Australian Journal of Crop Science AJCS

AJCS 10(2):144-151 (2016) ISSN:1835-2707

Genetic diversity and population structure in Indian wild rice accessions

Bikram Pratap Singh[1, 2], Balwant Singh[1], Shefali Mishra[1], Vinod Kumar[1] and Nagendra Kumar Singh*[1]

[1]National Research Center on Plant Biotechnology, New Delhi-110012, India
[2]Banasthali University, Rajasthan, India

***Corresponding author: nksingh@nrcpb.org**

Abstract

Assessment of genetic diversity is essential for germplasm characterization, utilization and conservation. Genetic diversity in a set of 132 Indian wild rice accessions belonging to *Oryza nivara* and *Oryza rufipogon* and eight cultivated rice varieties was evaluated with 25 highly variable simple sequence repeat (HvSSR) markers using agarose gels and 36 genome-wide SNP markers using Sequenom MassARRAY system. A total of 106 SSR alleles were amplified with an average of 4.24 allele per locus, but only 72 alleles with SNP markers because of its bi-allelic nature. PIC values for HvSSR markers ranged from 0.27 (HvSSR 11-24) to 0.71(HvSSR 05-39) with an average of 0.52 and for SNP it ranged from 0.01(11-3935) to 0.39 (09-209) with an average of 0.19, suggesting more allelic diversity for SSRs over SNP markers. Principal coordinate analysis with SSR and SNP markers revealed that the accessions were uniformly distributed across the two axes in both the plots with 66.53% and 77.65% of cumulative variation, respectively. The analysis of the population structure with SNP markers revealed three sub populations. The comprehensive molecular characterization of the selected set of rice germplasm contributes to the knowledge about levels and distribution of genetic diversity in the Indian wild rice.

Keywords: Genetic diversity, HvSSR, Population structure, SNP, Wild rice.
Abbreviations: HvSSR_Highly Variable Simple Sequence Repeat, PCoA_Principle Coordinate Analysis, PIC_Polymorphism Information content, SNP_Single Nucleotide Polymorphism.

Introduction

Rice is the world's most important food crop, serving as staple food for more than half of the world's population (Khush, 2005). With an increasing global population size, rice demand will continue to rise, which increases challenges for the production of high-yielding rice varieties. The continued development and utilization of the genetically diverse rice germplasm resources to constantly expand and enrich the genetic base of the breeding parents is extremely important for the development and promotion of rice breeding and production (Lu, 1998). Cultivated plants and their wild relatives often form a wild cultivated complex and constitute important genetic resources for plant breeding (Muller et al., 2006; Sang and Ge, 2007). Wild rice serves as a valuable reservoir of genes such as resistance to blast (Ram et al., 2007), resistant to brown plant hopper (Rongbai et al., 2001) and resistance to grassy stunt virus (Khush et al., 1977) for genetic improvement of cultivated rice. Likewise, Improved Sambha Mahsuri and Improved Pusa Basmati 1 are the rice varieties released in India through MAS carrying *Xa21* gene for bacterial blight resistance from *O. longistaminata* (Brar and Singh, 2011). Most of the wild rice germplasm, which represent the primary gene pool of rice crop, are seriously endangered by introgression (gene flow from cultivated rice) and human activity that lead to the extinction of diverse germplasm (Gao et al., 2000; Gao et al., 2002; Gao, 2004).

Genetic diversity refers to the total number of characteristics in the genetic makeup of a species. Successful breeding for crop improvement programmes depends on genetic variability that arises from genetic diversity (Rana and Bhat, 2004). Lack of genetic variability may limit breeding progress and gain from selection (Cornelious and Sneller 2002). So, knowledge of the genetic diversity of any germplasm collection provides a basis for improvement of crops and development of superior cultivars. Detailed understanding of the population structure and diversity is also needed for the conservation planning, management and utilization of rice germplasm (Hamrick and Godt, 1996; Frankham et al., 2002).

Simple sequence repeats (SSRs) are segments of DNA sequence consisting of nucleotide repeat units widely distributed across all eukaryotic genomes (Powell et al., 1996). SSRs are the most widely used DNA markers for the characterization of germplasm of different crops (Van Inghelandt et al., 2010). The most important features that make SSRs suitable for assessing the genetic diversity include multiallelism, co-dominance (Vignal et al., 2002), locus specificity, and high reproducibility (Jones et al., 2007; Smith et al., 1996). SSR markers have been applied to detect genetic variations in wild rice (Kuroda et al., 2007; Gao, 2004; Song et al., 2003), cultivated rice (Garris et al., 2005; Thomson et al., 2007, 2009) and weedy rice *O. sativa spontanea* (Yu et al., 2005; Cao et al., 2006; Prathepha, 2011). HvSSR markers used in this study have higher polymorphism rate than other markers (Singh et al., 2010).

Single nucleotide polymorphisms (SNPs) are considered as a robust class of molecular markers that have great importance in plant genetics and breeding because of their excellent distribution throughout the genome and suitability for genetic diversity analysis, evolutionary relationships and

genetic population substructure estimation (Rafalski, 2002; Garris et al., 2003; Varshney et. al., 2008). The present study aimed to investigate the allelic diversity prevalent in a collection of 132 wild rice accessions belonging to *O. nivara* and *O. rufipogon* species from five states of India and eight cultivated rice varieties, using SSR and SNP markers.

