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. The American Genetic Association: 371-373.

Yu, J., Xue, J.H. and Zhou, S.L. 2011. New universal matK primers for DNA barcoding angiosperms. Journal of Systematic and evolution 49 (3): 176-181.