INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
18–1574/201x/00–0–000–000
DOI: 10.17957/IJAB/15.1003
http://www.fspublishers.org

Full Length Article

Comparative Analysis of Genomic Variation and Population Structure of Rice (*Oryza sativa* L.)

Kashif Aslam[1], Ghulam Shabir[1], Muhammad Shahid[2], Muhammad Sabar[3], Shahzad Amir Naveed[4], Shahid Masood Shah[5*], Abdul Rehman Khan[5] and Muhammad Arif[6]

[1]*Institute of Molecular Biology and Biotechnology, BZU, Multan, Pakistan*
[2]*Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan*
[3]*Rice Research Institute, Kala Shah Kaku, Lahore, Pakistan*
[4]*Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China*
[5]*Department of Environmental Sciences, COMSATS University Islamabad, Abbottabad Campus, Pakistan*
[6]*National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, 38000, Pakistan*
[*]For correspondence: smasood@cuiatd.edu.pk; smasood.nibge@gmail.com

Abstract

Studying genomic variation and its understanding in cultivated rice (*Oryza sativa* L) can help in improving the crop yield by marker assisted selection. The current study was focused on finding the genomic variations among existing cultivated genotypes of *Oryza sativa* from different countries. Moreover, the study also compares the genomic variation found in Pakistani and International rice genotypes. A total of 95 genotypes were collected from sixteen countries, mainly from Pakistan, India and the Philippines. Evenly distributed 200 Simple Sequence Repeat (SSR) markers were surveyed and 142 SSR markers were found polymorphic, with an average of 4.53 alleles per locus while alleles per marker ranging from 2 to 14. Neighbor joining cluster analysis based on genetic dissimilarity coefficients clearly separated all the genotypes into three groups but being unable to separate the genotypes based on their geographical area reflected their single origin. The IRBL lines from IRRI being part of the aromatic group are the incongruity of results which reflects conserved genome for blast resistance in basmati group. The model-based structure and principal component analysis are in harmony with neighbor joining results. Overall results suggested a narrow genetic variation among rice genotypes taken from different countries. This information obtained from the investigated genotypes will provide a baseline to incorporate preferred variation in future breeding programs to tailor rice genotypes with better stand against unforeseen biotic or abiotic stresses.

Keywords: Genetic diversity; Phylogenetic analysis; Principal component analysis; Rice; SSR markers

Introduction

Rice (*Oryza sativa* L.) is grown in all continents except for Antarctica and is among the most important food crops. Presently, the world population consumes 503.0 million tons of rice per year (FAO, 2018). Among the major crops, rice has been subjected to ex situ conservation after recognition of plant genetic diversity as critical for improvement of agronomic traits (Liu *et al.*, 2015). Exploring diversity among the landraces results in genes/alleles identification that can be part of crop improvement regarding better yield and resistance (Thomson *et al.*, 2007). A study based on domesticated Asian rice landraces with the help of Simple Sequence Repeats (SSRs) markers resulted in a classification of these landraces into five major groups indica, aus, temperate japonica, tropical japonica and aromatic (Garris *et al.*, 2005).

The SSR markers are tandemly arranged repeats of 1-6 bp short DNA motifs which at a locus often exhibit variation. SSR markers are admired for evaluating genetically related rice genotypes as being co-dominance and highly polymorphic (Shabir *et al.*, 2017). In rice, thousands of SSR have been identified and developed SSR markers exceeds over 25,000 (Orjuela *et al.*, 2010). These are acknowledged for assessing diversity and population structure (Shah *et al.*, 2013; Jasim Aljumaili *et al.*, 2018), marker assisted breeding (Luo *et al.*, 2016), QTL mapping (Galeng-Lawilao *et al.*, 2018) and association mapping (Liu *et al.*, 2017).

Pakistan has a unique climate for basmati rice landraces with distinctive characteristics of basmati fragrance. During the last few decades for the conservation of basmati traits; selected basmati lines were used as parents resulting in narrowing of the genetic base of the crop as they all are genetically related. On the contrary, the diversity analysis of landraces of Pakistan revealed diverse genetic base. So, conservation of the landraces is obligatory along with their future utilization in the breeding of high yielding, superior quality and better-adapted cultivars in the country (Pervaiz *et al.*, 2010).

To cite this paper: Aslam, K., G. Shabir, M. Shahid, M. Sabar, S.A. Naveed, S.M. Shah, A.R. Khan and M. Arif, 201x. Comparative analysis of genomic variation and population structure of rice (*Oryza sativa* L.). *Intl. J. Agric. Biol*. 00: 000–000

Recently, drastic climatic variations led to an uncertain situation in agricultural countries, that also includes Pakistan, resulting in pressure building on the crop breeding community to develop cultivars that can sustain harsh environmental conditions (Saif, 2017). Therefore, it is mandatory to explore the genetic makeup of the existing cultivars for future selection and breeding purposes. All the studies conducted so far in Pakistan for diversity assessment in rice used a relatively smaller number of genotypes and markers (Pervaiz *et al.*, 2010; Rabbani *et al.*, 2010; Shah *et al.*, 2013). Till date, none of the studies has compared the existing cultivated diversity with cultivars from other agricultural countries, therefore, the present study was designed for comparative analysis of genomic variation and population structure of *Oryza sativa* L. from Pakistan and other countries of the world. The information obtained from these investigated genotypes will provide a baseline to incorporate preferred variation in future breeding programs to tailor more diverse rice genotypes.

