*Indian J. Genet*., 76(3): 246-254 (2016)
DOI: 10.5958/0975-6906.2016.00039.0

# Genetic diversity in Indian rice germplasm set using phenotypic and genotypic variables simultaneously

**Nivedita Singh[#], Shana Majumder, Sanjay Singh[1] and Prashant Vikram*[2]**

Sharda University, Greater Noida, Uttar Pradesh; [1]National Research Centre on Plant Biotechnology, Indian Agericultural research Institute, New Delhi 110 012; [2]International Rice Research Institute, Los Banos, Philippines

(Received: October 2015; Revised: July 2016; Accepted: August 2016)

**Abstract**

**Approximately 11% of the Indian rice varieties notified from 1969 to 2012 have IR-8 in their parentage. More than 30% notified varieties during this period are derived from germplasm with 'IR' in their parentage. India is one of the centers of diversity in rice and broadening of genetic base of cultivated rice can be successfully performed using indigenous germplasms. Genetic diversity among 42 Indian rice varieties (*Oryza sativa* L.), including landraces, *aus* cultivars and modern varieties of India were investigated for phenotype as well as genotype diversity. Genetic diversity analysis using genome-wide SNP panel suggested that *aus* cultivars, landraces and indica rice germplasm are genetically very close. The SNP panel was based on single copy genes thereby reflecting functional diversity of the germplasm set. The dendrogram based on molecular marker analysis grouped the 42 rice cultivars into 4 diverse groups. Further, population structure analysis also resulted to 4 sub populations. These sub population differentiated *aus* germplasm set from other accessions. Further, principal component analysis based on phenotype also suggested the closeness of *aus* cultivars, landraces and indica rice germplasm sets. Information generated through genetic, phenotypic and comprehensive diversity profiles in the present study can be efficiently utilized in breeding high yielding rice varieties.**

**Key words**: Rice, landraces, structure, molecular tags, genetic diversity

**Abbreviations**: AMOVA: Analysis of Molecular Variance; PIC: Polymorphism Information Content; PCoA: Principle Co-ordinate Analysis

## Introduction

Rice (*Oryza sativa* L.) is one of the earliest domesticated crop species and has become the one of the world's most widely grown crops. Rice consumption constitutes about 20% of the world's calorie intake, and in Asian countries, where over half of the world's population lives, rice represents over 50% of the calories consumed (Agrama et al. 2010).

In addition to the two major subspecies, japonica and indica, (Agrama et al. 2010) several other minor rice types have been identified with genetic markers (Londo et al. 2006; Bautista et al. 2001) which include the short duration *aus* germplasm of India and Bangladesh, the deep-water rice of Bangladesh, and the aromatic rice. Extensive genotypic and phenotypic diversity exists within *O. sativa*, due to adaptations to different habitats, resulting in about 120,000 different accessions (Second et al. 1982). These accessions include traditional rice landraces and commercially bred cultivars. Landraces of rice have been selected over generations by farmers are the one of the most important repositories of genetic diversity. Their value in modern plant breeding is massive. Non-elite natural populations such as landraces harbour rare alleles besides other high- and low-frequency alleles. These germplasm sets can play a significant role in expanding the genetic base of cultivated rice varieties. The crosses between parents with maximum genetic divergence are generally the most responsive for genetic improvement (Harlan et al. 1975). Genetic diversity is the most important pre-requisite for a rice improvement program as it helps in broadening the genetic base, identifying recombinants and therefore, giving chance to breeders in selecting favorable alleles.

*Corresponding author's e-mail: p.vikram@cgiar.org
Present address: [#]National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110 012
Published by the Indian Society of Genetics & Plant Breeding, F2, First Floor, NASC Complex, PB#11312, IARI, New Delhi 110 012
Online management by indianjournals.com

Molecular markers, especially DNA-based markers, have been used extensively for the study of genetic diversity. Mackill et al. (1995) using methodology as proposed by Arunachalam et al. (1981) classified traditional and modern japonica rice cultivars using RAPD and successfully distinguished the temperate and tropical japonica genotypes. SSR markers have been used for evaluating rice genetic diversity by a number of groups (Wu et al. 1993; Panaud et al. 1996; Xiao et al. 1996; Olufowote et al. 1997; Thanh et al. 1999; Yang et al. 1994, Akagi et al. 1997; Herrera et al. 2008) for estimating variation between the indica and japonica subspecies and classifying *Oryza sativa* L. genotypes. At national level breeding programs diversity analyses have been performed over a period of time (Rathi et al. 2014; Krishnamurthy et al. 2014). Herrera et al. (2008) used 48 simple-sequence-repeat (SSR) markers to assess the genetic diversity of 11 Venezuelan rice cultivars released by the National Rice Breeding Program between 1978 and 2007 and detected 203 alleles. Zhao et al. (2011) recently classified a global set of rice germplasm in to five groups using high density SNP array.

Genus *Oryza* is highly rich in intra specific as well as inter specific diversity allowing its diverse cultivation pattern. Rate of increase in yield potential of rice globally is less than required (Ray et al. 2013) to feed burgeoning food demands and challenges related to climate change are on card. Narrow genetic base in the breeding materials confines the genetic gain during improvement of genotype. Knowledge of genetic diversity and relatedness is an important factor for adopting a systematic and effective breeding strategy for improving genotype. Broadening cultivated rice gene pool through bridging utilized and under-utilized diversity can help greatly in addressing such problems. Balance between genetic drift, inbreeding, recombination, gene flow, mutation and selection determines the genetic diversity present within and among populations (Loveless et al. 1984; Hartl et al. 1997) and is affected by both natural and anthropogenic factors. Identification of diverse sources of genes for resistant to stress tolerance and yield potential has become important now. One way to identify alleles/genes for various qualitative and quantitative traits has been through use of various types of molecular markers (Pachauri et al. 2013). Therefore, focusing a comprehensive approach involving molecular markers applications along with large scale phenotyping platforms is required to expand the genetic base of breeding germplasm pool of rice.

