Pak. J. Bot., 47(5): 1773-1782, 2015.

POPULATION STRUCTURE AND DIVERSITY OF THE AA GENOME OF RICE BASED ON SIMPLE SEQUENCE REPEATS VARIATION IN ORGANELLE GENOME

SHAHID MASOOD SHAH[1,2*], KASHIF ASLAM[2,3*], GHULAM SHABIR[2], ABDUL REHMAN KHAN[1], BILAL HAIDER ABBASI[4], ZABTA KHAN SHINWARI[4] AND MUHAMMAD ARIF[2]

[1] Biotechnology Program, Department of Environmental Sciences, COMSATS Institute of Information Technology, Abbottabad, Pakistan
[2] National Institute for Biotechnology and Genetic Engineering (NIBGE) Faisalabad, Pakistan
[3]Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan[4]
[4]Department of Biotechnology, Quaid-i-Azam University, Islamabad 45320, Pakistan
**Corresponding authors Email address; smasood@ciit.net.pk, dr.kaslam@gmail.com , Phone No; +923335273893*

Abstract

Maternally inherited mitochondrial and chloroplast genomes based Simple Sequence Repeat (SSR) variations were examined for their contribution to diversity of rice genome. Population structure and diversity analysis based on mitochondria and chloroplast inherited genome has been studied less as compared to nuclear genome inheritance. The present study was designed to evaluate the population structure and diversity of rice grown in Pakistan along with other countries based on maternally inherited mitochondria and chloroplast genome. The mitochondrial and chloroplast genomes were analyzed by using 42 mitochondrial and 20 chloroplast pairs of SSR primers. A slightly higher percentage of polymorphism was observed in chloroplast (30%) than mitochondria (28.57%). The average gene diversity for both mitochondrial and chloroplast was 0.32 oscillating from 0.041 to 0.620. The Polymorphism Information Content (PIC) value ranged from 0.040 to 0.543 with an average of 0.282, while the allelic richness ranged from two to four alleles with an average of 2.779 alleles. Mononucleotide repeats stood first (50% polymorphic) for detecting polymorphism for organelle genomes followed by tri- (25%), tetra- (14.29%) and dinucleotide (12.5%), respectively. Cluster and population structure analysis revealed two groups of accessions. On the basis of our results the AA genome of Asian cultivated rice diverges from the same origin during evolution.

Key words; Mitochondrial and chloroplast genomes; SSR variation; *Oryza sativa*; Population structure; Genetic diversity.

Introduction

Genetic diversity of plants is vital to withstand in a wider climate and their adaptability across changing environment. This aspect of life has a positive impact on the sociocultural environment of life. Genetic diversity is the backbone for variation within the population and among populations which ultimately lead to the development of the diverse population structure of life.

Oryza genus comprised of 23 species (2 cultivated and 21 wild). Out of two cultivates species Asian cultivated specie (*Oryza sativa* L.) comprised of AA genome, helps to maintain mankind's survival since its evolution. A huge number of determinations have been given to evaluate the population structure and genetic diversity in rice based on the nuclear genome around the world (Lu *et al.,* 2005; Jin *et al.,* 2010; Agrama *et al.,* 2010; Courtois *et al.,* 2012; Das *et al.,* 2013) and in Pakistan as well (Pervaiz *et al.,* 2010; Ashfaq and Khan 2012; Shah *et al.,* 2013). While on the other hand a few examples are available to evaluate population structure and genetic diversity of rice based on chloroplast and mitochondrial genomes around the world (Nishikawa *et al.,* 2005; Rabbani *et al*., 2010) while in the case of Pakistan no study is available so far.

Organelle genomes (chloroplast and mitochondrial) are naturally non-recombinant and effectively diploid (Taberlet *et al.,* 2007). They contain genes that are coded for functional and structural component of other organelles. The uniparental inheritance of these genomes makes them suitable for tracing phylogenetic relations, genetic diversity and population structure. The determination of diversity and population structure is crucial for efficient utilization of germplasm in the breeding programs.

Different markers such as RFLP (Restriction Fragment Length Polymorphism), AFLP (Amplified Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), SSR (Simple Sequence Repeat) and SNPs (Single Nucleotide Polymorphism) are being utilized by scientist to explore the genome variability. Out of these the Simple Sequence Repeats have been used as potential markers to find out genetic diversity and population structure (Zhao *et al.,* 2009) association mapping (Borba *et al.,* 2010) and QTL mapping (Li *et al.,* 2011) etc. since their identification and development in plant and animal sciences. These are present in all forms of DNA i.e. both nuclear and organelle.

Mitochondrial and chloroplast genomes are being analyzed through different computer search to find out the presence of SSRs. The first attempted was made by Powell *et al.* (1995). Ishii *et al.,* 2001 studied 59 accessions of rice with the help of 24 nuclear and 10 chloroplast SSR markers. Rajendrakumar *et al.* (2007) used sequences of rice chloroplast and mitochondrial genome to identify SSRs with the help of a Software SSR Identification tool. In case of mitochondrial genome 2528 SSR markers were identified, while in case of chloroplast 870 SSR were identified. These SSRs are present in both genic and intergenic region of organelle genome. Out of these only few were designed for further study.

The wild rice is thought to be the progenitor of cultivated rice. The determination of relation between wild and cultivated rice is obligatory to answer the question of rice domestication. The maternally inherited

genomes, that are highly conserved as these are not subjected to recombination unlike the nuclear genome, needs to be investigated. The present investigation was carried out to find and compare the allelic variation at SSR loci in cp (Chloroplast) and rmt (Mitochondrial) genome from a representative set of *O. sativa* and wild AA genome. In addition cultivated accessions are investigated for phylogenetic and population structure relationship with the wild progenitors.

Materials and Methods

Plant material: A total of 95 genotypes, including wild rice, land races, cultivated varieties and lines were analyzed for mitochondrial and chloroplast genomes (Supplementary Table 1). The accessions from Pakistan were collected from Rice Research Institute, (RRI) Kala Shah Kaku, Lahore and from Plant Genetic Resource Institute, National Agriculture Research Centre (NARC), Islamabad. The cultivated accessions across the world were provided by the International Rice Research Institute (IRRI), Philippines, for research purposes. The wild rice accessions were taken from T.T. Chang Genetic Resource Center, IRRI, Philippines. The rice genotypes from different countries were characterized in to indica, japonica, aus and aromatic on the basis of SNP variation based on RiceOPA2.1 (Shah *et al.,* 2015).

DNA extraction: DNA was isolated from 21 days fresh leaves with the help of CTAB method (Doyle & Doyle, 1987). The extracted DNA was run on 1% agarose to check quantity and quality. The DNA concentration was adjusted to 30ng/µl for PCR amplification.

Simple sequence repeat markers: A total of 62 SSR markers were examined, including 42 mitochondrial and 20 chloroplast SSR markers, which were previously reported by Rajendrakumar *et al.* (2007), Nishikawa *et al.* (2005) and Ishii & McCouch (2000). Their sequences along with other information are provided in Supplementary Table 2.