Materials and Methods

Ninety-five rice genotypes were analyzed with the help of 200 SSR markers. These genotypes include lines, cultivars, landraces and wild rice populations (Supplementary Table 1). The genotypes relating to Pakistan were collected from Plant Genetic Resource Institute (PGRI), National Agriculture Research Centre (NARC), Islamabad and Rice Research Institute (RRI), Kala Shah Kaku, Lahore. International Rice Research Institute (IRRI), Philippines, being the world largest seed bank for rice, serve as a source for rice genotypes other than Pakistan including wild rice.

Genomic DNA Extraction

Twenty seed of each rice genotype was sown in lines on one m^2 raised bed surrounded with drainage channels to avoid submergence. From each genotype, after twenty-one-days, five seedlings were selected for total genomic DNA extraction. DNA was extracted with the help of CTAB method as described by Doyle (1987). The DNA quality and quantity were examined with the help of agarose gel electrophoresis and NanoDrop® ND1000 and followed by template dilution to 20 ηg/μL.

PCR Amplification

PCR was performed for genotyping by utilizing SSR markers. PCR reaction was dispensed in 96 well plate. A total volume of 15 μL reaction mixture was used. Following were the component, DNA template (2 μL), $MgCl_2$ (1 μL of 25 m*M*), PCR buffer (1 μL of 10X), dNTPs (1 μL of 1 m*M*), primers (0.5 μL of 5 μM) and Taq DNA polymerase (0.25 μL of 10 U/μL) and sterile nano-pure H_2O (8.75 μL). To prevent evaporation one drop of mineral oil was added. The PCR plate was placed inside the thermal cycler (G-Storm, Syrrey, UK).

The following PCR profile was used: Initial denaturation was done at 94ºC for 5 min at once followed by denaturation at 94ºC for 30 sec for 35 cycles, annealing temperature ranged from 55ºC to 65ºC for 30 sec followed by extension at 72ºC for 30 sec, a final extension was done for 5 min at 72ºC. The product was stored at 4ºC.

PCR Product Resolution and Detection

To resolve the PCR products 6% polyacrylamide gel electrophoresis (PAGE) was used. An electric field was applied to the tank with a voltage of 120 volts for ninety minutes. Polyacrylamide gel was separated from the glass plates and soaked under 10,000 times dilution Sybr® Safe staining solution in a dark box for 25 min. Amplified DNA products resolved on the gel were illuminated through the UV light cabinet and the image was captured with the viewing software, AlphaImager®. The bands on the gel were scored from lowest to highest as A/A, B/B up to N/N for each of the genotypes.

Data Analysis

A V3.25 of Power Marker developed by Liu and Muse (2005) (http// www.powermarker.net) was used to calculate allele number, major allele frequencies, polymorphic information content, and genetic diversity. Genetic distance was calculated based on the "C.S. Chord 1967" (Cavalli-Sforza and Edwards, 1967). An unrooted Neighbor-Joining phylogenetic tree was constructed by applying to bootstrap at 1,000 iterations in PowerMarker V3.25. TREEVIEW was used to view the phylogenetic tree.

Principal component analysis (PCA) was performed with the help of NTSYSpc version 2.10t (Rohlf, 2000). The PCA is based on covariance matrices. The NTSYSpc was based on binary format i.e. for presence or absence of bands. The allele frequency data of PowerMarker was transferred to NTSYSpc as in binary format. Similarity matrix with the SIMQUAL subprogram using Dice coefficient was calculated and used for PCA with DCENTER, EIGEN and MXPLOT subprograms in NTSYSpc.

In addition, model, based STRUCTURE (Pritchard *et al.*, 2000) was used to infer population structure for admixture and correlated allele frequencies using a 20,000 burn in length at run length of 10,000. The model was tested for K = 2 to K = 10 for five independent runs (Falush *et al.*, 2003). Evanno *et al.* (2005) parameters were calculated by using Structure Harvester V6.0 (Earl, 2012).

Results

A total of 200 SSR markers were used by covering the 12 chromosomes. The highest number of SSR markers were 22 distributed on chromosome 2 and the lowest number of SSR markers 7 were found distributed on chromosome 11 (Fig. 1). Among the 200 SSR markers, 58 were found monomorphic and were excluded from data analysis.

Remaining 142 SSRs revealed 644 alleles among 95 genotypes. A variable number of alleles per locus were observed for each marker and number of alleles per primer ranged from 2 to 14 alleles with an average of 4.53 alleles per locus. The minor allele frequency ranged from 0.0105 (RM403 and RM344) to 0.7263 (RM481) with an average of 0.354. The average gene diversity was 0.41, varying from 0.0208 (RM403 and RM344) to 0.8295 (RM404). Polymorphic information content (PIC) is the comparative measure of how much a marker is informative to distinguish between different populations and it depends upon the number of alleles detected by this marker and their relative frequency. Its values ranged from 0.0206 (RM403 and RM344) to 0.8103 (RM404), with an average of 0.368 (Supplementary Table 2).