Results

Genetic diversity in the Indian wild rice

A total of 25 SSRs were amplified with contrasting alleles across all the accessions in the current set of rice germplasm (Fig.1). All the HvSSR markers were found polymorphic and a total of 106 alleles were amplified ranging from 3 to 6 with a mean of 4.24 alleles per locus. Details of these SSR and SNP markers are given in Supplementary Table 3 and 4. Due to bi-allelic nature of SNP, 72 alleles were amplified with 36 markers. The major allele frequency (MAF) at the SSR loci varied between 0.34 (HvSSR12-31) to 0.81 (HvSSR11-24) with a mean value 0.54 (Table 1). For SNPs, the average major allele frequency was 0.85 which ranged between 0.51 (09-209) to 0.99 (11-3935) (Table 2). PIC values, which represent allelic diversity and frequency, had an average value of 0.52 and 0.19 per marker for SSR and SNP, respectively across all the accessions. The range of PIC was from 0.27 (HvSSR 11-24) to 0.71 (HvSSR 05-39). In case of SNPs, range of PIC was from 0.01 (11-3935) to 0.38 (09-209). Out of 25 SSRs, 8 didn't show any heterozygosity and in the remaining it ranged from 0.01 (HvSSR 01-86) to 0.13 (HvSSR 06-01) with an average of 0.02. While for SNP markers, four markers didn't show any heterozygosity while in other it ranged from 0.01 to 0.45 with an average of 0.06 (Table 2). The gene diversity for SSRs ranged from 0.31 (HvSSR11-24) to 0.75 (HvSSR05-39) with an average of 0.58. For SNP markers, gene diversity was ranged from 0.01 (11-3935) to 0.50 (09-209) with an average of 0.22.

The dendrogram shown in fig. 2 was made from genetic dissimilarity matrix derived from 25 SSR markers using CS chord genetic distance method. In the SSR-based UPGMA tree, rice genotypes were grouped into 14 clusters, however, 6 of them contained less than 5 accessions. Other 8 clusters were categorized as major groups containing cultivated rice varieties (*O. sativa*) and wild rice accessions in separate groups. Wild rice accessions were represented into eight major sub-groups of different intra-specific and eco-geographic regions. Cluster 3 was grouped with all the cultivated rice varieties (*O. sativa*) including one land race (IC 296609) resistance to rice tungro virus. Only one landrace (IC296608) was grouped with four wild rice accessions from UP in cluster 5. Cluster 6 was made up of four *O. nivara* accessions, of which 3 were from Orissa while one was from UP. Cluster 8 and 9 were characterized as major groups of *O. nivara*, respectively containing 42 and 7 accessions from UP. All the accessions from West Bengal were restricted into two major clusters 10 and 13. Cluster 12 was housed with 23 accessions of which 18 belonged to UP, 2 to West Bengal and 3 to Orissa. Most of the *O. nivara* accessions from Orissa were clustered along with West Bengal accessions in cluster 13. A total of four *O. rufipogon* accessions were grouped together into cluster 14 of which 3 belonged to Andhra Pradesh and one to Orissa.

SNP-based tree was subdivided into 7 clusters, of which 3 were grouped with more than 5 accessions (Fig.3). Likewise SSRs most of the cultivated rice varieties (*O. sativa*) were clustered into the same group with comparable level of diversity. Cluster 2, 3 and 5 each contained individual accession. Cluster 6, second largest group of SNP-based tree, was made up of 46 accessions most of them were from Orissa and West Bengal. One of the landraces, IC296609, resistant to rice tungro virus was housed at initial of the cluster 6 as a separate entry. Cluster 7 was grouped with highest number of accessions (80) and most of them were from UP, similar to group I of wild rice derived from 50 K SNPs (Singh et al., 2015). However, genetic similarity level among most of the accessions collected from UP in this group was found very high. Three *O. rufipogon* accessions were also grouped in this cluster. SSR-based principal coordinates analysis clustered the accessions closely agreed with the UPGMA-based tree. The first three coordinate axes accounted for 66.53% of the variation observed (Fig.4A). The first axis explained 35.53% of genetic variation followed by 17.80% explained by second axis. Most of the accessions from UP were plotted in first half of the coordinates while other accessions from Andhra Pradesh, Orissa and West Bengal were plotted in second lower half of the coordinates.

In SNP-based PCoA, due to low diversity, most of the variation (48.73%) was accounted only by first axis (Fig.4B). First three axes cumulatively explained 77.65% of genetic variation. As compared to SSRs, SNP-based intra-specific and geographical categorization of accessions in PCoA was not much descriptive. Accessions from West Bengal were uniformly distributed in both SSR and SNP based PCoA analysis.

In addition, analysis of molecular variance (AMOVA) procedure was used to estimate the partitioning of genetic variance among and within the populations (Table 3). According to HvSSR and SNPs, percentages of genetic variation within population were 88% and 81%, respectively (Fig.5A and B). However, the remaining genetic variation, 12% and 19% respectively for SSR and SNPs were observed among populations.