PCR: PCR reagents were dispensed in 96-well PCR plates. 20 µl reaction mixture including 2 µl of DNA template, 2 µl of 10X PCR buffer (containing 200 mM Tris-HCl pH 8.3, 500 mM KCl, 15 mM $MgCl_2$), 1µl of 1 mM dNTPs, 1 µl of $MgCl_2$, 0.5 µl each of 5 µM forward and reverse primers and 0.25 µl of 1 U/µl Taq DNA polymerase with 12.75 µl sterile nano-pure water were dispensed in each well adding with 1 drop of mineral oil and then covered with aluminum foil. The PCR plate was placed in automated G-Storm thermal cycler. The following PCR profile was used: initial denaturation at 94°C for 5 minutes; then 35 cycles with denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 30 seconds; final extension at 72°C for 5 minutes and storage at 4°C.

Amplified products were subjected to 6% polyacrylamide gel electrophoresis. The gel was stained with 10% solution of SYBR® Safe and visualized with the help of UV light. The image was taken and scored starting from bottom to top as A/A, B/B ….N/N.

Data analysis: Using the Power Marker version 3.25 software package (Liu & Muse, 2005; http// www.powermarker.net), the genetic diversity of each accession was analyzed on the basis of four statistical parameters: major allele frequencies, allele number, gene diversity and PIC. For the unrooted phylogenetic tree, genetic distance was calculated using the "C.S. Chord 1967" distance (Cavalli-Sforza & Edwards, 1967) followed by phylogeny reconstruction using UPGMA as implemented in Power Marker and dendogram was viewed by using TREEVIEW software. Bootstrapping of the UPGMA tree was performed using The Power Marker with 1,000 iterations.

In addition model based STRUCTURE (Pritchard *et al.,* 2000; Falush *et al.,* 2003) software was used to infer population structure using a burn in length of 20,000, run length about 10,000 and a model allowing for admixture and correlated allele frequencies and testing for K = 2 to K = 10. Five independent runs yielded consistent results. To determine the best K value for the material under study Structure Harvester (Earl & vonholdt, 2012) was used which is based on Evanno method (Evanno *et al.,* 2005).

Results

Nature of SSR: Chloroplast and mitochondrial genomes of 95 accessions from the AA genome of rice were analyzed for phylogenetic relationships with 62 pair of SSR markers. The repeat classes of study nucleotides ranged from mononucleotide to hexanucleotide. Out of 62 under investigation SSRs, only 22 belonged to mononucleotide class, eight from di-, sixteen from tri-, seven from tetra-, eight from penta- and only one was from hexanucleotide (Fig. 1).

Microsatellite allelic diversity was examined on the basis of amplified banding patterns. Out of 42 microsatellite loci for mitochondria only 12 SSR markers showed polymorphism whereas in case of chloroplast 6 out of 20 were found polymorphic. The polymorphic markers were utilized for the all analysis and the rest of the markers are discarded. Their information is provided in Supplementary Table 3. Microsatellites were amplified clearly in all 95 accessions showing differences on genetic pattern of 62 SSR loci.

Mononucleotide repeats stood first (50% polymorphic) in detecting polymorphism for organelle genome followed by trinucleotide (25%). Poly (A) or (T) repeats of mononucleotide were found more polymorphic as compared to poly (C) or (G). Di- (12.50%) and tri-nucleotide repeats (14.29%) resulted in an almost equal level of polymorphism. Pentanucleotide repeats were amplified in all accessions but did not show any polymorphism. Only one hex nucleotide repeat was analyzed which resulted in polymorphic products. The total number of repeats and proportion of repeats found to be polymorphic is shown in Figure 1.

Supplementary Table 1. List of 95 accessions along with their country of origin, type and group used for SSR analysis.

S. No.	Varieties	Country	Type	Group
1.	IRBL12-M	Philippines	Cultivated	Japonica
2.	IRBLk-Ka	Philippines	Cultivated	Japonica
3.	IRB LK S-F5	Philippines	Cultivated	Japonica
4.	IRBLz5-CA	Philippines	Cultivated	Japonica
5.	IR 79906-B-192-2-3	Philippines	Cultivated	Indica
6.	PSBRC80	Philippines	Cultivated	Indica
7.	IRBLt-K59	Philippines	Cultivated	Aromatic
8.	IRBL9-W	Philippines	Cultivated	Japonica
9.	IRBLi-F5	Philippines	Cultivated	Japonica
10.	IRBLz-Fu	Philippines	Cultivated	Japonica
11.	Brown Gora	India	Cultivated	Aus
12.	IR 74371 -54-1-1	Philippines	Cultivated	Indica
13.	Vandana	India	Cultivated	Indica
14.	IR 71525-19-1-1	Philippines	Cultivated	Indica
15.	IRBB 57	Philippines	Cultivated	Indica
16.	Super fine	Pakistan	Cultivated	Indica
17.	IR55419-04	Philippines	Cultivated	Indica
18.	UPLRI7	Philippines	Cultivated	Indica
19.	IR 78908-263-2-2-3	Philippines	Cultivated	Japonica
20.	IR 74371-46-1-1	Philippines	Cultivated	Indica
21.	IR 80021 -B-86-3-4	Philippines	Cultivated	Indica
22.	IR 78875-131-B-1-4	Philippines	Cultivated	Indica
23.	IR 78877-181 -B-1-2	Philippines	Cultivated	Indica
24.	Way Rarem	India	Cultivated	Indica
25.	KSK-133	Pakistan	Cultivated	Indica
26.	KSK-282	Pakistan	Cultivated	Indica
27.	IR-72	Philippines	Cultivated	Indica
28.	DR-92	India	Cultivated	Indica
29.	Sufaid 86	Pakistan	Cultivated	Indica
30.	PK.386	Pakistan	Cultivated	Indica
31.	Supra	Pakistan	Cultivated	Indica
32.	APO	Philippines	Cultivated	Indica
33.	B 6144 F-MR-6-0-0	Philippines	Cultivated	Indica
34.	IR 72667-18-1-B-B-3	Philippines	Cultivated	Indica
35.	IR 74963-262-5-1-3-3	Philippines	Cultivated	Indica
36.	IR 78875-131-B-1-1	china	Cultivated	Indica
37.	KCD-1	Philippines	Cultivated	Indica
38.	IR 78878-53-2-2-2	Philippines	Cultivated	Indica
39.	IR 64683-87-2-2-3-3	Philippines	Cultivated	Indica
40.	IR-6	Pakistan	Cultivated	Indica
41.	DR-58	Philippines	Cultivated	Indica
42.	IR-64	India	Cultivated	Indica
43.	Pusa Basmati-1	Philippines	Cultivated	Aromatic
44.	Salumpikit	Philippines	Cultivated	Indica
45.	IRBL11-Zh	Philippines	Cultivated	Aromatic
46.	IRBL19-A	Philippines	Cultivated	Aromatic
47.	IRBLkh-K3	Philippines	Cultivated	Aromatic

Supplementary Table 1. (Cont'd.)..