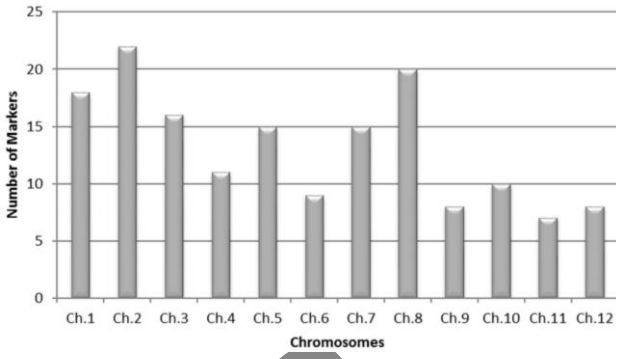

Fig. 1: Graphical presentation of the number of SSR markers per chromosome

Phylogenetic Analysis

In the phylogenetic analysis based on C.S. Cord distance at 1000 bootstrap divided 95 genotypes into three major groups. Group-I (G-I), composed of only *indica* genotypes, Group-II (G-II), composed of *indica* from India and Philippines, *aromatic* from Pakistan and India and *japonica* genotypes form the Philippines, while Group-III (G-III), comprised of wild rice. Furthermore, two genotypes were shown admixture between other groups. Three major groups formed by NJ grouped 95 genotypes at 100% confidence interval. The *aromatic* and *japonica* rice cultivars were effectively differentiated from non-aromatic and wild rice genotypes (Fig. 2 and Table 1).

In Group-I, 41 genotypes were clustered together, which again split up into seven subclusters (G-Ia to G-Ig. Group-II was further subdivided into two subgroups. G-IIa and G-IIb. G-IIa consisted of aromatic, IRBL genotypes from IRRI and *Japonica* group with two major clusters comprising of 22 genotypes. The G-IIb mainly clustered IRBL genotypes from IRRI along with Indian. This group consisted of 14 genotypes.

Wild rice bunched into Group-III, which further divided into two subclusters (G-IIIa and G-IIIb). G- IIIa combining *O. rufipogon* and *O. nivara*, while the other G-IIIb shared *O. barthii, O glumaepatula* and *O. meridionalis*. However, aromatic and genetically related genotypes were separated from those of non-aromatic groups. The non-aromatic *indica* group was clustered far away from *japonica* genotypes to aromatic cultivars. Dendrogram showed that the genotypes clustered together were the progeny of genetically similar types. Genotypes sharing a high percentage of lineages came under the same subgroup.

Genetic Distance Analysis for 95 Genotypes

All genotypes were also analyzed for relatedness on the basis of the ratio of shared SSR fragments by calculating a genetic distance matrix. The average genetic distance coefficient ranged from 0.3715 to 0.9987. In this assay, G-I and G-II

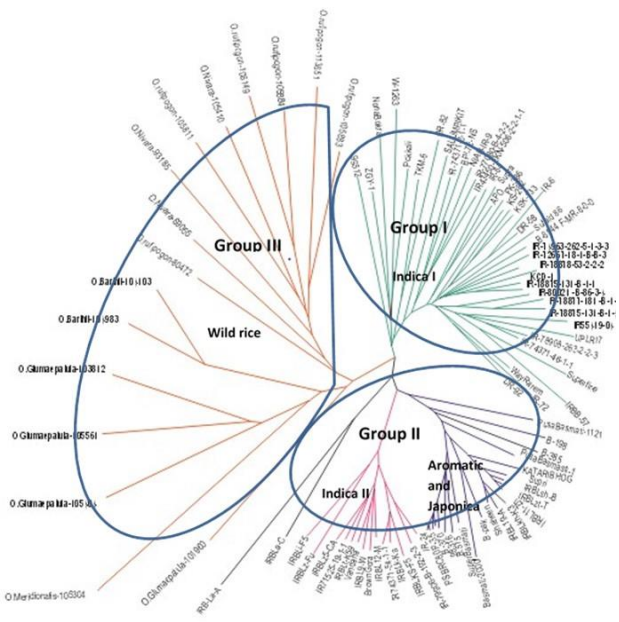

Fig. 2: An unrooted NJ tree showing the genetic relationships between the 95 rice genotypes based on 159 microsatellite markers

were closely related with 0.3639 GD, while G-I and G-III was most extensively divergent with a 0.5040 GD. For analysis of within group GD, the highest level was observed in G-II (0.226) followed by G-I (0.3006) and G-III (0.4482) (Table 2).

The GD ratio in rice genotypes revealed an average of 0.369 for all genotypes. The maximum degree of GD was 0.63 between *O. meridionalis*-105304 (G-III) vs. IRBB-57 (G-I). On the other hand, the minimum GD ratio of 0.01 was observed in P-35 vs. IR-24. A genetic distance of 0.62 was observed in *O. meridionalis*-105304 vs. IR55419-04. The same GD was also observed in *O. meridionalis*-105304 vs. ZGY-1. Overall *O. meridionalis*-105304 had shown the highest degree of GD ratio among all the genotypes under study with an average of 0.560. The lowest level of GD was found in P-35 vs. IR-24 followed by IRBLkm-Ts vs.