Population structure analysis of the Indian wild rice

SNP markers were used for the estimation of population structure among the current set of germplasm due to their better resolution at the population structure level over SSR markers (Singh et al., 2013). Considering 36 SNP markers, 3 populations were estimated through Structure Harvester. In Structure analysis, accessions were further categorized as pure or admixture, accessions with more than 0.80 score were considered as pure and less than 0.80 as admixture.

In SNP-based structure analysis, population I consisted 50.7% of accessions (71) including 9 accessions of *O. nivara*, 61 NKSWR accessions from UP and 1 accession NKSWR 43 from Bihar (Fig.6). In population I, 60 accessions were pure and 11 were admixed. Population II comprised of total 51 accessions, 49 wild rice accessions and two landraces, out of which 7 accessions were found admixed. In population III, 8 out of 9 *O. rufipogon* accessions were grouped with the six *O. sativa* varieties. Out of 18 accessions in population III, five *O. rufipogon* and three *O. nivara* accessions were found admixed. Inferred ancestry details with SNP markers for determination of population structure of 140 rice genotypes are given in supplementary Table 2.

Discussion

Genetic diversity plays a major role in survival and adaptability of a species because when a species environment changes, slight gene variations are needed to make changes in the organism phenotype that enables it to adapt and survive in unfavorable conditions. Species with good amount of genetic

Table 1. Major allele frequency (MAF), gene diversity, heterozygosity and polymorphism information content (PIC) of the SSR markers.

S. No.	Marker	MAF	Gene Diversity	Heterozygosity	PIC
1	HvSSR01-02	0.5071	0.6431	0.0000	0.5882
2	HvSSR 01-04	0.4357	0.6764	0.0143	0.6190
3	HvSSR01-24	0.4679	0.7040	0.0286	0.6662
4	HvSSR01-86	0.6607	0.4802	0.0071	0.4073
5	HvSSR02-01	0.4607	0.6792	0.0214	0.6271
6	HvSSR02-12	0.4857	0.6510	0.0000	0.5933
7	HvSSR02-14	0.3750	0.7338	0.0071	0.6895
8	HvSSR02-56	0.4786	0.6255	0.0000	0.5554
9	HvSSR03-23	0.7643	0.3721	0.0000	0.3210
10	HvSSR 03-41	0.4286	0.7279	0.0214	0.6901
11	HvSSR03-74	0.7036	0.4733	0.0071	0.4414
12	HvSSR04-17	0.5214	0.6451	0.0500	0.5970
13	HvSSR05-08	0.7214	0.4407	0.0000	0.3995
14	HvSSR05-39	0.3500	0.7545	0.0071	0.7167
15	HvSSR05-66	0.5500	0.5985	0.0000	0.5357
16	HvSSR06-01	0.7679	0.3727	0.1357	0.3282
17	HvSSR09-07	0.3607	0.7172	0.0500	0.6645
18	HvSSR09-19	0.5750	0.5060	0.0214	0.3979
19	HvSSR09-40	0.5750	0.5299	0.0571	0.4373
20	HvSSR10-35	0.6250	0.5183	0.0214	0.4460
21	HvSSR11-24	0.8107	0.3154	0.0071	0.2795
22	HvSSR12-10	0.4786	0.5634	0.0286	0.4662
23	HvSSR12-31	0.3429	0.7268	0.0000	0.6776
24	HvSSR12-35	0.6214	0.5568	0.0071	0.5112
25	HvSSR12-37	0.6643	0.5080	0.0000	0.4619
	Mean	0.5493	0.5808	0.0197	0.5247

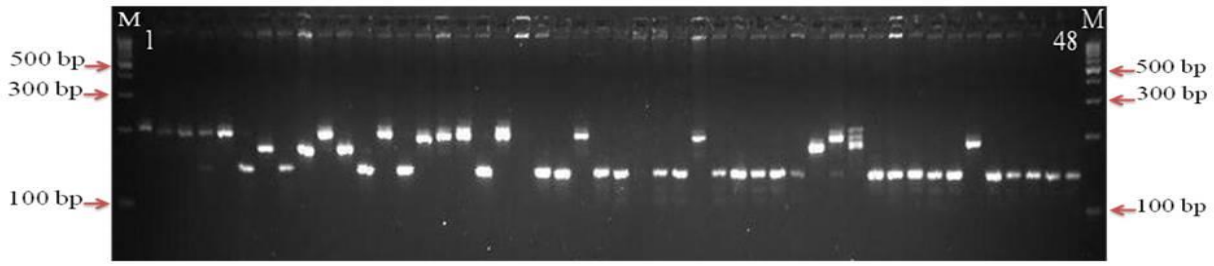

Fig 1. Gel picture of marker HvSSR 03-41 for wild rice accessions (1- 48) resolved in 4% metaphor agarose gel. M=100bp DNA marker.

diversity among populations have more variations to choose the fittest alleles. Species with little genetic diversity are at a greater risk of extinction. Broadening the genetic base of core breeding material requires the identification of diverse parents for hybridization with cultivated rice.