S. No.	Varieties	Country	Type	Group
48.	IRBLkm-Ts	Philippines	Cultivated	Aromatic
49.	IRBLsh-B	Philippines	Cultivated	Aromatic
50.	IRBLzt-T	Philippines	Cultivated	Aromatic
51.	KATARI BHOG	India	Cultivated	Aromatic
52.	Supri	Pakistan	Cultivated	Aromatic
53.	*O. rufipogon*-80472	India	Wild	Wild
54.	BPI 76 NS	Philippines	Cultivated	Indica
55.	IR43450 SKN-506-2-2-1-1	Philippines	Cultivated	Indica
56.	IRB La-A	Philippines	Cultivated	Japonica
57.	NIAB-IR-9	Pakistan	Cultivated	Indica
58.	IR 74371-3-1-1	Philippines	Cultivated	Indica
59.	IR 77080-B-4-2-2	Philippines	Cultivated	Indica
60.	Pusa Basmati -1121	India	Cultivated	Aromatic
61.	Basmati-Pak	Pakistan	Cultivated	Aromatic
62.	Shaheen Basmati	Pakistan	Cultivated	Aromatic
63.	DR-82	Pakistan	Cultivated	Indica
64.	IRBLa-C	Philippines	Cultivated	Japonica
65.	TKM 6	India	Cultivated	Aromatic
66.	Basmati-198	Pakistan	Cultivated	Aromatic
67.	Basmati 2000	Pakistan	Cultivated	Aromatic
68.	Basmati-370	Pakistan	Cultivated	Aromatic
69.	Basmati-385	Pakistan	Cultivated	Aromatic
70.	KSK-98316	Pakistan	Cultivated	Aromatic
71.	KSK-99417	Pakistan	Cultivated	Aromatic
72.	KSK-99512	Pakistan	Cultivated	Indica
73.	Basmati.515	Pakistan	Cultivated	Aromatic
74.	Super Basmati	Pakistan	Cultivated	Aromatic
75.	P-35	Philippines	Cultivated	Indica
76.	IR-24	Philippines	Cultivated	Japonica
77.	ZGY-1	china	Cultivated	Indica
78.	Nona Bokra	India	Cultivated	Indica
79.	Pokkali	India	Cultivated	Indica
80.	W 1263	Malaysia	Cultivated	Indica
81.	*O. barthii*-104103	chad	Wild Progenitor	wild
82.	*O. barthii*-104983	Niger	Wild Progenitor	Wild
83.	*O. glumaepatula*-103812	Venezuela	Wild Progenitor	Wild
84.	*O. glumaepatula*-105561	Colombia	Wild Progenitor	Wild
85.	*O. glumaepatula*-101960	Brazil	Wild Progenitor	Wild
86.	*O. glumaepatula*-105404	China	Wild Progenitor	Wild
87.	*O. meridionalis*-105304	Australia	Wild Progenitor	Wild
88.	*O. nivara*-89055	Laos	Wild Progenitor	Wild
89.	*O. nivara*-93185	Nepal	Wild Progenitor	Wild
90.	*O. nivara*-105410	Sri Lanka	Wild Progenitor	Wild
91.	*O. rufipogon*-105884	Bangladesh	Wild Progenitor	Wild
92.	*O. rufipogon*-106149	Laos	Wild Progenitor	Wild
93.	*O. rufipogon*-105811	Thailand	Wild Progenitor	Wild
94.	*O. rufipogon*-105883	Bangladesh	Wild Progenitor	Wild
95.	*O. rufipogon*-113651	Vietnam	Wild Progenitor	Wild

Supplementary Table 2. SSR primers information used for analysis of Mitochondrial and Chloroplast genomes.