Most of the marker assay/systems used till date are based on non-genic region because of the presence of higher polymorphisms in those regions. Study based on single copy genes should provide a good estimate of functional diversity. The polymorphism present among the candidate genes may be more enlightening than anonymous DNA markers with respective to functional diversity. Single copy genes are important source of variation because they represent the functional aspects of plant metabolism. Therefore, single nucleotide polymorphism (SNP) markers derived from functional region of genomes should provide a robust estimate of genetic diversity relevant to plant breeders. In this study SNP information based on single copy genes have been implied for diversity analysis along with phenotype information. The present study was conducted on 42 Indian rice varieties in order to study the extent of genetic variation and relatedness between genotypes.

## Materials and methods

### *Phenotypic evaluation of rice varieties for yield related traits*

The forty two different rice varieties including traditional cultivars, cultivated varieties and landraces were evaluated in randomized block design with two replications. Twenty one day old seedlings were transplanted in 2.0-meter 3-row plots and middle row was harvested for measuring the 1000 grain weight. Standard agronomic practices were followed for field trial management. Data pertaining to grain yield related traits were recorded, *viz*., days to flowering, plant height, panicle length, ear bearing tillers, grain yield and the thousand grain weight.1000-grain weight was measured in grams by randomly weighing 1000 filled grains. Observations were recorded on five plants per genotype and averaged over five samples.

### *Genotypic analysis*

Six to nine fully mature healthy rice grains were dehulled/dehusked in a mechanical Satake dehuller with the resulting brown rice ground into fine powder in a tissue lyser (Tissue lyser II Retsch, Germany). The fine powder was then used to isolate DNA using QIAGEN DNeasy plant mini kit (Hilden, Germany). Manufacturers' instructions were followed accordingly.

A total of 36 SNPs from conserved single copy rice genes with an average of three genes per rice

chromosome (one each from short arm, long arm and centromeric region from all 12 rice chromosomes) were designed and used for genotyping using Sequenom MALDI-TOF MassARRAY system (Vikram et al. 2016) (http://www.sequenom.com). The multiplex assays and 30-mer pre-amplification primers were designed using MassARRAY Assay Design 3.1 software and optimized premixed primers (2 x 36 plex assays) were procured from Sequenom (San Diego, California). PCR reactions were done as per Sequenom iPLEX Gold amplification kit instruction manual. Mass ARRAY workstation ver.3.3 was used to analyze the results of iPLEX spectrochip bio-arrays.

### *Statistical analysis of massARRAY data*

MassARRAY Typer 3.4 Software was used for visualization of SNPs and allele calling. Power Marker 3.5 software was used to obtain major allele frequency, gene diversity, heterozygosity and PIC values for each locus for SNP as well as SSR markers (Liu et al. 2005) and genetic distances across the genotypes. The dissimilarity matrix generated by Power Marker was used to construct un-weighted neighbor joining (NJ) tree using DARwin 5.0.158 software. GenA1Ex V6.5 software (Peakall et al. 2012) was used to calculate Principle Coordinate Analysis (PCoA) and Analysis of Molecular Variance (AMOVA). For SNP data, the four nucleotide bases were numerically coded *viz*., A=1, C=2, G=3 and T=4 while any missing data was coded as 0.

### *Genetic diversity analysis*

STRUCTURE V2.3.1 software was used to cluster similar rice genotypes on the basis of their historical lineages. To obtain genetic clusters, K value was set from 1 to 20 using admixture model with correlated allele frequency; with each set replicated three times. Each run was implemented with a burn-in period of 100,000 steps followed by 100,000 Monte Carlo Markov Chain replicates (Pritchard et al. 2000). Ln(PD) derived for each K was then plotted to obtain plateau of the ÄK values (Evanno et al. 2005). Final population structure was constructed using open web source program STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu). The proportion of the genome of individual rice genotypes belonging to the admixture of each inferred population was also estimated (Pritchard et al. 2000; Evanno et al. 2005). A principal component analysis was also done based on the grain yield related traits *viz*., days to flowering, plant height, panicle length, ear bearing tillers, grain yield and the thousand grain weight) using R-project (*www.r-project.org*).

## Results and discussion

Enhancement of the genetic yield potential of rice is a major concern for global food security. Genetic improvement of Indian rice germplasm has been centered mainly to the green revolution/pre-green revolution rice varieties IR8, TKM6 or Mahsuri (Upadhyay et al. 2011). Most of the high yielding varieties released from 1970s to till date have IR8, TKM6 and Mahsuri either in their parentage or in their pedigree. We compiled the total number of rice varieties notified from 1969 to 2012 (http://drdpat.bih.nic.in/Rice%20Varieties%20in%20India.htm)

and found that 11% of them had IR-8 as direct parentage. A total of 11% notified varieties had parentage with TN-1, TKM-6, Mahsuri or IR36. Interestingly, 34% notified varieties have been derived from IRRI derived lines i.e. with 'IR' pedigree. Recent reports have clearly delineated the negative selection of rice breeders against drought tolerance during and after green revolution (Vikram et al. 2016; Vikram et al. 2016; Vikram et al. 2016). Continuous use of the similar parents in the national breeding programs is attributed to enhanced homogeneity in the Indian rice germplasm. Therefore stagnation in the genetic yield potential is another probable threat for further yield enhancement that too under rain-fed environments. Development of new rice varieties with comparable yield potential under irrigated and enhanced yields under drought stress is the ultimate need of plant breeders. India is the home for a number of *aus* germplasm as well as landraces (Mc Nally et al. 2009). Landraces in India and other Southeast Asian countries have been maintained by farmers. Since these landraces didn't undergo selection, they endow tremendous genetic variability. The *aus* group of rice cultivars and locally adapted traditional varieties/land races has immense genetic potential for the yield related parameters which can be potentially utilized for genetic improvement of rice yields. This genetic variability could be utilized efficiently in genomics studies as well as molecular breeding for improving yield and yield related parameters in rice genotypes. A total of 42 rice varieties were subjected to genotypic as well as phenotypic diversity analysis in the present study. Genotypic and phenotypic analyses were carried out with SNP markers and grain yield related parameters respectively. Significant variation was observed among traits under study. Among these traits plant height ranged 73-136 (cm); days to flowering