This study investigated the allelic diversity existing among a collection of 132 wild rice accessions and 8 cultivated rice varieties using 25 HvSSR markers and 36 SNP markers. Both of the DNA-based co-dominant markers offer great variability in crop germplasm which makes them more suitable in molecular breeding. Comparatively, SSRs were more polymorphic over SNP markers even when used in less numbers. Both of the markers have successfully been applied in rice to decipher genetic diversity (Das et al., 2013; Singh et al., 2013; Emanuelli et al., 2013). Our study also implied that SSR-based clustering was more informative as compared to SNP-based clustering. Despite higher number of SNP markers over SSRs in current study, they were unable to classify some of the wild rice accessions whereas lesser SSRs categorized them successfully with greater variability. PIC values for SSRs were found better over SNP markers which indicate that multi-allelic SSRs show high allelic diversity over bi-allelic SNP markers in rice. In case of SSRs, on an average 50% SSRs showed PIC value more than 0.5 which entails the utility of SSRs in rice diversity analysis.

All the 25 SSR markers used for this study revealed a clear and consistent amplification profile. However, both of the SSR and SNP markers did not detect private alleles for *O. rufipogon* and *O. nivara* species separately so they can be identified as distinct groups. Better classification of these two species might be observed if we have included some of the markers which can specifically classify them in separate groups. Also the lower resolution might be due to the loci involved in this study revealed only a small part of the genotypic constitution. The trees generated by SSR and SNP markers were in agreement with the structure results. In case of SSRs, all the accessions from cluster 7, 8 and 9 were allocated in population I of structure. Cluster 9 based on SSR markers was similar to cluster 7 based on SNP and allotted to population I. Cluster 3 of SSR derived tree containing *O. sativa* varieties was corresponding to cluster 1 of SNP derived tree and grouped together in population structure analysis. Population structure analysis based on SNP markers showed that 7% of accessions were highly admixed. Due to low variability in SNP dataset, most of the genetic variation has been explained by first axis of PCoA analysis. However, both of the markers were found important in grouping of the

Table 2. Major allele frequency, gene diversity, heterozygosity and polymorphism information content of the SNP markers.

S. No.	Marker	MAF	Gene Diversity	Heterozygosity	PIC
1	01-37451-338	0.8571	0.2552	0.1857	0.2403
2	01-6084-375	0.7786	0.3448	0.0429	0.2854
3	01-63511-202	0.6250	0.4688	0.0786	0.3589
4	02-267	0.8679	0.2341	0.0357	0.2149
5	02-30291-474	0.7964	0.3243	0.0786	0.2717
6	02-43331-293	0.8857	0.2024	0.0286	0.1820
7	03-16911-373	0.8893	0.1969	0.0357	0.1775
8	03-34781-206	0.9321	0.1265	0.0643	0.1185
9	03-46601-355	0.5679	0.4908	0.0786	0.3704
10	04-180120-428	0.7571	0.3678	0.3714	0.3001
11	04-194-240	0.8857	0.2089	0.0857	0.1985
12	04-37873-358	0.9036	0.1755	0.0357	0.1624
13	05-26921-109	0.6893	0.4283	0.1071	0.3366
14	05-41921-280	0.9143	0.1600	0.0143	0.1532
15	05-481-279	0.6357	0.4632	0.0571	0.3559
16	06-12561-147	0.8179	0.2979	0.0357	0.2536
17	06-17761-501	0.9750	0.0490	0.0071	0.0483
18	06-25091-497	0.8857	0.2081	0.0571	0.1964
19	07-290439-299	0.9786	0.0419	0.0143	0.0411
20	07-29312-368	0.9929	0.0142	0.0000	0.0141
21	07-43201-192	0.9429	0.1090	0.0000	0.1054
22	08-27652-360	0.8750	0.2212	0.0071	0.2010
23	08-42185-129	0.7643	0.3695	0.0714	0.3153
24	08-8476-113	0.8750	0.2188	0.0357	0.1948
25	09-209	0.5107	0.5067	0.1214	0.3853
26	09-21075-145	0.9821	0.0351	0.0071	0.0345
27	09-27164-457	0.9393	0.1141	0.0071	0.1076
28	10-11927-178	0.9000	0.1841	0.0286	0.1745
29	10-1881	0.9750	0.0488	0.0214	0.0476
30	10-2723	0.7036	0.4324	0.4500	0.3605
31	11-1849	0.9929	0.0142	0.0000	0.0141
32	11-3935	0.9964	0.0071	0.0071	0.0071
33	11-5221-214	0.9571	0.0820	0.0143	0.0787
34	12-1794	0.9786	0.0421	0.0143	0.0417
35	12-32002-389	0.7643	0.3636	0.0000	0.3025
36	12-400	0.9214	0.1479	0.0429	0.1426
	Mean	0.8532	0.2210	0.0623	0.1887

Table 3. AMOVA summary for SSR and SNP markers.