S.No.	SSR marker	Sequence(5'-3')		Repeat motif
		Forward primers	Reverse primers	
1.	RMT 01	TTCATACGGCGGGAGTC	AGCTCTCAGACGAGCTG	(GTAG)4
2.	RMT 02	GGAACTCAGACCCGATC	ATTTATTGCCCGTCGAG	(ACA)4
3.	RMT 03	ATGGGGATCCGGTTGTG	ACAGAAAGCGTGACATG	(ATA)4
4.	RMT 04	GGTGGTTGACAAGCCAC	TTCTCTGGTACGCCGAG	(CT)6
5.	RMT 05	GTTGAAGCTTGGCAGTG	TACGAATCGCTACGCTC	(AGT)4
6.	RMT 06	GGGTTTAGAGTCGCCAC	GATGGTTTGGAAGGCTG	(AT)6
7.	RMT 07	GAGGATTTCGAGTCCTC	GAATTCTTCGAGGCCTG	(AG)6 & (TC)6
8.	RMT 08	AGAACAGAGGGAGGCTC	AACCATCCGGACGATTC	(CTC)4
9.	RMT 09	GACCAAAGGCCTTGTTG	ATAGCTCGGCTTTCGAG	(TTG)4 & (AGT)4
10.	RMT 10	AACCCAATGACGCGTTG	TTGCGTACCAACCCAAG	(ATT)4
11.	RMT 11	GCCACATAGAGCTGTCGAC	GAGCGTAGTTCTCTGGTACG	(CT)6
12.	RMT 12	TCATTACTTTGGCCACCTAAGC	GGCTTTCGTGAAAGCACC	(AT)6
13.	RMT 13	GCTCTAACCAGCCAGAACC	CCATAGAATTCTTCGAGGCCTG	(AG)6 & (TC)6
14.	RMT 14	TCTAGCCGAACGGATGC	GGTACTCAACGTTGAAGCCAC	(ACA)4
15.	RMT 15	ATGGGGATCCGGTTGTG	CCTCTTATCACACAGAAAGCGTG	(ATA)4
16.	RMT 16	TGGAGCCAAACCGAAGG	GAGAAAGCACGCCAGTG	(AGA)4
17.	RMT 17	CGATGACGTGGAACCTACC	CAAAGCCTTCCAAAGGTCC	(GAA)4
18.	RMT 18	CTGTTGAAGCTTGGCAGTG	AGAGGAAGCTTCGATTGGTG	(AGT)4
19.	RMT 19	TCCGGCTTTGGCGAATC	TCTCCACGGCAACAACG	(CTC)4
20.	RMT 20	ACCAGGGTTTGGGACAC	TCACCGTCAAGATCGCAG	(TTG)4 & (AGT)4
21.	RMT 21	TGAGAATCCGCCTTCACC	TGTGGGTATGTCGGTTGG	(AAG)4
22.	RMT 22	ACCAGGGTTTGGGACAC	CAGTTCACAAAGCACACCAAC	(TTG)4
23.	RMT 23	GCCATGTTACCACGTTCG	CCTGAGTTGTACTGGGTCG	(TAATT)3
24.	RMT 24	GCCAATGGTAAATCCGAGTTGG	TTGTCCGCCGTACTTCC	(CAATT)3
25.	RMT 25	TAGGCAAGCACCCACTC	GCTCAATCCGTTCACCAGG	RMT 25
26.	RMT 26	CCTTATATGCTCCAAGGATTCAGC	GACCAATAAGGAAAGTACACGGAG	(ATCCA)3
27.	RMT 27	CTTTATGGTACAGAAGGAGGTGAG	AAGAAGTTCCCTACCACGAAC	(TCTTA)3
28.	RMT 28	TCCTGAAGTTCTCGCGAAG	GCTTTCCTGTGTTGCACTG	(CGGGC)3
29.	RMT 29	GTTGCCAAAGAAGCTCAAACC	CACGTAAGATGAGGCCTATCC	(TAGAA)3
30.	RMT 30	TCCTGAAGTTCTCGCGAAG	AAGCCTCTGGGAAGCTG	(CGGGC)3
31.	RMT 31	TGACTTGACTGCTGGTTGAG	GGTCTTGTTGCCGTAAGC	(GGAA)3 & (AAAG)
32.	RMT 32	AGTAAAGGGTCTTCCTTTGTGG	CGAAGCTAGCTCTCAGACG	(GTAG)4
33.	RMT 33	ACTTGACAAATCGGGATTCCTC	GGAATAGAATAACCCGGTTGAAGG	(AATA)3
34.	RMT 34	AGAGGCTTCAGCTGACTTGC	CAGTGGTTTGTGGACCGATA	(CCT)3 & (AAATAA)3
35.	RMT 35	ATGTCGTCGGCGTATCG	AGATGGGTCGTTCCAGC	(GAGC)3
36.	RMT 36	GGTCCGCCTTTCTCTACTAT	TGTCTTTCTTGCATACATCG	(T)10
37.	RMT 37	ACTGCTTTTAAGCCTGTTTG	TAGGATCTCCCATTCGTAAA	(T)12
38.	RMT 38	TAACGGCTACAAGGGATAAA	GTGATGTGAGCGGTTCTATT	(G)12
39.	RMT 39	CCAAGAGAGGACAACCTGT	ATTCCTCACCTATCCTGTCA	(T)10
40.	RMT 40	CAAACTCAAGATGCCTATCC	CAACCCGGAATATTGATTTA	(T)10
41.	RMT 41	GATAAAAGATGATCCCCACA	AGTCCTTTTTCTGCTTGTTG	(A)10
42.	RMT 42	GAAAGCACCCTCTTTTTGTA	CGTGGTTATCTGAAGTGGAT	(G)11
43.	RCL 01	TGTACGAACGGCGGATG	AGTGGCAATGCACCGTG	(AT)5
44.	RCL 02	GCACTATTACGTGGCAG	GATAGTAGGAACGGCAC	(CT)5
45.	RCL 03	GTTTCCTTAGCCCACTC	GCATTCTACCCGCAATG	(TCT)4 & (T)11
46.	RCL 04	GTTAAAAGTGGCACCAATC	GATTTATGTCGTGCCAATC	(TAAA)4
47.	RCL 05	TTCATACGGCGGGAGTC	GATACGAGTCGAGGCTG	(GTAG)4
48.	RCL 06	GGGTGTAGGTAGGGCTAAAA	GACTTTTGAAAATGCGAAAT	(A)10
49.	RCL 07	TTCCTACGTGAACCAATTTT	TTCAAAGGGTTAGGTTTTTCT	(T)10
50.	RCL 08	CTTTGTTTATGCTTCGGATT	GTTCGCCTAGAGAATGACAC	(A)10
51.	RCL 09	AAACATATGCGGATCAAATC	CAACACAACATAGGTCATCG	(T)10
52.	RCL 10	TGAAGGAGGAGAAAGAAACA	TGATATCATCAACCGTGCTA	(A)10
53.	RCL 11	CATCCTTTTCAATCCAAAATCA	TGCCTGATGTAGGGAAAAGC	(A)10
54.	RCL 12	CTGGGGGGGATTATACCTGT	ATATCTCTCATTTCCGACGCA	(A)11
55.	RCL 13	TAGGCATAATTCCCAACCCA	CTTATCCATTTGGAGCATAGGG	(A)10
56.	RCL 14	ACGGAATTGGAACTTCTTTGG	AAAAGGAGCCTTGGAATGGT	(T)12
57.	RCL 15	ATTTGGAATTTGGACATTTTCG	ACTGATTCGTAGGCGTGGAC	(T)10
58.	RCL 16	GAATTTTAGAACTTTGAATTTTTTACCC	AAGCGTACCGAAGACTCGAA	(A)10
59.	RCL 17	GTGTCATTCTCTAGGCGAAC	AAATATGACAGAAAAGAAAAATAGG	(T)10
60.	RCL 18	ATAGTCAAGAAAGAGGATCTAGAAT	ACCGCGATTCAATAAGAGTA	(T)10
61.	RCL 19	ATAAGGTTATTCCCCGCTTACC	AAATTGGGGGAATTCGTACC	(T)10
62.	RCL 20	TCTTCATTTGGAATCTGGGC	CTATTGATGCAAACGCTGTACC	(A)10

RMT1 to RMT35 and RCL1 to RCL5 (Rajendrakumar *et al.*, 2007)
RMT36 to RMT42 and RCL6 to RCL10 (Nishikawa *et al.*, 2005)
RCL11to RCL 20 (Ishii & McCouch 2000)

Supplementary Table 3. Summary statistic of mitochondrial and chloroplast markers.

Marker	Major allele frequency	Allele No.	Gene diversity	PIC
RMT1	0.7474	3.0000	0.3953	0.3437
RMT2	0.8737	3.0000	0.2252	0.2076
RMT3	0.5368	3.0000	0.5845	0.5074
RMT12	0.6105	4.0000	0.5669	0.5203
RMT21	0.5474	3.0000	0.5576	0.4701
RMT34	0.9579	2.0000	0.0807	0.0774
RMT36	0.8842	3.0000	0.2070	0.1895
RMT37	0.7895	3.0000	0.3437	0.3025
RMT38	0.9368	2.0000	0.1183	0.1113
RMT40	0.9789	2.0000	0.0412	0.0404
RMT41	0.8526	2.0000	0.2513	0.2197
RMT42	0.8737	2.0000	0.2207	0.1964
RCL3	0.9579	2.0000	0.0807	0.0774
RCL6	0.5579	4.0000	0.5888	0.5228
RCL7	0.4632	4.0000	0.6201	0.5428
RCL8	0.9368	2.0000	0.1183	0.1113
RCL10	0.8421	3.0000	0.2717	0.2445
RCL18	0.6105	3.0000	0.4911	0.3894
Mean	0.7754	2.7778	0.3202	0.2819