ranged 66-103 (days), panicle length: 18-31 (cm), ear bearing tiller: 5-14, and grain yield: 1.1-5.6 t/ha$^{-1}$. Similarly, significant variation in 1000-grain weight in the rice varieties was reported (Pd"0.05). Basmati 370 had the lowest thousand-kernel weight (17.1g) whereas Jaya showed the highest value for thousand-kernel weight (30.57 g). Details of the phenotypic variability among the 42 rice genotypes are summarized in Tables 1a, 1b and 2. Coefficient of variation was maximum for grain yield and minimum for plant height in the germplasm set analyzed in this study. This analysis suggests utilizing larger collections for future diversity analyses in rice. These data sets were used to perform principal component analysis presented in

**Table 1a.** General morphological and grain features of the 42 rice genotypes recorded for two consecutive crop seasons

| S. No. | Genotype | Variety type | EVV _1 | EVV _2 | DTF _1 | DTF _2 | PH _1* | PH _2* | PL _1* | PL _2* | EBT _1 | EBT _2 | GY _1# | GY _2# | 1000-GW@ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Anjali | Indica | 3.7 | 3.0 | 84.0 | 86.7 | 95.7 | 99.5 | 25.5 | 25.7 | 9.0 | 8.3 | 4.1 | 4.4 | 27.94 |
| 2 | Annada | Indica | 5.7 | 5.0 | 82.7 | 80.0 | 84.7 | 83.5 | 25.5 | 25.5 | 10.0 | 9.7 | 4.5 | 4.7 | 25.97 |
| 3 | Apo | Indica | 5.7 | 5.0 | 95.0 | 94.3 | 104.0 | 104.0 | 23.2 | 21.7 | 9.3 | 8.0 | 5.2 | 5.3 | 26.7 |
| 4 | ASD-17 | Aus | 4.3 | 4.3 | 74.0 | 74.3 | 80.0 | 78.3 | 20.7 | 19.0 | 5.3 | 5.7 | 2.4 | 2.5 | 28.27 |
| 5 | Ashoka | Indica | 3.7 | 3.0 | 83.3 | 81.7 | 122.3 | 124.7 | 19.7 | 18.3 | 6.0 | 6.7 | 2.7 | 3.0 | 25.73 |
| 6 | Azucena | Japonica | 4.3 | 5.0 | 99.7 | 100.3 | 105.3 | 105.8 | 26.7 | 25.0 | 7.7 | 7.3 | 1.3 | 1.4 | 30.2 |
| 7 | Basmati-370 | Aromatic | 5.7 | 5.0 | 103.0 | 103.0 | 123.3 | 123.2 | 22.3 | 21.8 | 8.0 | 7.0 | 2.8 | 3.0 | 17.1 |
| 8 | BHOJ | Indica | 5.7 | 5.0 | 82.3 | 82.3 | 94.3 | 94.0 | 21.7 | 23.0 | 8.3 | 12.7 | 5.3 | 5.5 | 25.4 |
| 9 | Bhuvan | Indica | 6.3 | 7.0 | 91.7 | 90.7 | 114.7 | 115.2 | 22.8 | 22.5 | 9.7 | 8.7 | 3.4 | 3.5 | 24.87 |
| 10 | Birsagora | Aus | 5.7 | 3.7 | 84.3 | 87.3 | 123.3 | 120.0 | 23.2 | 23.2 | 5.0 | 5.7 | 1.6 | 1.7 | 25.17 |
| 11 | Dandi | Indica | 6.3 | 6.3 | 86.7 | 86.3 | 103.3 | 100.7 | 23.3 | 25.7 | 9.3 | 8.3 | 5.1 | 5.4 | 29.2 |
| 12 | Dehula | Indica | 5.7 | 5.0 | 83.7 | 82.3 | 129.0 | 128.7 | 22.3 | 23.2 | 7.3 | 8.3 | 5.6 | 3.6 | 24.37 |
| 13 | Dhalaheera | Indica | 5.7 | 5.0 | 66.7 | 66.7 | 89.0 | 89.5 | 27.0 | 26.8 | 7.3 | 7.3 | 2.4 | 2.4 | 25.43 |
| 14 | Dular | Aus | 5.7 | 7.0 | 80.3 | 81.7 | 115.7 | 112.7 | 26.5 | 27.7 | 5.7 | 7.0 | 3.6 | 3.8 | 25.23 |
| 15 | Durgabhog | Aus | 3.7 | 3.0 | 94.0 | 95.0 | 95.3 | 95.0 | 22.0 | 20.5 | 6.3 | 6.3 | 2.6 | 2.6 | 23.4 |
| 16 | Ganteshwari | Indica | 5.7 | 5.0 | 82.0 | 84.0 | 85.3 | 83.0 | 23.0 | 21.5 | 8.7 | 8.7 | 2.3 | 2.3 | 25.4 |
| 17 | IR-36 | Indica | 5.7 | 4.3 | 76.7 | 76.7 | 91.7 | 90.8 | 22.3 | 22.3 | 9.7 | 9.3 | 4.6 | 4.6 | 25.9 |
| 18 | IR-64 | Indica | 5.7 | 5.0 | 83.7 | 82.7 | 77.0 | 79.3 | 24.3 | 25.8 | 8.3 | 8.3 | 5.4 | 5.4 | 24.33 |