Table 1: Rice cultivars classification into indica, japonica or aromatic based on simple sequence repeat variations

Group	Type	Sub-cluster	Cultivars Names	Total Cultivars
Group I	*Indica* I	1	Sufaid-86, IR-72, DR-92	3
		2	IR-64683-87-2-2-3-3, IR-74963-262-5-1-3-3, IR-72667-18-1-B-B-3, IR-78878-53-2-2-2, KCD-1, IR-78875-131-B-1-1, PK-386, KSK-133, KS-282, IR-6, DR-58	11
		3	B-6144-F-MR-6-0-0, IR-64, Supra, APO	4
		4	IR-77080-B-4-2-2, IR-74371-3-1-1, BPI-76-NS, NIAB-IR-9, IR-43450-SKN-506-2-2-1-1, Salumpikit, line-99512	7
		5	ZGY-1, Nona Bokra, W-1263, Pokkali	4
		6	TKM-6, DR-82	2
		7	Way Rarem, IR-80021-B-86-3-4, IR-78877-181-B-1-2, IR-78875-131-B-1-4, IR-55419-04, UPLRI-7, IR-78908-263-2-2-3, IR-74371-46-1-1, Superfine, IRBB-57	10
Group II	Aromatic and *Japonica*	1	PusaBasamti-1121, B-198, B-385, PusaBasmati-1	4
		2	Katari Bhog, Supri, IRBLzt-T, IRBLsh-B, IRBLkm-Ts, IRBL11-Zh, IRBLkh-K3, IRBL19-A, Shaheen Basmati, B.Pak,	10
		3	line-99417, B-2000, B-370, Super Basmati, line-98316, B-515, P-35, IR-24,	8
	Indica II	4	IR-74371-54-1-1, Brown Gora, IRBLt-K59, IRBL9-W, IR-71525-19-1-1, Vandana, IRBLz-Fu, PSBRC80, IR-79906-B-192-2-3, IRBLKS-F5, IRBLk-Ka, IRBL12-M, IRBLz5-CA, IRBL1-F5	14
	Intermediate		IRBLa-C, IRBLa-A	2
Group III	Wild rice	1	*O. glumaepatula, O. meridionalis, O. barthii*	7
		2	*O. rufipogon, O. nivara*	9

Table 2: Genetic distances between groups based on 142 SSR analyses

Group	1	2	3
1	0.3006		
2	0.3639	0.2266	
3	0.5040	0.4852	0.4482

IRBLsh-B, IRBLsh-B vs. IRBLzt-T and P-35 vs. Bas-515. IRBLzt-T genotype showed the lowest level of GD ratio among all the genotypes ranging from 0.022 to 0.526.

Geographical Analysis of Diversity

The 142 SSR markers exhibited 370 alleles among 21 genotypes from Pakistan with an average of 2.605 alleles per locus, while among 11 genotypes from India, exhibited 364 alleles with an average of 2.563 alleles per locus. In the case of the Philippines, among the 44 genotypes, 142 SSR markers exhibited 438 alleles with an average of 3.084 per locus. The mean gene diversity for Pakistan, India and Philippines rice genotypes was 0.359, 0.401 and 0.378 for each SSR locus respectively.

The genotypes from Pakistan, India and the Philippines showed mean PIC value of 0.312, 0.347 and 0.331 respectively. The comparison of genetic diversity of the three countries revealed that the Philippines genotypes possessed a maximum number of mean alleles, while the Indian genotypes had higher values for gene diversity and PIC (Fig. 3). Based on region related tendency, the following genetic diversity order was observed; in mean alleles per locus the order was Philippines>Pakistan>India, while in case of average gene diversity and PIC value the order found was India>Philippines>Pakistan (Table 3).

Population Structure Analysis

In model-based clustering, comprising of 95 genotypes surveyed by 142 SSR markers, the analysis based on genetic distance as well as population structure revealed that rice genotypes under study had significant population structure. The highest log likelihood score was calculated when the number of populations was set to three (Fig. 4a).

Results of population structure were in accordance with NJ analysis based on genetic distance. The mean value of alpha at K=3 was 0.717, where K is the number of population. The fixation index (Fst) is a measure of population differentiation due to genetic structure. Fst is based on the variance of allele frequencies between populations. The Fst values found for the G-I, G-II and G-III was 0.053, 0.683 and 0.531 respectively. Forty-five (47.37%) genotypes were undoubtedly allocated to a single population, whereas, fifty genotypes (52.63%) under study had admixed ancestry. The admixed genotypes were mostly identified as an admixture between *indica* and aromatic groups, while the remaining admixtures were amalgamations of other groups.