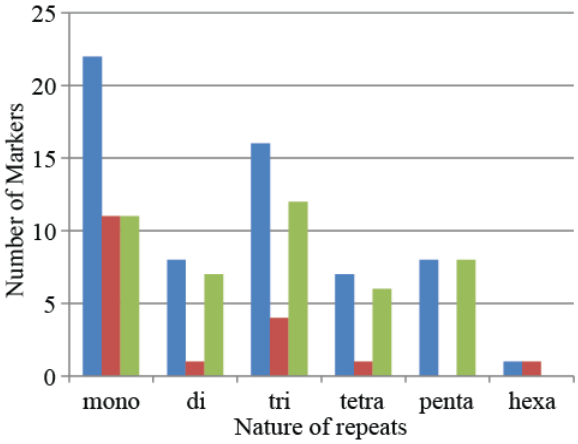

Fig. 1. Nature of repeat motif of mitochondrial and chloroplast SSR.

Allelic diversity: The markers showed a varying number of allelic richness on each locus. Out of 18 markers that showed polymorphism, four markers showed a maximum number of four alleles, while the remaining 14 markers showed two (seven markers) and three alleles (seven markers). The allelic richness ranged from 2-4 alleles with an average of 2.78 alleles. The average minor allele frequency for both mitochondrial and chloroplast genome was 0.225 while it ranged from 0.021 (RMT40) to 0.537 (RCL07).

The average gene diversity for both mitochondrial and chloroplast genomes was 0.32 by oscillating from 0.041 to 0.620. The PIC value ranged from 0.040 to 0.543 with an average of 0.282. The gene diversity for only mitochondrial markers varied from 0.0412 (RMT40) to 0.5845 (RMT3), with an average value of 0.3058, while in case of chloroplast it was 0.3168, ranging from 0.0807 (RCL3) to 0.6201 (RCl7). The PIC values ranged from 0.0404 to 0.5311, with an average of 0.2711 for mitochondrial genome while in case of chloroplast average PIC was 0.3147, ranging from 0.0744 to 0.5428. The highest PIC value was observed in RCl7 and RMT12. (Supplementary Table 3)

Cluster analysis: Phylogenetic analysis divided 95 accessions into two major groups (Fig. 2). Group-I consisted of only 11 accessions and Group-II included the rest of the accessions. In Group-I, 6 accessions belonged to Pakistan while 5 were native to the Philippines.

Wild rice from the AA genome of *Oryza* gathered into Group-II along with 68 accessions. Group-II did not have a clear grouping but was a mixture of subgroups and accessions which exactly came under a specific group. Group-II was further divided into three sub clusters. Sub cluster-IIa consisted of 23 accessions mostly developed by the International Rice Research Institute (IRRI) for bacterial leaf blight and blast resistance. Well known *indica* accessions like Vandana, PBSCR80, IR64, Brown Gora and IR6 were also part of this group. Pusa Basamti-1 was also included in this group but not in close ancestry. All wild rice was grouped into sub cluster-IIb. *O. rufipogon* from different countries were grouped together and showed a high level of parental conservation as compared to rest of the wild rice.

Genetic distance (GD) analysis: The ninety five accessions under study were divided into two major groups based on mitochondrial and chloroplast SSR. The group-I was most distantly linked with group-IIa at a GD of 0.4127 while the most closely associated groups were subgroup-IIb and subgroup-IIc which were linked at GD of 0.1497. A GD of 0.4052 was observed in group-I and subgroup-IIb while subgroup-IIc has a GD of 0.3863 with group-I. Similarly GD in subgroups of group-II was 0.3037 and 0.3168 between subgroup-IIa vs. subgroup-IIb and subgroup-IIa vs. subgroup-IIc respectively. The highest within group GD was found in group-I which was 0.4222 followed by subgroup-IIb with GD of 0.2650. The minimum within subgroup GD of 0.1846 was observed for subgroup-IIc.

The average GD ratio was 0.282 while oscillating from 0.027 to 0.65. As a whole, highest level of GD was observed in IR43450SKN-506-2-2-1-1 vs. Basmati Pak. On the other hand the lowest level of GD was observed by Vandana vs. IR 74371 -54-1-1 and NIAB-IR-9 vs. IR 78877-181 -B-1-2. IRBLa-C was found as an accession, having a maximum average GD among the others. The two wild accessions i.e. *O. glumaepatula*-105561 and *O. nivara*-89055 were found as having lowest average levels of GD (0.196).

Population structure analysis: The 95 accessions were investigated for population structure with the help of STRUCTURE software. The number of population was two when the highest log likelihood scores were obtained (Fig. 3a). These two populations were same as obtained when clustering based on genetic distance but the discrepancy was the number of accessions in each group. The alpha has a mean value of 0.0756 while the mean value of Fst for Group-I was 0.3443 and 0.2414 for Group-II.