Source	AMOVA for SSR					AMOVA for SNP			
	df	SS	MS	Est. Var.	%	SS	MS	Est. Var.	%
Among Pops	6	470.053	78.342	3.738	12%	108.190	18.032	0.988	19%
Within Pops	133	3526.175	26.513	26.513	88%	576.560	4.335	4.335	81%
Total	139	3996.229	104.855	30.251	100%	684.750	22.367	5.323	100%

current set of accessions to some extent according to their geographical locations. All the three methods (UPGMA-tree, PCoA and Structure) used in the current study to classify the cultivated and wild rice accessions according to their species and geographical locations, were found meaningful as most of the groups were co-linear in all the grouping methods.

NKSWR 48 in population III generated by SNP markers was very high grain yielding wild rice accession and grouping with cultivated rice varieties (*O. sativa*) suggests its possible use in breeding program of the yield related traits in cultivated rice. Considering AMOVA, both the markers showed a lower genetic differentiation among populations, and most of the diversity is due to individuals within the populations. The distribution of variation among and within the population was found proportional with earlier studies (Yamasaki and Ideata, 2013; Singh et al., 2013). These wild rice accessions are the rich source of potential alleles for biotic and abiotic stress and yield related traits, so can be used for rice improvement. The distribution of these wild accessions across India indicates existence of good amount of genetic diversity and need for their conservation. The diverse accessions from the current set of Indian wild rice germplasm analyzed by molecular markers can be utilized to enhance the genetic base of the cultivated rice through breeding.

Materials and Methods

Plant material and DNA extraction

Forty *O. nivara* and six *O. rufipogon* and eight agronomically improved cultivated rice accessions obtained from NBPGR gene bank including two landraces resistance to rice tungro virus along with their passport data and 86 wild rice accessions (NKSWR) collected from their natural habitats UP and Bihar (India), were used in this study. These wild rice accessions were characterized based on pSINE markers (Yamanaka et al., 2003) and found belonging to 83 *O. nivara* and three *O. rufipogon* accessions. The details of these wild rice accessions are given in supplementary Table1 and in our wild rice database http://nksingh.nationalprof.in/.

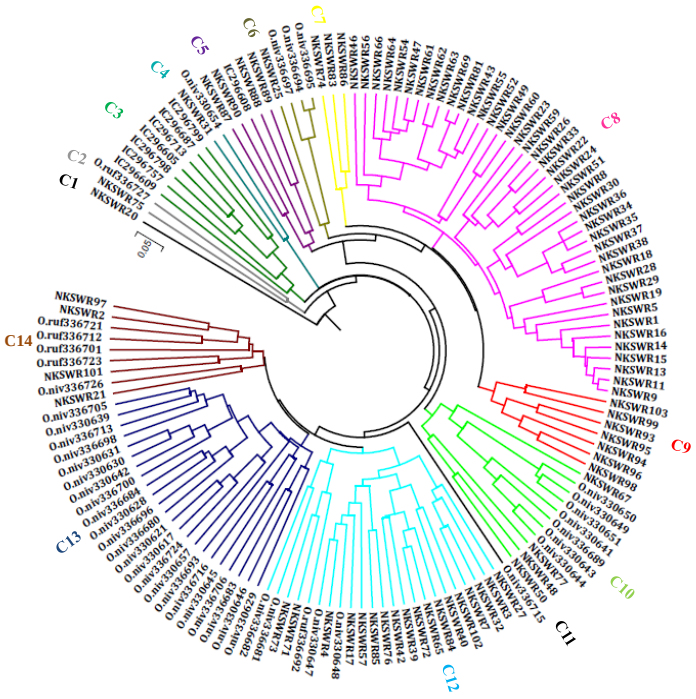

Fig 2. UPGMA tree based on dissimilarity index of 25 genome-wide SSR markers for 140 rice genotypes.

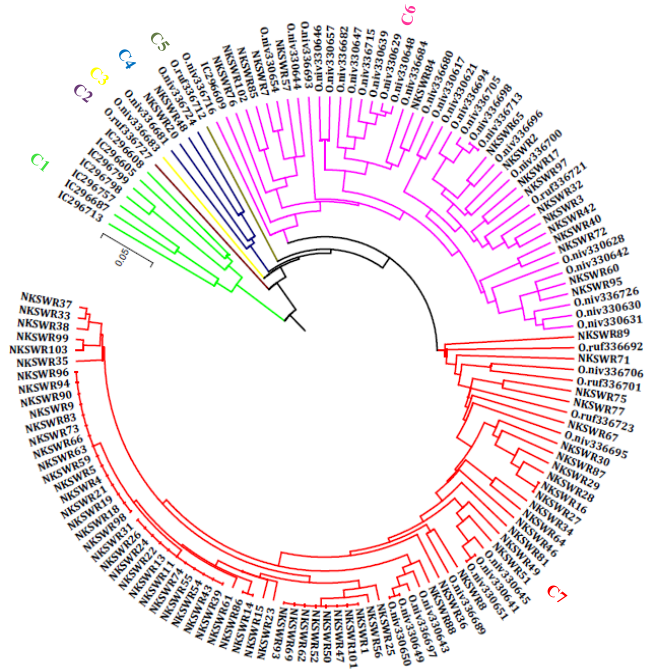

Fig 3. UPGMA tree based on dissimilarity index of 36 genome-wide SNP markers for 140 rice genotypes.