Specifically, at K=2, structure analysis separated all genotypes into two main varietal groups. G-I consisted of thirty-six genotypes, while G-II comprised of fifty-nine genotypes. Wild rice was an unpredictable mixture of *indica* and *japonica* and laid in G-II. When K was increased from two to three, G-I remained the same,

Table 3: Summary statistics of cultivars from different countries as revealed by SSR marker analysis

	Cultivars	pSSR	MaAF	MAF	Avg. Alleles	GD	PIC
All	95	142	0.69	0.31	3.956	0.404	0.361
Pakistan	23	127	0.736	0.264	2.692	0.334	0.289
Philippines	44	127	0.736	0.264	2.692	0.334	0.289
India	10	117	0.728	0.272	2.277	0.345	0.295
Wild + Others	18	139	0.619	0.381	3.358	0.477	0.429

Here pSSR = Polymorphic SSR; MaAF = Major allele Frequency; MAF = Minor Allele Frequency; Avg. Alleles = Average Number of Alleles; GD = Gene Diversity; PIC = Polymorphic Information Content

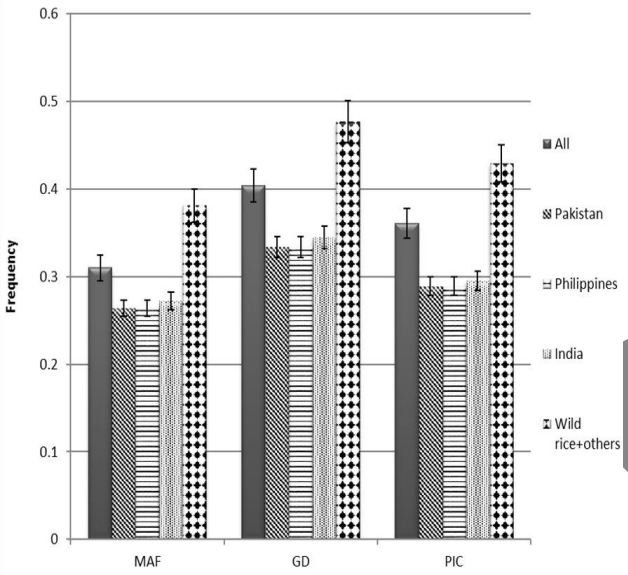

Fig. 3: Graphical presentation of minor allele frequency (MAF), genetic diversity (GD) and PIC value among different countries. Error bar represents the standard error of means (n=3)

while the G-II split into two subgroups and each new cluster corresponded to one of the two main populations, *indica* and wild rice. Aromatic genotypes from Pakistan and India laid in G-I along with IRBL genotypes from IRRI Philippines. Non- aromatic genotypes from Pakistan were included in G-II (Fig. 4b).

Principal Component Analysis (PCA)

The PCA analysis resulted in almost the same groupings among genotypes as identified by the NJ and STRUCTURE analysis except that the PCA amalgamated *indica*-II and aromatic in one group (Fig. 5). PCA resulted in three major clusters. C-I comprised of all *indica* genotypes, while C-II was made with Basmati and *indica*-II, whereas wild rice belonged to C-III.

Unlike the NJ analysis, in which all genotypes were allocated to a group, except IRBLa-A and IRBLa-C, six other genotypes were not in closeness with other groups in PCA. These are from *indica*-I (DR-82, Nona Bokra and W-1263), Aromatic group (Pusa Basmati-1121 and B-198) and from

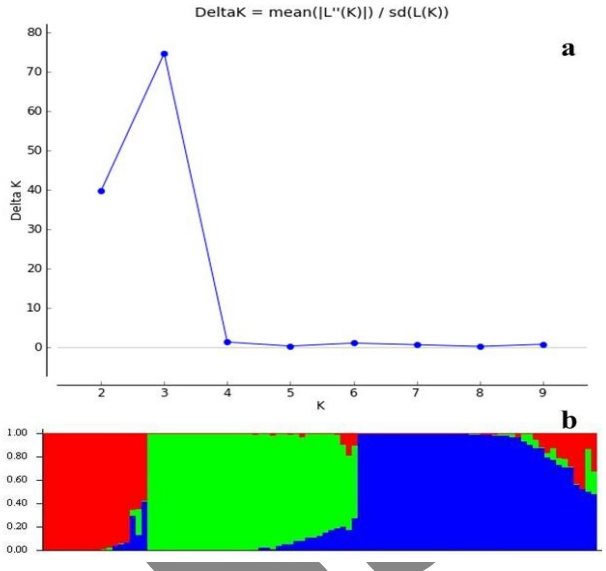

Fig. 4: Population structure analysis of 95 genotypes based on 142 SSR markers showing (a) value of LnP (D) for determining the optimum number of subpopulation for total population (b) Bar plot showing groups at different values of K