| 19 | JAYA | Indica | 7.0 | 9.0 | 93.7 | 91.0 | 92.3 | 90.3 | 23.3 | 25.2 | 9.0 | 9.0 | 5.1 | 5.4 | 30.57 |
| 20 | Kakro | Landrace | 3.7 | 3.0 | 68.7 | 66.7 | 136.0 | 133.8 | 23.3 | 24.5 | 10.7 | 10.7 | 1.5 | 1.6 | 24.4 |
| 21 | Kalakeri | Aus | 5.7 | 6.3 | 89.0 | 87.0 | 118.3 | 119.7 | 23.0 | 23.2 | 7.3 | 7.3 | 1.7 | 1.7 | 21.93 |
| 22 | KALINGA III | Indica | 5.7 | 4.3 | 83.7 | 80.0 | 120.3 | 120.3 | 22.7 | 22.0 | 10.0 | 10.0 | 1.6 | 1.7 | 23.07 |
| 23 | Khandagiri | Indica | 4.3 | 4.3 | 81.7 | 81.0 | 89.3 | 88.8 | 20.5 | 21.8 | 10.3 | 10.0 | 4.2 | 4.6 | 27.47 |
| 24 | Lalat | Landrace | 4.3 | 3.7 | 91.0 | 89.0 | 103.0 | 102.3 | 24.3 | 26.3 | 8.7 | 9.0 | 4.0 | 5.1 | 25.4 |
| 25 | Lalnakanda-41 | Indica | 3.0 | 3.0 | 76.0 | 76.3 | 127.0 | 126.3 | 31.0 | 30.5 | 11.7 | 10.3 | 1.4 | 1.4 | 24.4 |
| 26 | Lalsar | Indica | 3.7 | 4.3 | 66.0 | 65.7 | 86.7 | 85.0 | 25.7 | 27.3 | 6.7 | 6.7 | 2.2 | 2.2 | 23.97 |
| 27 | Mehar | Indica | 7.0 | 7.0 | 81.7 | 81.7 | 75.3 | 72.7 | 18.3 | 19.3 | 9.7 | 9.3 | 4.5 | 3.8 | 24.67 |
| 28 | N22 | Aus | 3.7 | 3.7 | 82.3 | 80.3 | 99.7 | 99.7 | 22.0 | 22.5 | 9.3 | 9.0 | 3.6 | 2.9 | 19.8 |
| 29 | Naveen | Indica | 5.0 | 5.0 | 94.7 | 94.7 | 96.7 | 97.0 | 25.3 | 24.7 | 7.7 | 8.7 | 3.8 | 5.2 | 25.17 |
| 30 | Ranjit | Indica | 5.0 | 5.0 | 80.3 | 80.7 | 82.3 | 82.0 | 21.0 | 22.0 | 12.7 | 12.7 | 4.8 | 4.9 | 28.23 |
| 31 | Ratna | Indica | 3.0 | 3.0 | 80.3 | 80.3 | 92.7 | 92.3 | 23.3 | 24.0 | 10.0 | 10.7 | 4.8 | 4.8 | 24.87 |
| 32 | Sadabahar | Indica | 4.3 | 4.3 | 87.3 | 88.0 | 94.3 | 90.7 | 22.0 | 22.0 | 9.7 | 8.3 | 4.2 | 4.2 | 27.2 |
| 33 | Sahabhagi | Indica | 5.0 | 5.0 | 92.7 | 94.7 | 95.0 | 95.5 | 22.7 | 23.2 | 12.0 | 14.0 | 4.5 | 4.6 | 27.17 |
| 34 | Saita | Indica | 7.0 | 7.0 | 83.7 | 85.7 | 108.3 | 106.7 | 22.3 | 23.8 | 9.0 | 9.3 | 1.4 | 1.4 | 28.13 |
| 35 | Saket-4 | Indica | 3.7 | 4.3 | 95.0 | 94.7 | 91.0 | 88.0 | 23.2 | 23.5 | 9.3 | 9.0 | 3.6 | 4.7 | 23.53 |
| 36 | Samanta | Indica | 3.0 | 3.0 | 86.3 | 84.0 | 93.0 | 93.7 | 25.3 | 25.0 | 10.0 | 10.3 | 2.1 | 2.1 | 25.37 |
| 37 | Satabdi | Indica | 5.0 | 5.0 | 88.3 | 89.7 | 83.3 | 83.0 | 23.8 | 25.5 | 13.3 | 13.0 | 2.3 | 2.2 | 23.83 |
| 38 | Sathi | Landrace | 5.0 | 5.0 | 77.7 | 80.3 | 117.0 | 116.8 | 22.3 | 22.7 | 5.0 | 5.3 | 1.1 | 1.2 | 24.27 |
| 39 | Sukhawan | Indica | 4.3 | 4.3 | 88.0 | 90.0 | 96.7 | 99.3 | 22.5 | 23.5 | 8.3 | 8.3 | 2.8 | 2.8 | 25.47 |
| 40 | Udayagiri | Indica | 7.0 | 7.0 | 87.0 | 87.7 | 87.0 | 86.8 | 24.2 | 24.5 | 10.7 | 9.7 | 4.6 | 3.6 | 27.03 |
| 41 | Vasumati | Indica | 6.3 | 6.3 | 96.0 | 98.0 | 101.0 | 101.3 | 24.7 | 25.3 | 6.7 | 8.3 | 2.2 | 2.3 | 25.01 |
| 42 | Virendra | Indica | 5.0 | 5.0 | 83.7 | 81.3 | 103.0 | 101.0 | 23.3 | 25.3 | 8.7 | 9.0 | 3.6 | 4.6 | 26.23 |

EVV: Early vegetative vigor; DTF: Days to flowering; PH: Plant height; PL: Panicle length; EBT: Ear bearing tillers; GY: Grain yield; 1000-GW: Thousand grain weight; _1 and _2 represent two consecutive crop seasons; *In centimeters; #Grain yield in two consecutive crop seasons; @Mean of five independent determinations at 5% level of significance