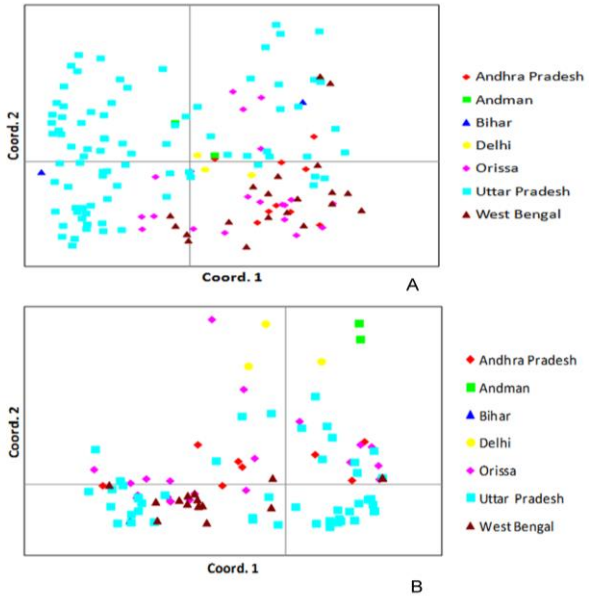

Fig 4. Principal coordinate analysis of 140 rice genotypes based on (A) SSR markers data and (B) SNP markers data.

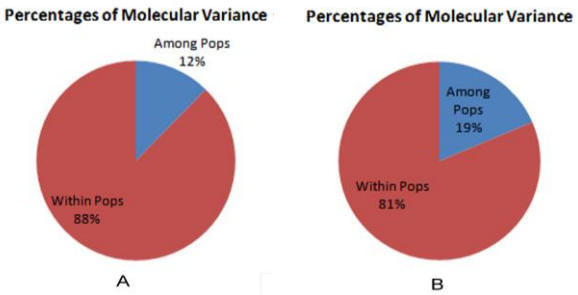

Fig 5. Analysis of Molecular variance of 140 rice genotypes based on (A) SSR markers data and (B) SNP markers.

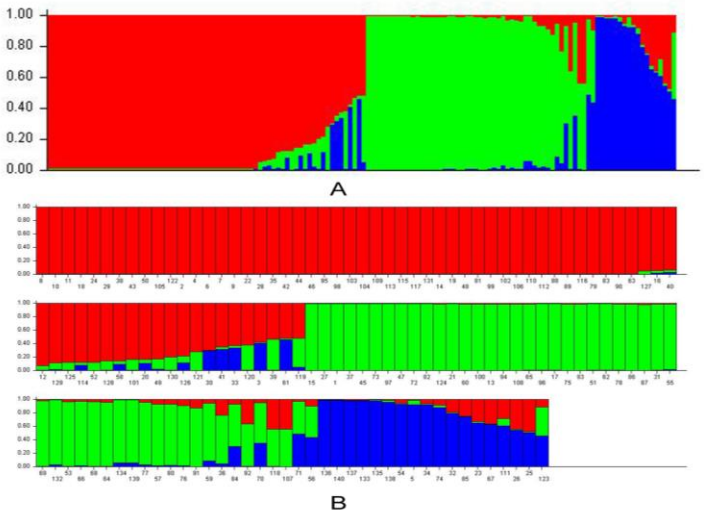

Fig 6. Model based population structure plot for each variety with K=3, using Structure with 36 SNP markers. Color codes are as follows: Population I red, Population II green, population III blue (A). Code of each wild rice accession corresponds to description in table (B).

The exact location of each collection site was documented using a global positioning system (GPS) receiver from GARMIN. Seeds of all the 140 rice accessions were grown during Kharif-2011 in the field of IARI. After 25 days fresh leaves of individual plant of each accession were collected for DNA isolation and seedlings were transplanted in the field for seed multiplication. Genomic DNA extractions of each individual leaf sample from all the accessions were performed using CTAB method as described by Murray and Thompson (1980).

Genotyping by SSR markers

Twenty five HvSSR primer pairs with relatively high polymorphism and distributed across the rice genome were selected for the genetic diversity analyses of this set of germplasm (Singh et al 2010). The PCR reactions were carried out in a volume of 15 µl reaction mixture each containing 1.5 µl of 10x reaction buffer, 0.20 µl of 10 mM dNTPs (133 µM), 1.5 µl each of forward and reverse primers (10 pmol), 2.0 µl of template genomic DNA (50 ng), 0.15 µl of Taq DNA polymerase (0.75 U) (Vivantis Technologies) and 9.65 µl MQ water. The PCR thermal profiles were: DNA denaturation at 94ºC for 4 min. followed by 34 cycles of 94ºC for 1 min., 55ºC for 30 sec.,72ºC for 1 min. and finally, 72ºC for a final extension of 10 min. The PCR products were separated by electrophoresis in 4% Metaphor agarose gels (Lonza, Rockland ME USA) contain 0.1 µg/ml ethidium bromide in 1x TBE buffer at 120 Volts for 3 hours. After electrophoresis DNA fragments were visualized and documented using ALPHA IMAGER gel documentation system (Alpha Innotech, USA).