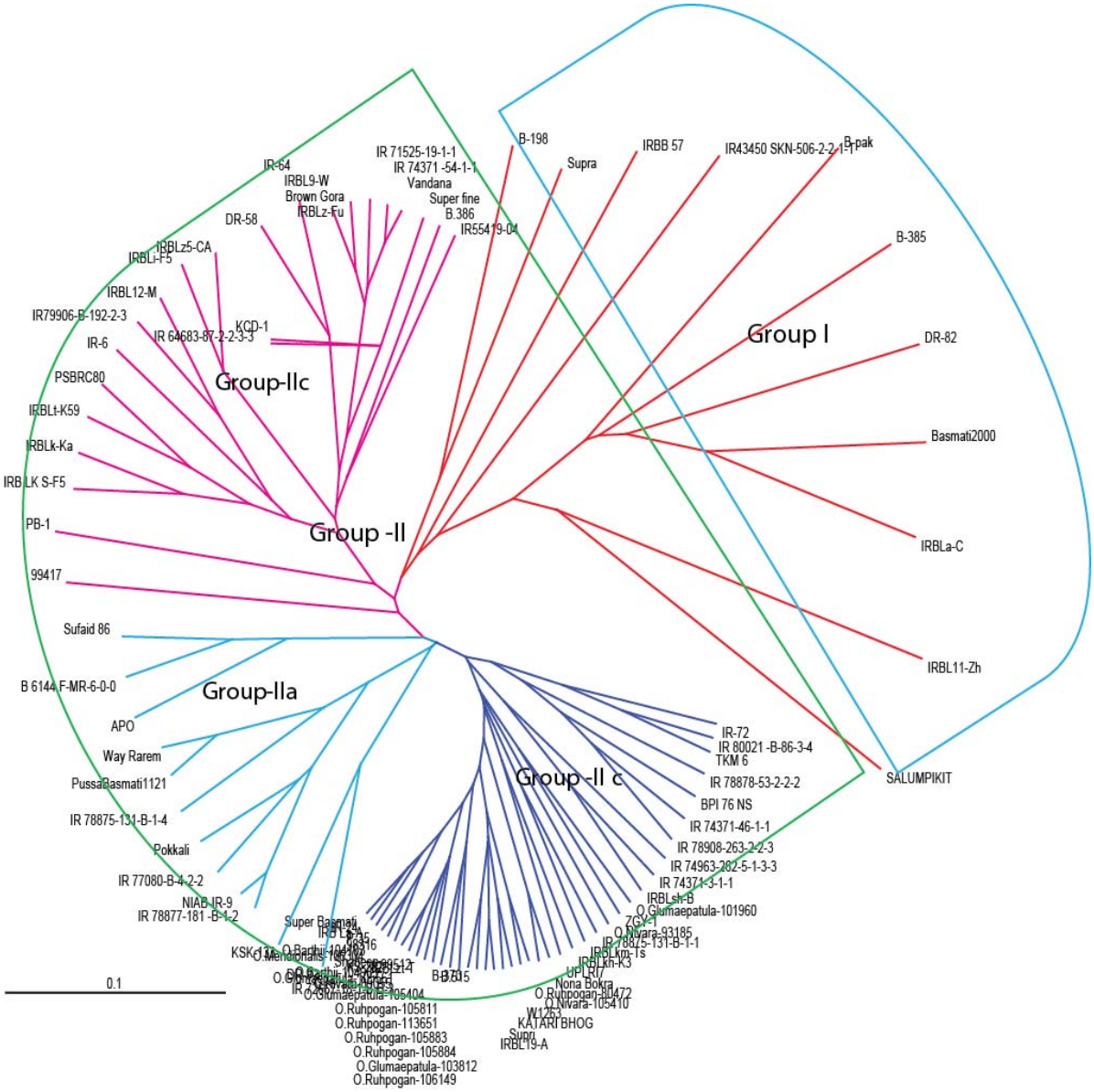

Fig. 2. UPGMA tree based on C. S. Cord (1967) ratio of individuals for mitochondrial and chloroplast variations. Each tip represents the single individual.

Mixed ancestry was observed in most accessions. Thirteen (13.68%) accessions were clearly allocated to a particular population, whereas 82 accessions (86.32%) were considered as having admixed ancestry. The percent of admix vary from 0.1% to 40% among groups. Most accessions were categorized under two groups, i.e. Group-I or Group-II. In group-I only three accessions out of twenty nine accessions showed 99.99% homozygosity while in case of Group-II only ten accessions out of sixty six were found homozygous (Fig. 3b). The accessions from different origins and types did not strictly follow their ancestry as normally observed in case of nuclear SSR analysis. All the wild rice was gathered under Group-II. The maximum mixed ancestry was observed for *O. nivara* followed by *O. glumaepatula*. The six accessions of *O. rufipogon* from different countries showed a close parentage. Almost all aromatic accessions from Pakistan and India were clustered in Group-II with the exception of Pusa Basmati-1. The typical *indica* accessions such as Vandana, IR-64, IR-6, APO and PSBCR80 came together under Group-1.

Discussion

Rice genetic diversity and population structure was studied well across the world to understand the different aspects of rice evolution (Chang, 1976; Cheng *et al.,* 2003) and improvement (Cui *et al.,* 2011, Shah *et al*., 2011; Kim *et al*., 2012). Most of those studies were based on the nuclear genome (Agrama & Eizenga, 2008; Xie *et al.,* 2012; Das *et al.,* 2013) and only a small proportion is

based on organelle genome (Guo & Ge, 2005; Duan *et al.,* 2007; Zhang *et al.,* 2012). Most of the organelle genome based studies revolve around the evolutionary path for cultivating rice and the specific location of the research group. The ultimate results of this thought result in negligence of some area and its varieties which need to be addressed. This study reports the insight into phylogenetic analysis and structure of the AA genome of cultivated rice from Pakistan and other countries along with wild rice based on organelle genome.

Our results revealed a low polymorphism rate by organelle genome. Such a lower number of polymorphism was also observed previously by Wang *et al.* (2012) which shows that both maternally inherited chloroplast and mitochondrial genomes are highly conserved across different accessions.

Our analysis generated 1 to 4 alleles with an average of 2.78. PIC value ranged from 0.040 to 0.531, with an average of 0.271 for mitochondrial genome while in case of chloroplast average PIC was 0.314, ranging from 0.074 to 0.543. This allelic diversity of organelle genome was much less as compared to nuclear genome due to its maternal inheritance in nature. The PIC values for chloroplast SSR were higher as compared to mitochondrial SSR, which reflected more diversity of chloroplast than mitochondria. Ishii & McCouch (2000) investigated ten chloroplast microsatellites in 13 diverse *O. sativa* cultivars, 19 accessions of wild rice and eight other Gramineae species. The number of alleles per markers ranged from one to four which were consistent with present results, but the PIC values range (0 to 0.710) was higher. While the average PIC value was 0.267 which was slightly less than 0.271, observed in the present study.

Nishikawa *et al.* (2005) sequenced SSR and their flanking regions in the chloroplast and mitochondrial genomes of *Oryza* species to investigate sequence variation and observed highest polymorphism for RMt17, of which 8 to 14 repeats were observed. Similarly, other plants observed in the organelle genome also supported our results for organelle genomes. Wills *et al.* (2005) investigated 5 wild *Helianthus annuus* for chloroplast variation by 36 primer pairs and detected two to five alleles at each locus. Wang *et al.* (2012) studied allo-cytoplasmic male sterile types in cabbage and amplified 32 polymorphic bands with 11 primer pairs of chloroplast and mitochondrial genome with an average 2.91 bands per primer pair.