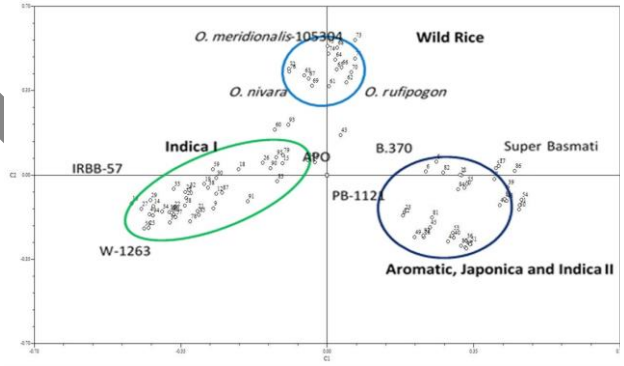

Fig. 5: Two-dimensional scaling resulting from principal component and coordinate analysis (NTSYSpc) of 95 genotypes using genetic diversity data for 644 alleles at 142 simple sequence repeat loci. Circles *indica*te three major clusters formed

wild rice (*O. rufipogon*-80475). These genotypes seemed to be little different in genetic makeup from the rest of genotypes. Out of 95 genotypes, 38 were included in C-I, which had a resemblance with G-I of NJ.

G-II of the dendrogram had a close relationship with the C-II detected in the PCA. A total of thirty-four genotypes were included in the second cluster. The famous aromatic rice from Pakistan and India were also a part of this C-II. Two genotypes from Philippines (IRBLa-A and IRBLa-C) were separated from the other genotypes and laid in between Basmati, *indica* and wild groups. These findings were consistent among the results obtained in the PCA and hierarchical analysis. The remaining 16 wild rice originating from different countries formed a separate cluster. The wild

rice cluster was comparable with C-III of the dendrogram, except for *O. rufipogon*-80475, which was separated. Wild rice genotypes originated long before than other cultivated genotypes and had diverse alleles as compared to other genotypes under investigation. These wild rice genotypes were not close to each other as the other genotypes did in the other two groups.

Discussion

The genetic diversity is crucial for existence of a genotype in a defined environment. It is one of the key components that ultimately contributes to the survival of genotypes against biotic and biotic stress. Crop improvement in term of yield or stress withstand is the prime objective of breeders. This objective could be only possible by exploring a variation on phenotypic as well as on genetic level and incorporating the desirable traits in upcoming cultivars through breeding. In the current study, 95 genotypes were evaluated by 200 SSR markers which were distributed across 12 chromosomes. PIC values determine the effectiveness of a marker to differentiate within and between groups. PIC values varied considerably for studies SSR loci and were similar to other reports (Wang *et al.,* 2014; Salgotra *et al.,* 2015; Ekka *et al.,* 2016; Anupam *et al.,* 2017). An average PIC value observed by Singh *et al.* (2013, 2016) and Shah *et al.* (2015) were lower than reported in this study. A higher PIC value was observed by Lee *et al.* (2011), Upadhyay *et al.* (2012) and Roy *et al.* (2016) This difference in average PIC value with other researcher might be associated with a diverse set of cultivars under analysis and the choice of different markers.

PIC value is highly dependent upon the number of alleles detected. The highest number of alleles observed was 14 which was higher than reported by Huang *et al.* (2010), Pervaiz *et al.* (2010) and Upadhyay *et al.* (2011) by using a different set of rice germplasm. However, other researchers obtained a higher number of alleles such as 21, 39 and 16 (Shu *et al.,* 2009; Zhao *et al.,* 2009; Chuang *et al.,* 2011). These paradoxes amongst reports might be the results of different number, origin or countries of genotypes under analysis. The second probable reason might be depending upon the type of microsatellite primers *i.e.*, mono-, di-, tri-, etc. as reported in different studies. The third reason might be the scorable alleles across all genotypes. The finding of this study was consistent with that mentioned above. The multiple natures of the markers involved in the present investigation might have also been added to potential diversity, as the discrete repeats involved are prone to DNA slippage and size mutations. These findings also designated that SSR markers are highly explanatory and dominant to evaluate the genetic diversity of rice.

The average genetic distance coefficient ranged from 0.3715 to 0.9987. Other studies based on SSR markers observed varying levels of genetic distance comparable to our observations. Joshi *et al.* (2010), Behera *et al.* (2012) observed genetic similarity coefficient from 0.239 to 0.827 and 0.041-0.728 respectively. This low genetic diversity in cultivated rice is not native to Pakistan only. The low genetic diversity was observed in rice germplasm of Taiwan (Lin *et al.,* 2012). Similarly, on analysis of genetic diversity and population structure of landraces of Eastern and Northeastern state of India more diverse nature of landraces were observed from West Bengal than other states (Das *et al.,* 2013). This low genetic divercification is the results of high preferences of breeders and farmers for a desirable traits and utilization of a superior genotype for breeding from last couple of decades.

The genotypes were grouped into three groups based on their genetic distance. The two genotypes from IRRI, (IRBLa-A and IRBLa-C) were separated and placed in-between Basmati, *indica* and wild groups which may be due to their composite ancestry from diverse origin. The same results for these genotypes were obtained from the study of different sets of popular rice cultivars from India, including cultivars released by IRRI (Upadhyay *et al.,* 2012). In reality, wild rice genotypes studied in current investigation were found to have diverse alleles as compared to the other genotypes. Moreover, diverse wild rice genotypes were found different from each other as compared to cultivated cultivars.