Genotyping by SNP markers

The Sequenom MassARRAY system was used for the high-throughput SNP genotyping. Sequenom MassARRAY multiplex assays were designed for 36 SNPs, representing conserved single copy rice genes (Singh et al., 2007) taking three genes from each of the rice chromosome. This 36-plex assay was designed and validated by Sequenom Corporation (San Diego). The 30-mer pre-amplification primers and variable length genotyping primers generated by the Assay Design 3.1software were used for the validation of this SNP assay according to the Sequenom user manual (www. sequenom.com).

Statistical analysis

The SSR profiles were scored based on the size of fragments amplified across all accessions of wild rice and cultivated varieties. In case of SNP, for the visualization of SNPs and allele calling Mass ARRAY Typer 3.4 Software was used. The major allele frequency, polymorphism information content (Table 1) and genetic distance based clustering was performed with UPGMA tree using Power Marker v3.25 (Liu and Muse 2004), and the dendrogram was constructed using MEGA 4.0 software (Tamura et al 2007). Software GenAlEx V6.5 (Peakall et al., 2012) was used for the Principle Coordinate Analysis and Analysis of Molecular Variance. The population structure was inferred using Structure 2.3.4 (Pritchard et al., 2000). For Evanno plot, the Structure outputs were visualized using Structure Harvester (Earl et al., 2012). In Structure, we assumed an admixed model with independent allele frequency and a uniform prior probability of the number of populations, *K*. All the runs with 50,000 MCMC replicates after a burn-in of 50,000 replicates were

conducted for $K = 2$ to 10. Five independent runs were done for each value of K to generate our estimate of the true number of sub populations (Pritchard et al., 2000).

Conclusion

Breeding programmes depend on a high level of genetic diversity for achieving progress through wide hybridization. Wild rice is known as rich source of genes for genetic improvement of cultivated rice varieties. Genetic studies of wild rice accessions are necessary to provide information for their utilization by hybridization, proper management and conservation programs. In this study, we analyzed genetic diversity and population structure of Indian wild rice, collected from UP, Bihar, Orissa, West Bengal and Andhra Pradesh. The SSR and SNP based analysis shows that a high level of genetic diversity exists in Indian wild rice collection and population structure analysis reveals the occurrence of three different gene pools of wild rice accessions among the examined populations, one of which is closely related to the Indica group of rice cultivars. This set of wild rice germplasm with sufficient variability provides opportunity of their possible use in allele mining and rice breeding.

Acknowledgements

The financial assistance received from Indian Council for Agricultural Research funded "Network Project on Transgenic Crops" project is gratefully acknowledged. We are thankful to NBPGR gene bank for providing seeds of 40 *O. nivara*, 6 *O. rufipogon* accessions and 8 cultivated rice varieties used in this study.

References

Brar DS, Singh K (2011) Wild Crop Relatives: Genomic and Breeding Resources, Cereals. Springer-Verlag Berlin Heidelberg.

Cao Q, Lu BR, Xia H, Rong J, Sala F, Spada A, Grassi F (2006) Genetic diversity and origin of weedy rice (*Oryza sativa* f. *spontanea*) populations found in North-eastern China revealed by simple sequence repeat (SSR) markers. Ann Bot. 98: 1241- 1252.

Cornelious BK, Sneller CH (2002) Yield and molecular diversity of soybean lines derived from crosses of Northern and Southern elite parents. Crop Sci. 42: 642-647.

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

Earl DA, Von H, Bridgett M (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 4: 359-361.

Emanuelli Francesco, et al.,(2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC plant biol. 13: 39.

Frankham R, Ballou JD, Briscoe DA (2002) Introduction to conservation genetics. Cambridge University Press, Cambridge, United Kingdom.

Gao LZ (2004) Population structure and conservation genetics of wild rice *Oryza rufipogon* (*Poaceae*): a region-wide perspective from microsatellite variation. Mol Ecol. 13:1009-10024.

Gao LZ, Chen W, Jiang WZ, Ge S, Hong DY, Wang XK (2000) Genetic erosion in the Northern marginal population of the common wild rice *Oryza rufipogon* Griff. and its conservation, revealed by the change of population genetic structure. Hereditas. 133: 47–53.

Gao LZ, Ge S, Hong DY, Lin R, Tao G, Xu Z (2002) Allozyme variation and conservation genetics of common wild rice (*Oryza rufipogon* Griff.) in Yunnan, China. Euphytica. 124: 273–281.

Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics. 165:759–769.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics. 169:1631-1638.

Hamrick JL, Godt MJW (1996) Effect of life history traits on genetic diversity in plant species. Phil Trans R Soc Lond B. 351:1291-1298.

Jones ES, Sullivan H, Bhattramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet. 115: 361-371.

Khush GS (2005) What it will take to feed five billion rice consumers by 2030. Plant Mol Biol. 59: 1-6.