**Table 1b.** Statistical analysis presenting variability for phenotypic traits

| Trait | Mean | SD based on total SS | SD based on residual SS | CV (%) |
|---|---|---|---|---|
| EVV | 4.9536 | 1.26 | 0.511 | 10.3 |
| DTF | 84.918 | 8.21 | 1.159 | 1.4 |
| PH | 100.29 | 15.35 | 1.204 | 1.2 |
| PL | 23.557 | 2.29 | 0.811 | 3.4 |
| EBT | 8.7929 | 1.94 | 0.713 | 8.1 |
| GY | 3.3881 | 1.37 | 0.381 | 11.2 |
| TKW | 25.42 | 2.43 | 0.657 | 2.59 |

**Table 2.** Analysis of variance among the genotypes

| Source | df | SS | MS | Est. Var. | % |
|---|---|---|---|---|---|
| Among populations | 1 | 31.519 | 31.519 | 0.91 | 18% |
| Among individuals | 40 | 317.945 | 7.949 | 3.754 | 74% |
| Within individuals | 42 | 18.5 | 0.44 | 0.44 | 9% |
| Total | 83 | 367.964 | | 5.104 | 100% |

Fig. 1.

Thirty six unlinked SNP markers distributed on

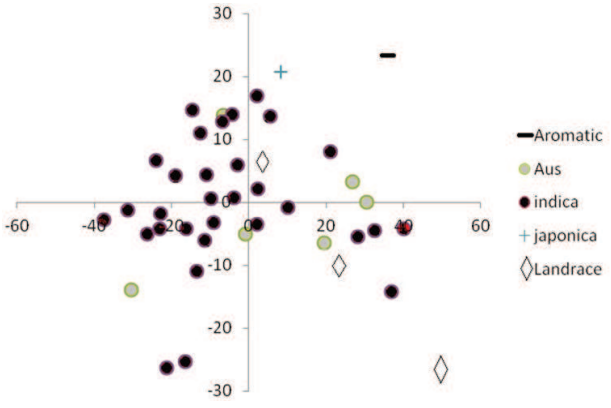


**Fig. 1. Principal component analysis of 42 rice genotypes based on grain yield related traits**

all 12 chromosomes were used for the genetic diversity analysis. A total of 64 alleles were amplified with an average of 2 alleles per locus in all the varieties. Four SNPs were found to be monomorphic. Alleles were scored as wild, heterozygous and alternate alleles across all the 42 rice varieties and the percent allele was calculated for each loci. Major allele frequency ranged from 0.53 to 0.97 with an average of 0.80. Average PIC value was observed to be 0.22, with highest PIC value (0.37) recorded for SNP primer 11-521_C_214 located on chromosome 11, while primers 11-1849 and 11-3935, both on chromosome 11 recorded lowest PIC value (0.04). Gene diversity ranged from 0.04 to 0.49 with an average gene diversity of 0.27. The average value of PIC for SNP was 0.22, similar reports of low PIC value for SNP marker was also reported in rice by (Chen et al. 2011). This low average PIC indicates that lower genetic diversity exists among the rice varieties. Heterozygosity ranged from 0.0 to 0.16 with an average value of 0.02. Highest number of heterozygous alleles (17%) were found associated with marker 10-188-1 located on chromosome 10. Details of the thirty two polymorphic SNP primers along with their allelic profile are listed in Table 3. Performing Analysis of Molecular Variance (AMOVA) with obtained SNP data showed that highest diversity existed at individual level (73%), while there was only 9% diversity level within individuals. At the population level, 18% diversity existed, highest diversity exists at individual level (73%) and lowest diversity exists within individual (9%) (Table 3). Genetic diversity in the present study was based on the SNPs derived from single copy genes. Such an approach to perform overall diversity analysis should provide robust estimate of genetic base of a breeding germplasm pool.

In order to study genetic relatedness between rice varieties, clustering pattern was seen in un-rooted tree. For clustering similar genotypes, NJ tree was constructed using dissimilarity matrix. Unrooted tree was drawn in order to predict genetic relationships among the 42 rice varieties (Fig. 2). The genotypes were grouped into four different clusters *viz*., I, II, III and IV. Clusters 1, 2, 3 and 4 harbored 2, 5, 13 and 4 varieties, respectively. There were six *aus* varieties; four were present in Cluster 2, one in cluster 3 and one in cluster 4. Further, to better understand the population structure and genetic relationship among individual rice varieties, a model based program STRUCTURE V2.3.1 was used. Genotypes were categorized as pure and admixed; varieties with >0.80 score were considered as pure while those scoring less than 0.80 were categorized as admixture. Populations one, two, three and four had 4, 6, 4 and 28 varieties respectively. Population 1, 2 and 3 were