At the selected core set the model-based structure analysis exposed the occurrence of three populations. Different studies assessed the genetic structure of rice to find out its origin, diversity, and domestication. Chakhonkaen *et al.* (2012) observed that highest likelihood of data ΔK was obtained when K=6, Huang *et al.,* 2010 observed the most suitable number of clusters was identified at K=4. All the studies are done, so far, with different types of markers for analysis with STRUCTURE yielded at least two main groups of rice i.e. *indica* and *japonica*. Our result showed that, out of three, two populations are either *indica* or *japonica* which might be the primary reason for population structure. This *indica* or *japonica* species differentiation is the outcome of adaptive phenomena for rice under changing climatic condition. A lot of attempts have been made, so far, by the breeding community of rice to cross *indica* and *japonica*, but they faced a strong reproductive barrier which could be second reason of this population structure. Our results showed that there are some genotypes which have the blood of both *indica* and *japonica* at varying levels. These genotypes could be used as a valuable source for crossing between *indica* and *japonica* to overcome reproductive barriers for hybrid seed development and crop improvement. The third population exclusively consisted of wild rice with the highest level of allele conservation demonstrate their diverse origin.

Conclusions

Present analysis of genetic diversity depicted that high preference of certain agronomic traits over the last decades leads to conservation of certain loci. The diverse loci were

loosed in Pakistani cultivars as compared to the other countries cultivars understudy. Overall the Indian genotypes are more diverse followed by Philippines and Pakistan. However, in case of Pakistan, all the rice genotypes and exiting cultivars and most of the genotypes from IRRI are sister lines provided for breeding purposes. The genetic diversity explored for studied genotypes can be used to breed rice cultivars that can be capable of higher production along with the ability to withstand biotic and abiotic stress in new challenging environmental conditions.

Acknowledgments

I pay my special thanks to Dr. Micheal J. Thomson and his team for their guidance and help during research activities at IRRI. Yoo Jin Lee was highly acknowledged for the provision of wild rice.

References

Anupam, A., J. Imam, S.M. Quatadah, A. Siddaiah, S.P. Das, M. Variar and N.P. Mandal, 2017. Genetic diversity analysis of rice germplasm in Tripura State of Northeast India using drought and blast linked markers. *Rice Sci.*, 24: 10–20

Behera, L., B. Patra, R. Sahu, A. Nanda, S. Sahu, A. Patnaik, G. Rao and O. Singh, 2012. Assessment of genetic diversity in medicinal rices using microsatellite markers. *Aust. J. Crop. Sci*., 6: 1369–1376

Cavalli-Sforza, L.L. and A.W. Edwards, 1967. Phylogenetic analysis: models and estimation procedures. *Evolution*, 21: 550–570

Chakhonkaen, S., K. Pitnjam, W. Saisuk, K. Ukoskit and A. Muangprom, 2012. Genetic structure of Thai rice and rice accessions obtained from the International Rice Research Institute. *Rice*, 5: 19

Chuang, H.Y., L. Huu-Shen, H. Kae-Kang and M.C. Chang, 2011. Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers. *Bot. Stud*., 52: 393–405

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, doi: 10.1186/1471–2156–1114–1171

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

Earl, D.A., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.*, 4: 359–361

Ekka, R.E., A. Sarawgi and R.R. Kanwar, 2016. Microsatellite marker based genetic diversity analysis in brown plant hopper resistant rice genotypes. *Bang. J. Bot.*, 45: 1203–1210

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

FAO, 2018. FAO Cereal Supply and Demand Brief. http://www.fao.org/worldfoodsituation/csdb/en/. 24 February 2018

Galeng-Lawilao, J., A. Kumar and D. De Waele, 2018. QTL mapping for resistance to and tolerance for the rice root-knot nematode, Meloidogyne graminicola. *BMC Genet.*, 19: 53

Garris, A.J., T.H. Tai, J. Coburn, S. Kresovich and S. McCouch, 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics*, 169: 1631–1638

Huang, M., F.M. Xie, L.Y. Chen, X.Q. Zhao, L. Jojee and D. Madonna, 2010. Comparative analysis of genetic diversity and structure in rice using ILP and SSR markers. *Rice Sci.*, 17: 257–268

Jasim Aljumaili, S., M.Y. Rafii, M. Latif, S.Z. Sakimin, I.W. Arolu and G. Miah, 2018. Genetic Diversity of Aromatic Rice Germplasm Revealed By SSR Markers. *Biomed. Res. Intl.*, 2018: 7658032

Joshi, R.K., E. Subudhi, B. Kar and S. Nayak, 2010. Comparative genetic analysis of lowland rice cultivars of India using microsatellite markers. *Biores. Bull.*, 4: 213–223

Lee, H., P. Neoh, W. Bong, J. Puvaneswaran, S. Wong, P. Yiu and A. Rajan, 2011. Genotyping of Sarawak rice cultivars using microsatellite markers. *Pertanika J. Trop. Agric. Sci.*, 34: 123–136

Lin, H.Y., Y.P. Wu, A.L. Hour, S.W. Ho, F.J. Wei, Y.I.C. Hsing and Y.R. Lin, 2012. Genetic diversity of rice germplasm used in Taiwan breeding programs. *Botanical Stud.*, 53