Khush GS, Ling KC, Aquino RC, Aguiero VM (1977) Breeding for resistance to grassy stunt in rice. Proceedings of the 3rd International Congress of the Society for the Advancement of Breeding Researchers in Asia and Oceania (SABRAO). 12-13 Feb 1977, Canberra, Australia.

Kuroda Y, Sato YI, Bounphanousay C, Kono Y, Tanaka K (2007) Genetic structure of three *Oryza* AA genome species (*O. rufipogon*, *O. nivara* and *O. sativa*) as assessed by SSR analysis on the Vientiane Plain of Laos. Conserv Genet. 8: 149-158.

Liu K, Muse S (2004) Power Marker: new genetic data analysis software, version 2.7 (http://www.powermarker.net).

Lu BR (1998) Diversity of rice genetic resources and its utilization and conservation. Chin Biodiv. 6(1): 63–72.

Muller MH, Poncet C, Prosperi JM, Santoni S, Ronfort J (2006) Domestication history in the *Medicago sativa* species complex: inferences from nuclear sequence polymorphism. Mol Ecol. 15:1589–1602.

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acid Res. 8: 4321-4325.

Peakall R, Smouse P (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics. 28: 2537-2539.

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends in Plant Sci. 1: 215-222.

Prathepha P (2011) Microsatellite analysis of weedy rice (*Oryza sativa* f. *spontanea*) from Thailand and Lao PDR. Aust J Crop Sci. 5:49-54.

Pritchard JK, Matthew S, Peter D (2000) Inference of population structure using multilocus genotype data. Genetics. 155: 945–959.

Rafalski JA (2002) Application of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol. 5:94–100.

Ram T, Majumder ND, Mishra B, Ansari MM, Padmavathi G (2007) Introgression of broad-spectrum blast resistance gene(s) into cultivated rice (*Oryza sativa* ssp *indica*) from wild rice *O. rufipogon*. Curr Sci. 92:225-228.

Rana MK, Bhat KV (2004) A comparison of AFLP and RAPD markers for genetic diversity and cultivar identification in cotton. J Plant Biochem Biotech. 13: 19-24.

Rongbai L, Xueyi Q, Sumei W, Pandey MP, Pathak PK, Fenguan H, Qing L, Shanyu L (2001) Inheritance of tresistance to brown plant hopper in an *Oryza rufipogon* (Griff.)-derived line in rice. Curr Sci. 80:1421-1423.

Sang T, Ge S (2007) The puzzle of rice domestication. J Integr Plant Biol. 49:760–768.

Singh H, Deshmukh RK, Singh A, Gaikwad K, Sharma TR, Mohapatra T, Singh NK (2010) Highly variable SSR markers suitable for rice genotyping using agarose gels. Mol Breeding. 25:359–364.

Singh NK, Dalal V, Batra K, Singh BK, Chitra G et al.,(2007) Single copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Funct Integr Genomics. 7: 17‑35.

Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Singh NK, Singh R. (2013) Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS One. 8(12): e84136.

Singh N, Jayaswal PK, Panda K, Mandal P, Kumar V, Singh B, Mishra S, Singh Y, Singh R, Rai V, Gupta A, Raj Sharma T, Singh NK (2015) Single-copy gene based 50K SNP chip for genetic studies and molecular breeding in rice. Sci. Rep. 5,11600; doi: 10.1038/srep11600.

Smith S, Helentjaris T (1996) DNA fingerprinting and plant variety protection. In AH Paterson, ed, Genome Mapping in Plants, Landes Company, Texas, pp 95-110.

Song ZP, Xu X, Wang B, Chen JK, Lu BR (2003) Genetic diversity in the northernmost *Oryza rufipogon* Griff. populations estimated by SSR markers. Theor Appl Genet. 107: 1492–1499.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596‑1599.

Thomson MJ, Polato NR, Prasetiyono J, Trijatmiko KR, Silitonga TS, McCouch SR (2009) Genetic diversity of isolated populations of Indonesian landraces of rice (*Oryza sativa* L.), collected in East Kalimantan on the Island of Borneo. Rice. 2: 80-92.

Thomson MJ, Septiningsih EM, Suwardjo F, Santoso TJ, Silitonga TS, McCouch SR (2007) Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using micosatellite markers. Theor Appl Genet. 114: 559-568.

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet. 120: 1289-1299.

Varshney RK, Thiel T, Sretenovic-Rajicic T, Baum M, Valkoun J, Guo P, Grando S, Ceccarelli S, Graner A (2008) Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. Mol Breed. 22:1–13.

Vignal A, Milan D, San CM, Eggen A (2002) A review of SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol. 34: 275-305.

Yamanaka S, Nakamura I, Nakai H, Sato YI (2003) Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial strains of wild rice species, *Oryza nivara* and *O. rufipogon*. Genet Resour Crop Evol. 50:529–538.

Yamasaki M, Ideta O (2013) Population structure in Japanese rice population breeding. Science. 63: 49-57.

Yu GQ, Bao Y, Shi CH, Dong CQ, Ge S (2005) Genetic diversity and population differentiation of Liaoning weedy rice detected by RAPD and SSR markers. Biochem Genet. 43**:** 261–270.