**Table 3.** Details of SNP primers and genotyping data of 42 rice varieties

| Chr.* | Primer ID | Wild alleles (%) | Alternate alleles (%) | Heterozygous allele (%) | Frequency of major allele | Gene diversity | Hetero-zygosity | PIC |
|---|---|---|---|---|---|---|---|---|
| 1 | 01-3916-1_C_156.fasta | 69 | 29 | 2 | 0.7 | 0.42 | 0.02 | 0.33 |
| | 01-608-4_C_375.fasta | 62 | 33 | 5 | 0.64 | 0.46 | 0.05 | 0.35 |
| | 01-6351-1_C_202.fasta | 88 | 12 | 0 | 0.88 | 0.21 | 0 | 0.19 |
| 2 | 02-267.fasta | 67 | 33 | 0 | 0.67 | 0.44 | 0 | 0.35 |
| | 02-3029-1_C_474.fasta | 64 | 33 | 2 | 0.65 | 0.45 | 0.02 | 0.35 |
| | 02-4333-1_C_293.fasta | 86 | 14 | 0 | 0.86 | 0.24 | 0 | 0.21 |
| 3 | 03-1691-1_C_373.fasta | 79 | 21 | 0 | 0.79 | 0.34 | 0 | 0.28 |
| | 03-3478-1_C_206.fasta | 67 | 29 | 5 | 0.69 | 0.43 | 0.05 | 0.34 |
| | 03-4660-1_C_355.fasta | 90 | 10 | 0 | 0.9 | 0.17 | 0 | 0.16 |
| 4 | 04-1801-20_C_428.fasta | 95 | 5 | 0 | 0.95 | 0.09 | 0 | 0.09 |
| | 04-3787-3_C_358.fasta | 81 | 17 | 2 | 0.82 | 0.29 | 0.02 | 0.25 |
| 5 | 05-2692-1_C_109.fasta | 83 | 17 | 0 | 0.83 | 0.28 | 0 | 0.24 |
| | 05-4192-1_C_280.fasta | 64 | 31 | 5 | 0.67 | 0.44 | 0.05 | 0.35 |
| | 05-48-1_C_279.fasta | 93 | 5 | 2 | 0.94 | 0.11 | 0.02 | 0.11 |
| 6 | 06-1776-1_C_501fasta | 86 | 12 | 2 | 0.87 | 0.23 | 0.02 | 0.2 |
| | 06-2509-1_C_497.fasta | 93 | 2 | 2 | 0.96 | 0.07 | 0.02 | 0.07 |
| 7 | 07-2904-39_C_299.fasta | 67 | 26 | 7 | 0.7 | 0.42 | 0.07 | 0.33 |
| | 07-293-12_C_368.fasta | 90 | 10 | 0 | 0.9 | 0.17 | 0 | 0.16 |
| | 07-4304_new | 57 | 38 | 5 | 0.6 | 0.48 | 0.05 | 0.37 |
| 8 | 08-2765-2_C_360.fasta | 93 | 2 | 5 | 0.95 | 0.09 | 0.05 | 0.09 |
| | 08-4218-5_C_129.fasta | 69 | 26 | 5 | 0.71 | 0.41 | 0.05 | 0.32 |
| | 08-847-6_C_113.fasta | 67 | 31 | 2 | 0.68 | 0.44 | 0.02 | 0.34 |
| 9 | 09-209.fasta | 95 | 5 | 0 | 0.95 | 0.09 | 0 | 0.09 |
| | 09-2107-5_C_145.fasta | 95 | 5 | 0 | 0.95 | 0.09 | 0 | 0.09 |
| | 09-2716-4_C_457.fasta | 83 | 14 | 2 | 0.85 | 0.26 | 0.02 | 0.23 |
| 10 | 10-1192-7_C_178.fasta | 55 | 38 | 7 | 0.58 | 0.49 | 0.07 | 0.37 |
| | 10-188-1.fasta | 81 | 2 | 17 | 0.89 | 0.19 | 0.17 | 0.17 |
| | 10-2723.fasta | 86 | 12 | 2 | 0.87 | 0.23 | 0.02 | 0.2 |
| 11 | 11-1849.fasta | 98 | 2 | 0 | 0.98 | 0.05 | 0 | 0.05 |
| | 11-3935.fasta | 98 | 2 | 0 | 0.98 | 0.05 | 0 | 0.05 |
| | 11-522-1_C_214.fasta | 50 | 43 | 7 | 0.54 | 0.5 | 0.07 | 0.37 |
| 12 | 12-400.fasta | 95 | 5 | 0 | 0.95 | 0.09 | 0 | 0.09 |
| | Mean | 0.81 | 0.27 | 0.03 | 0.22 | | | |

Chr. = Chromosome

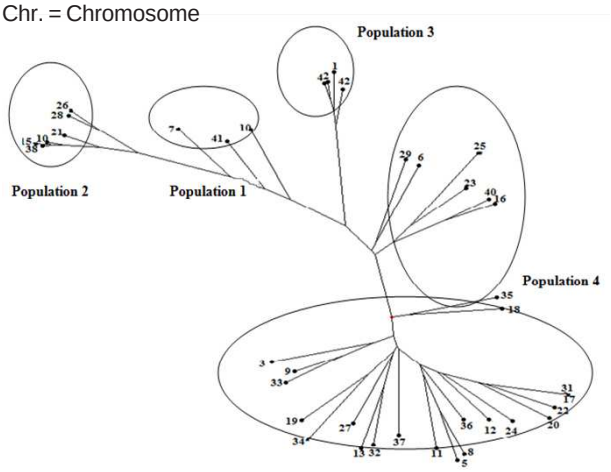


**Fig. 2. NJ tree constructed for SNP data (4, 10, 14, 15, 21, 28 are *aus* varieties)**

pure and population 4 contained six admixed varieties (Figs. 3 and 4). Varieties in cluster 2 and cluster 3 of NJ tree correspond to varieties in population 2 and population 3. Varieties in cluster 1 and 4 correspond to population 1 and 4 with few exceptions. Cluster 2 and population 2 in diversity and population structure analysis were therefore specific to *aus*. Also, diversity analysis suggested that the *aus* and *indica* group of accessions are genetically highly closer to each other. Varieties were grouped into four clusters, but *indica* and *aus* varieties were mainly grouped together. Principal Coordinate Analysis (PCoA) exhibited large diversity among the 42 rice varieties.