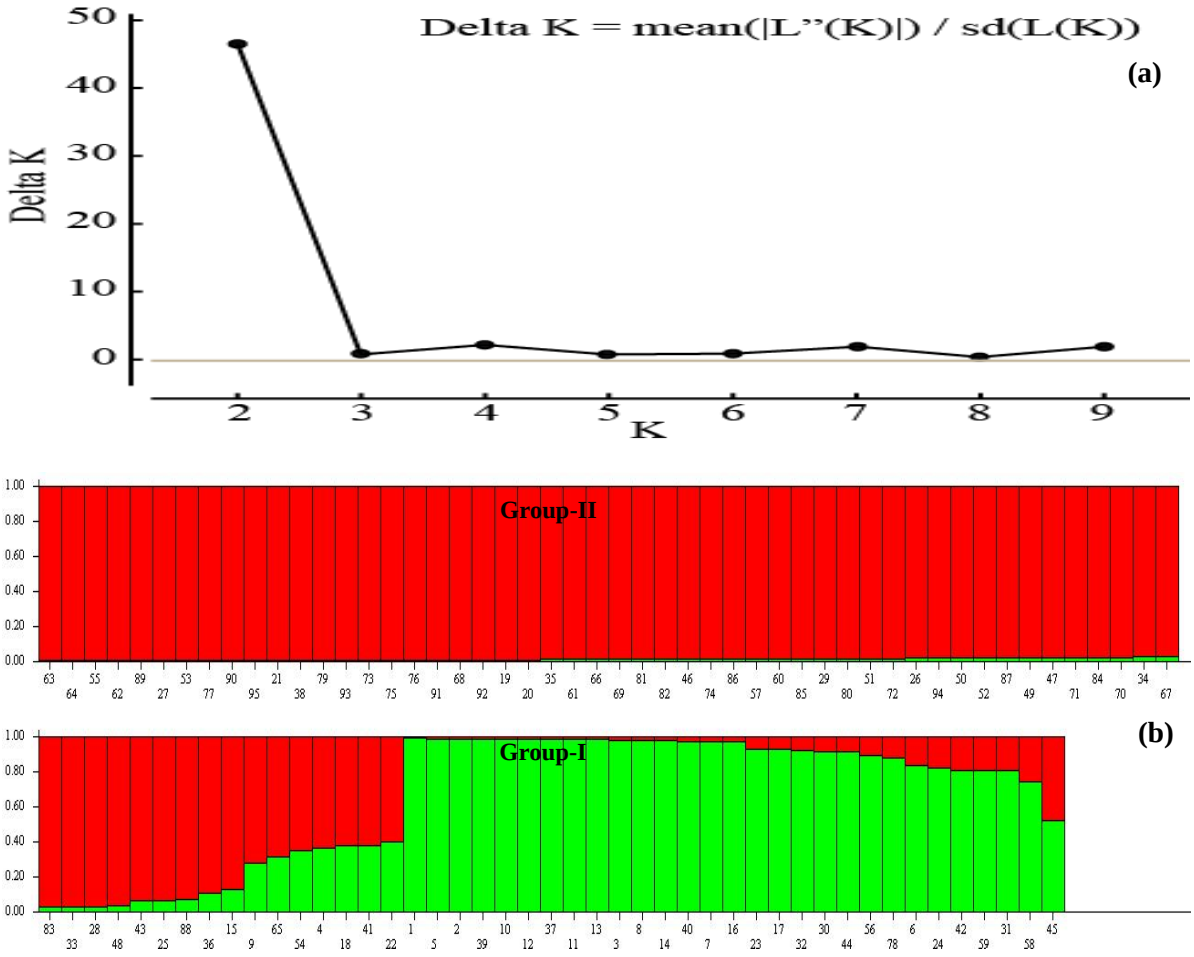

Fig. 3. Graph showing likely hood of ΔK (a) The highest log of likely hood at K=2 (b) Population structure analysis plots for mitochondrial and chloroplast genomes at K=2

In the cluster analysis of mitochondrial and chloroplast genome based on UPGMA, a close ancestry was observed between *O. rufipogon* and *O. nivara*, between *O. barthii* and *O. meridionalis* while *O. glumaepatula* did not show higher levels of linkage with any other wild rice except that the *O. glumaepatula*-105561 had close lineage or relationship with *O. nivara*-89055. A study conducted by Kumagai *et al.* (2010) to evaluate phylogeny of the genus *Oryza* by using chloroplast DNA sequences supported our results for wild rice. A contradiction between association of *O. barthii and O. meridionalis* was observed by Duan *et al.* (2007) on phylogenetic analysis done by means of DNA sequences form mitochondrial, chloroplast and nuclear genome. These differences in results might be due to type of marker used and kind of accessions under analysis.

The present investigation revealed a low gene diversity of 0.32 across the accessions. This overall lower gene diversity in spite of their diverse origin is the result of their close ancestry in case of organelle genome and their low substitution rate. This low diversity of organelle genome gives a close look into speciation of AA genome across the different countries and suggests close natural hybridization among the AA genome. These results are supported by Daun *et al.* (2007).Our results showed that the SSR marker for chloroplast and mitochondrial genome has much less resolving power and a very high number of these markers would be required when used as selectable markers.

The phylogenetic analysis and model based clustering as implemented in STRUCTURE result in two major groups. These groups do not follow the rule of nuclear genome evolution for the AA genome of Asian cultivated rice and do not separate the tested accessions into japonica and indica, the two most authenticated groups for Asian cultivated rice.

The Asian cultivated rice is still in discrepancy regarding its origin. The previous studies resulted in two contrasting hypothesis, i.e. monophyletic and diphyletic origin. The domestication and selection results into two main groups of Asian cultivated rice, Indica and Japonica. According to monophyletic hypothesis the *O. rufipogan* or *O. nivara* are the wild progenitor of indica and the japonica is the selection of indica by the farmer's community in the lateral time span (Chang, 1976, Wang *et al*., 1992). On the other hand the diphyletic origin suggests the independent evolution of both indica and japonica (Cheng *et al*., 2003). On the basis of our results the AA genome of Asian cultivated rice diverges from the same origin during evolution. These results are in accordance with the previous findings (Vaughan *et al.,* 2008; Molina *et al.,* 2011). These results also have contradictions with the finding of multiple origins of rice as investigated by Zhu & Ge (2005), Londo *et al.* (2006) and Rakshit *et al.* (2007).

Conclusion

This study overall suggests that a very large number of SSR markers are required to find out the contribution of organelle genome in the inheritance among accessions as the very low polymorphism rate was detected. The uniparental inheritance of these genome results in conservation and low substitution rate. Our findings are also in support of monophyletic origin of rice.

Acknowledgement

This work was financially supported by Higher Education Commission (HEC) of Pakistan and International Rice Research Institute (IRRI) Philippines. This study is part of PhD research work of corresponding author.

References

Agrama, H.A. and G.C. Eizenga. 2008. Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its wild relatives. *Euphytica,* 160: 339-355.

Agrama, H.A., W.G. Yan, M. Jia, R. Fjellstrom and A.M. McClung. 2010. Genetic structure associated with diversity and geographic distribution in the USDA rice world collection. *Nat. Sci.,* 2: 247-291.

Ashfaq, M. and A.S. Khan. 2012. Genetic diversity in basmati rice (*Oryza sativa* L.) germplasm as revealed by microsatellite (SSR) markers. *Russ. J. Genet.,* 48: 53-62.

Borba, T.C.O., R.P.V. Brondani, F. Breseghello, A.S.G. Coelho, J.A. Mendonça, P.H.N. Rangel and C. Brondani. 2010. Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). *Genet. Mol. Biol.,* 33: 515-524.

Cavalli-Sforza, L.L. and A.W.F. Edwards. 1967. Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.,* 19: 233-257.

Chang, T.T. 1976. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica*, 25: 425-441.

Cheng, 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.