Liu, E., S. Zeng, X. Chen, X. Dang, L. Liang, H. Wang, Z. Dong, Y. Liu and D. Hong, 2017. Identification of putative markers linked to grain plumpness in rice (*Oryza sativa* L.) via association mapping. *BMC Genet.*, 18: 89

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

Liu, W., M.Q. Shahid, L. Bai, Z. Lu, Y. Chen, L. Jiang, M. Diao, X. Liu and Y. Lu, 2015. Evaluation of genetic diversity and development of a core collection of wild rice (*Oryza rufipogon* Griff.) populations in China. *PLoS One*, 10: e0145990

Luo, Y., T. Ma, A. Zhang, K.H. Ong, Z. Li, J. Yang and Z. Yin, 2016. Marker-assisted breeding of the rice restorer line Wanhui 6725 for disease resistance, submergence tolerance and aromatic fragrance. *Rice*, 9: 66

Orjuela, J., A. Garavito, M. Bouniol, J.D. Arbelaez, L. Moreno, J. Kimball, G. Wilson, J.F. Rami, J. Tohme and S.R. McCouch, 2010. A universal core genetic map for rice. *Theor. Appl. Genet.*, 120: 563–572

Pervaiz, Z.H., M.A. Rabbani, I. Khaliq, S.R. Pearce and S.A. Malik, 2010. Genetic diversity associated with agronomic traits using microsatellite markers in Pakistani rice landraces. *Electron. J. Biotechnol.*, 13: 4–5

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. Shinozaki, 2010. Genetic analysis of basmati and non-basmati Pakistani rice (*Oryza sativa* L.) cultivars using microsatellite markers. *Pak. J. Bot.*, 42: 2551–2564

Rohlf, F.J., 2000. NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System Version 2.1. Exeter Publishing Setauket, New York

Roy, S., B. Marndi, B. Mawkhlieng, A. Banerjee, R. Yadav, A. Misra and K. Bansal, 2016. Genetic diversity and structure in hill rice (*Oryza sativa* L.) landraces from the North-Eastern Himalayas of India. *BMC Genet.*, 17: 107

Saif, U., 2017. Climate change impact on agriculture of Pakistan- A leading agent to food security. *Intl. J. Environ. Sci. Nat. Res.*, 6: 555690

Salgotra, R., B. Gupta, J.A. Bhat and S. Sharma, 2015. Genetic diversity and population structure of Basmati rice (Oryza sativa L.) germplasm collected from North Western Himalayas using trait linked SSR markers. *PloS One*, 10: e0131858

Shabir, G., K. Aslam, A.R. Khan, M. Shahid, H. Manzoor, S. Noreen, M.A. Khan, M. Baber, M. Sabar, S.M. Shah and M. Arif, 2017. Rice molecular markers and genetic mapping: Current status and prospects. *J. Integr. Agric.*, 16: 1879–1891

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: 423–431

Shah, S.M., K. Aslam, G. Shabir, A.R. Khan, B.H. Abbasi, Z.K. Shinwari and M. Arif, 2015. Population structure and diversity of the AA genome of rice based on simple sequence repeats variation in organelle genome. *Pak. J. Bot*., 47: 1773–1782

Shu, A.P., J. Hwan Kim, S.Y. Zhang, G.L. Cao, Z.H. Nan, L. Qin and L.Z. Han, 2009. Analysis on genetic similarity of japonica rice variety from different origins of geography in the world. *Agr. Sci. Chin.*, 8: 513–520

Singh, N., D.R. Choudhury, A.K. Singh, S. Kumar, K. Srinivasan, R. Tyagi, N. Singh and R. Singh, 2013. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One*, 8: e84136

Singh, N., D.R. Choudhury, G. Tiwari, A.K. Singh, S. Kumar, K. Srinivasan, R. Tyagi, A. Sharma, N. Singh and R. Singh, 2016. Genetic diversity trend in Indian rice varieties: an analysis using SSR markers. *BMC Genet.*, 17: 127

Thomson, M.J., E.M. Septiningsih, F. Suwardjo, T.J. Santoso, T.S. Silitonga and S.R. McCouch, 2007. Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. *Theor. Appl. Genet.*, 114: 559–568

Upadhyay, P., C. Neeraja, C. Kole and V.K. Singh, 2012. Population structure and genetic diversity in popular rice varieties of India as evidenced from SSR analysis. *Biochem. Genet*., 50: 770–783

Upadhyay, P., V.K. Singh and C. Neeraja, 2011. Identification of genotype specific alleles and molecular diversity assessment of popular rice (*Oryza sativa* L.) varieties of India. *Intl. J. Plant Breed. Genet*., 5: 130–140

Wang, C., X. Zheng, Q. Xu, X. Yuan, L. Huang, H. Zhou, X. Wei and S. Ge, 2014. Genetic diversity and classification of Oryza sativa with emphasis on Chinese rice germplasm. *Heredity*, 112: 489

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 Genom.*, 31: 283–292

(Received 03 November 2018; Accepted 1 January 2019)