Phenotypic diversity was estimated based on

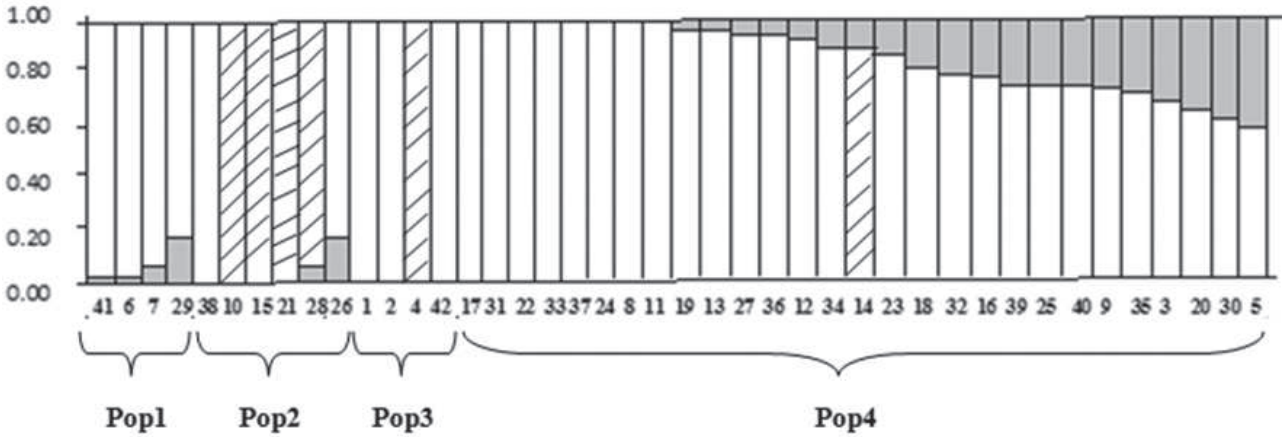


**Fig. 3. Model based clustering of SNP (K=4) of *indica* rice varieties, (varieties showing lines indicates *aus* varieties and shaded portion show percentage of admixture)**

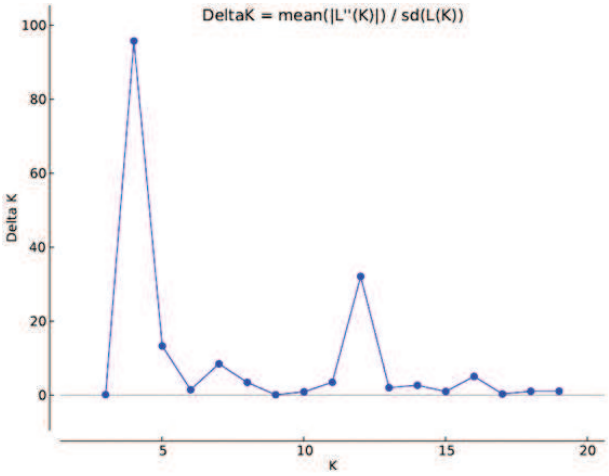


**Fig. 4. Estimation of population using LnP(D) derived Δk for k from 1 to20**

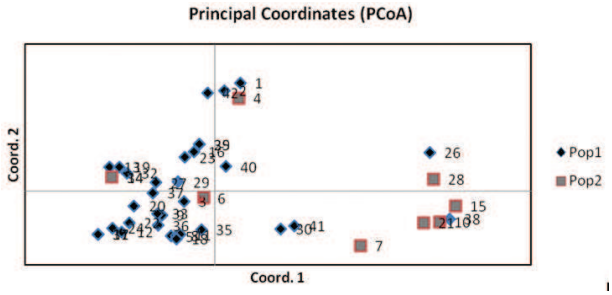


**Fig. 5. Principal Coordinate Analysis (PCoA) of 42 rice varieties based on SNP data (Population1 indicates *indica* varieties and Population 2 indicates *aus* varieties)**

grain yield related parameters of rice accessions subjected to analysis. The principal component analysis performed using grain yield related traits did not revealed clear groups and no specific cluster of *aus*/*indica* was observed (Fig. 1). However, another classification led some of the *aus* to form a clear cluster (Fig. 5). Interestingly, some of the *aus* accessions grouped with *indicas* (Fig. 5). These results suggest that that using *Aus* and landrace cultivars in breeding programs don't have a detrimental effect on yield. As an example, well known rice landrace, Durgabhog is grouped with a recently released variety, Sahbhagi. This study provides an indication, however, extended work using large number of individuals and traits would further confirm. Usually landraces are known for the low yield potential and not preferred by breeders for development of the high yielding rice varieties. However, our analysis with the grain yield related traits clearly suggest here that not necessarily we may have a detrimental effect on yield potential after using *aus* cultivars and the landraces. It could therefore be recommended that following a focused breeding strategy for yield enhancement in rice using *Aus* cultivars and/or landraces may help significantly. It is interesting to note that genetic classification of rice reported earlier by Kato et al. (1928) and Glazman et al. (1987) suggested that there are two broad groups-*indica* and *japonica*. Genotypic as well as phenotypic diversity results obtained in the present study are in full agreement of this. Overall, we recommend large scale use of *Aus* and traditional germplasm groups in rice for genetic improvement.

**Authors' contributions**

Conceptualization of research (PV, SS); Designing of the experiments (NS, SS, PV); Contribution of experimental materials (NS, SS); Execution of field/ lab experiments and data collection (NS, SS); Analysis of data and interpretation (NS, SM); Preparation of the manuscript (NS, SM, PV).

**Declaration**

Authors declare no conflict of interest.

**Acknowledgements**

The authors are thankful to the various regional farmers of Uttar Pradesh from where the rice varieties were collected. We are also thankful to ICAR for providing financial assistance to conduct the field trials and research work. We like to thank Dr. V. K. Singh for his support in providing useful suggestions during course of study.

**References**

Agrama H. A., Yan W., Jia M., Fjellstrom R. and McClung A. M. 2010. Genetic structure associated with diversity and geographic distribution in the USDA rice world collection. Natural Sci., **2**(4): 247-291.

Akagi K., Sandig V., Vooijs M., Van der Valk M., Giovannini M., Strauss M. and Berns A. 1997. Pre-mediated somatic site-specific recombination in mice. Nucleic Acids Res., **25**: 1766-1773.

Arunachalam V. 1981. Genetic distance in plant breeding. Indian J. Genet., **41**: 226-236.

Bautista N. S., Solis R., Kamijima O. and Ishii T. 2001. RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice genes. Genet. Syst., **76**(2)**:** 71-79.