Courtois, B., S. Negrão, G. Bruschi, E. Lupotto, N. Ahmadi, R. Greco, S. van Coppenole, C. Matos, P. Piffanelli and D. Katsantonis. 2012. Genetic Diversity and Population Structure in a European Collection of Rice. *Crop Sci.,* 52: 1663-1675.

Cui, J., B. Chen, Y. Shi, R. Zhang, H. Wang, Y. Gao, Y. Qian, H. Liu and L. Zhu. 2011. Genetic Diversity of Involved Varieties and Improvement of Elite Restorer of Indica Rice (*Oryza sativa* L.) Using Backcross Introgression. *Mol. Plant Breed.,* 6(1): 25-31.

Das, B., S. Sengupta, S.K. Parida, B. Roy, M. Ghosh, M. Prasad and T.K. Ghose. 2013. Genetic diversity and population structure of rice landraces from Eastern and North Eastern States of India. *BMC Genet.,* 14: 71.

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

Duan, S., B. Lu, Z. Li, J. Tong, J. Kong, W. Yao, S. Li and Y. Zhu. 2007. Phylogenetic analysis of AA-genome Oryza species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. *Biochem. Genet.,* 45: 113-129.

Earl, D.A. and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.,* 4(2), 359-361.

Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, 14: 2611-2620.

Falush, D., M. Stephens and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics,* 164: 1567-1587.

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

Ishii, T. and S.R. McCouch. 2000. Microsatellites and microsynteny in the chloroplast genomes of Oryza and eight other Gramineae species. *Theor. Appl. Genet.,* 100: 1257-1266.

Ishii, T., Y. Xu and S.R. McCouch. 2001. Nuclear-and chloroplast-microsatellite variation in A-genome species of rice. *Genome*, 44: 658-666.

Jin, L., Y. Lu, P. Xiao, M. Sun, H. Corke and J. Bao. 2010. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor. Appl. Genet.,,* 121: 475-487.

Kim Yoon-Ha, A.L. Khan, Z.K. Shinwari, Duck-Hwan Kim, M. Waqas, M. Kamran and In-Jung Lee. 2012 . A Silicon Treatment to Rice (*Oryza sativa* L. Cv 'Gopumbyeo') Plants During Different Growth Periods and its Effects on Rice Growth and Grain Yield. PAK. *J. Bot*., 44(3): 891-897

Kumagai, M., L. Wang and S. Ueda. 2010. Genetic diversity and evolutionary relationships in genus Oryza revealed by using highly variable regions of chloroplast *DNA. Gene,* 462: 44-51.

Li, X., W. Yan, H. Agrama, L. Jia, X. Shen, A. Jackson, K. Moldenhauer, K. Yeater, A. McClung and D. Wu. 2011. Mapping QTLs for improving grain yield using the USDA rice mini-core collection. *Planta,* 234: 347-361.

Liu, K. and S.V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21: 2128-2129.

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

Lu, H., M.A. Redus, J.R. Coburn, J.N. Rutger, S.R. McCouch and T.H. Tai. 2005. Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. *Crop Sci.*, 45: 66-76.

Molina, J., M. Sikora, N. Garud, J.M. Flowers, S. Rubinstein, A. Reynolds, P. Huang, S. Jackson, B.A. Schaal and C.D. Bustamante. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Nati. Acad. Sci.,* 108: 8351-8356.

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

Pervaiz, Z.H., M.A. Rabbani, Z.K. Shinwari, M.S. Masood and S.A. Malik. 2010. Assessment of genetic variability in rice (*Oryza sativa* L.) germplasm from Pakistan using RAPD markers. *Pak. J. Bot*., 42: 3369-3376.

Powell, W., M. Morgante, R. McDevitt, G.G. Vendramin and J. Rafalski. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc. Nati. Acad. Sci.,* 92: 7759-7763.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959.

Rabbani M.A., M.S. Masood, Z.K. Shinwari and K.Y. Shinozaki. 2010. Genetic Analysis of Basmati and Non-Basmati Pakistani Rice (*OryzaSativa* L.) Cultivars using Microsatellite Markers. *Pak. J. Bot*., 42(4): 2551-2564

Rajendrakumar, P., A.K. Biswal, S.M. Balachandran, K. Srinivasarao and R.M. Sundaram. 2007. Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics,* 23: 1-4.

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

Shah, S.M., S.A. Naveed and M. Arif. 2013. Genetic diversity in basmati and non-basmati rice varieties based on microsatellite markers. *Pak. J. Bot.*, 45(SI): 423-431.

Shah, S. M., M. Arif, K. Aslam, G. Shabir and M. Thomson (2015) Genetic diversity analysis of Pakistan rice (*Oryza sativa*) germplasm using multiplexed single nucleotide polymorphism markers. *Genet. Resour Crop Evol.,* 1-14. DOI: 10.1007/s10722-015-0304-9.

Shah, S.M.A., H. Rahman, F.M. Abbasi, M.A. Rabbani, I.A. Khan and Z.K. Shinwari. 2011 Inter specific variation of total seed protein in wild rice germplasm using SDS-PAGE. *Pak. J. Bot*., 43(4): 2147-2152

Taberlet, P., E. Coissac, F. Pompanon, L. Gielly, C. Miquel, A. Valentini, T. Vermat, G. Corthier, C. Brochmann and E. Willerslev. 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.,* 35: e14-e14.

Vaughan, D.A., B.R. Lu and N. Tomooka. 2008. The evolving story of rice evolution. *Plant Science*, 174: 394-408.

Wang, Q., Y. Zhang, Z. Fang, Y. Liu, L. Yang and M. Zhuang. 2012. Chloroplast and mitochondrial SSR help to distinguish allo-cytoplasmic male sterile types in cabbage (*Brassica oleracea* L. var. capitata). *Mol. Breed.,* 30: 709-716.

Wills, D.M., M.L. Hester, A. Liu and J.M. Burke. 2005. Chloroplast SSR polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus*. *Theor. Appl. Genet.,* 110: 941-947.

Xie, F., L. Guo, G. Ren, P. Hu, F. Wang, J. Xu, X. Li, F. Qiu and M.A. dela Paz. 2012. Genetic diversity and structure of indica rice varieties from two heterotic pools of southern China and IRRI. Plant Genet. *Resour. Charact. Util.,* 10: 186-193

Zhang, T., S. Hu, G. Zhang, L. Pan, X. Zhang, I. S. Al-Mssallem and J. Yu. 2012. The Organelle Genomes of Hassawi Rice (*Oryza sativa* L.) and Its Hybrid in Saudi Arabia: Genome Variation, Rearrangement, and Origins. *PloS one,* 7: e42041.

Zhao, W., J. W. Chung, K. H. Ma, T. S. Kim, S. M. Kim, D. I. Shin, C. H. Kim, H. M. Koo and Y. J. Park. 2009. Analysis of genetic diversity and population structure of rice cultivars from Korea, China and Japan using SSR markers. *Genes & Genomics,* 31: 283-292.

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

(Received for publication 20 October 2014)