Chen H., He H., Zou Y., Chen W. and Yu R. 2011. Development and application of a set of breeder friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). Theor. Appl. Genet., **123**: 869-879.

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

Glaszmann J. C. 1987. Isozymes and clasification of rice varieties. Theor. Appl. Genet., **74**: 21-30.

Harlan J. R. 1975. *Crops and Man*. Madison, Wisconsin: American Society of Agronomy and Crop Science Society of America.

Hartl D. L. and Clark A. G. 1997. Principles of population. Genetics. Sinauer Associates, Inc. Sunderland, MA. 542 pp.

Herrera T. G., Duque D. P., Almeida I. P., Núñez G. T., Pieters A. J., Martinez C. P. and Tohme J. M. 2008. Assessment of genetic diversity in Venezuelan rice cultivars using simple sequence repeats markers. Electron. J. Biotechnol., **11**(5): 1-14.

http://drdpat.bih.nic.in/Rice%20Varieties%20in%20India.htm

Kato S., Kosaka H. and Hara S. 1928. On the affinity of rice varieties as showed by fertility of hybrid plants, Bulletin Of Science, Faculty of Agronomy, Kyushu Imperial University, **3**(2): 132-147.

Krishnamurthy S. L., Sharma S. K., Kumar V., Tiwari1 S., Batra V. and Singh N. K. 2014. Assessment of genetic diversity in rice genotypes for salinity tolerance using Saltol markers of Chromosome 1. Indian J. Genet., **74**(2): 243-247. DOI: 10.5958/0975-6906.2014.00167.9.

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

Londo J. P., Chiang Y. C., Hung K. H., Chiang T. Y. and Schaal B. A. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice *Oryza sativa*. PNAS, **103**(25): 9578-9583.

Loveless M. D. and Hamrick J. L. 1984. Ecological determinants of genetic structure in plant populations. Annual Rev. Ecol. Syst., **15**: 65-95.

Mackill D. J. 1995. Plant genetic resources. *Classifying japonica rice cultivars with RAPD markers*. Crop. Sci., **35**: 889-894.

McNally K. L., Childs K. L., Bohnert R., Davidson R. M., Zhao K., Ulat V. J., Zeller G., Clark R. M., Hoen D. R. and Bureau T. E. 2009. Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. PNAS, **106**: 12273-12278.

Olufowote J. O., Xu Y., Chen X., Park W. D., Beachell H. M., Goto M. and McCouch S. R. 1997. Comparative evaluation of within cultivar variation of rice (Oryza sativa L.) using microsatellite and RFLP markers. Genome., **40**: 370-378.

Pachauri V., Taneja N., Vikram P., Singh N. K. and Singh S. 2013. Genetic diversity analysis of Indian farmers' rice (*Oryza sativa* L.) varieties based on microsatellite markers. A.J.C.S., **7**(7): 923-932.

Panaud O., McCouch S. R. and Chen X. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol. Gen. Genet., **252**: 597-607.

Peakall R. and Smouse P. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for

teaching and research - an update. Bioinformatics, **28**: 2537-2539.

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

Rathi Sunayana, Yadav R. N. S., Pathak K. and Sarma R. N. 2014. Genetic diversity in upland rice of Assam assessed through SSR markers. Indian J. Genet., **74**(2): 238-242. DOI: 10.5958/0975-6906.2014. 00161.8

Ray D., Mueller N., West P. and Foley J. 2013. Yield trends are insufficient to double global crop production by 2050. PLoS One, **8**(6): e66428.

Second G. 1982. Origin of the genic diversity of cultivated rice (*Oryza* spp.). Study of the polymorphism scored at 40 isozyme loci. Jpn. J. Genet., **57**: 25-57.

Thanh N. D., Zheng H. G., Dong N. V., Trinh L. N., Ali M. L. and Nguyen H. T. 1999. Genetic variation in root morphology and microsatellite DNA loci in upland rice (*Oryza sativa* L.) from Vietnam. Euphytica, **105**: 43-51.

Upadhyay P., Singh V. K. and Neeraja C. N. 2011. Identification of Genotype Specific Alleles and Molecular Diversity Assessment of Popular Rice (*Oryza sativa* L.) Varieties of India. Int. J. Plant Breed. Genet., **5**: 130-140.

Vikram P., Kadam S., Singh B. P., Lee Y. J., Pal J. K., Singh S., Singh O. N., Swamy B. P. M., Thiyagarajan K., Singh S. and Singh N. K. 2016. Genetic Diversity Analysis Reveals Importance of Green Revolution Gene (*Sd1* Locus) for Drought Tolerance in Rice. Agric. Res., **5**(1): 1-12.

Vikram P., Swamy B. P. M., Dixit S., Singh R., Singh B. P., Miro B., Kohli A., Henry A., Singh N. K. and Kumar A. 2016. Drought susceptibility of modern rice varieties: an effect of linkage of drought tolerance with undesirable traits. Nat. Sci. Rep., **5**: 14799.

Vikram P., Swamy B. P. M., Dixit S., Trinidad J., Sta Cruz M. T., Amante M. and Kumar A. 2016. Linkage and Interaction analysis of major effect drought grain yield QTLs in rice: PLoS One, **11**(3): e0151532.

Wu K. S. and Tanksley S. D. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet., **241**: 225-235.

Xiao J., Li J., Yuan L., McCouch S. R. and Tanksley S. D. 1996. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR based markers. Theor.Appl.Genet., **92:** 637-643.

Yang G. P., Maroof M. A. S., Xu C. G., Zang Q. and Baiyashev R. M. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Genet., **245**: 187-194.

Zhao K., Tung C., Eizenga G. C., Wright M. H., Ali M. L., Price A. H., Norton G. J., Islam M. R., Reynolds A., Mezey J., McClung A. M., Bustamante C. D. and McCouch S. R. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat. Com., **2**: 467