DOI 10.5958/0976-1926.2017.00037.7

SNP Marker Based Genetic Diversity and Population Structure Study of Rice Germplasm of Arunachal Pradesh

Debjani Roy Choudhury[1], Nivedita Singh[1], RK Tyagi[2], Kalyani Srinivasan[2], NK Singh[3] and Rakesh Singh*[1]
[1]*Division of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources (ICAR-NBPGR), New Delhi-110012*
[2]*Division of Germplasm Conservation, ICAR-NBPGR, New Delhi-110012*
[3]*ICAR-NRC on Plant Biotechnology, IARI, New Delhi-110012*

Rice genetic resource is the primary source for rice breeding. It also makes a concrete contribution to global wealth creation and food security. India is one of the centers of origin for rice and the nutritional security of the North-Eastern (NE) states of India predominantly depends on rice. The NE region especially Arunachal Pradesh is considered to be one of the hot pockets of rice genetic resources in the world and also has diverse rice growing conditions as compared to other parts of the country. Genetic diversity of 662 rice accessions originating from the state of Arunachal Pradesh was assessed using 36 single nucleotide polymorphism (SNP) markers distributed across all 12 chromosome of rice genome. These SNP markers generated a total of 36 allelic combinations. Polymorphic Information content (PIC) ranged from 0.006 to 0.375 and major allele frequency ranged from 0.51 to 0.99. Similarly, heterozygosity and gene diversity ranged from 0.003 to 0.162 and 0.006 to 0.49, respectively. A picture of genetic relatedness was inferred using Neighbor joining tree, which grouped all the genotypes in to three major clusters. AMOVA analysis showed that, among population 31% whereas, among individual 51% diversity. Population structure from K=1 to K=20 were tested and Δk was found maximum at three population (K=3). Principal Coordinate Analysis (PCoA) showed that rice germplasm were intermixed as per Neighbour joining tree but showed three major groups based three population structure information. Hence, PCoA supports the population obtained from the model based approach. This study shows that large diversity exists in the rice germplasm of Arunachal Pradesh and can be utilized for trait-specific breeding.

Key Words: Genetic diversity, Population structure, Rice, SNP markers

Introduction

Rice (*Oryza sativa*) feeds more than 50% of the world's population and is one of the most important crops in the world. Rice accounts for 21, 14 and 2% of global energy, protein and fat supply respectively (Kennedy *et al.,* 2003). Being a model organism with fully sequenced genome, various genomic approaches have been used to study its domestication, adaptive selection, and the history of crop improvement (Wing *et al.,* 2005, Zhang *et al.,* 2008). Rice genetic resource is the primary material for rice breeding and makes a concrete contribution to global wealth creation and food security (Zhang *et al,.* 2011). Numerous studies on genetic structure of rice cultivars at a local scale (within a country) have also been conducted (Gao *et al*., 2005, Barry *et al*., 2007, Thomson *et al.,* 2009, Zhang *et al.,* 2007, Li *et.al.,* 2014). Such local-scale studies not only provide a detail view of rice genetic diversity within a country, but it also gives a better understanding of complex interaction between rice genetic diversity and human cultivation practices (Thomson *et al.,* 2009), and for formulating *in situ* conservation strategies (Barry *et al.,* 2007). India is one of the centres of origin for rice. Phylo-geographical and archaeological evidence suggest that rice was domesticated about 10000 years ago from its wild ancestor *O. rufipugon* in the region south of Himalayan mountain range, likely in the present day Eastern and North Eastern India, extending eastward to Nepal, Myanmar and Thailand to Southern China (Chang, 1976; Khush 1997; Londo *et al.,* 2006) the eastern Himalayan region of North-east India, a geographical area of over 255000 Km^2 consisting of Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland and Tripura states is home to a large number of indigenous rice varieties (Choudhury *et al.,* 2013). The state of Arunachal Pradesh is situated on the eastern most corner of India sharing international borders with Bhutan, China and Myanmar. It has a geographical area of 83,743 sq. km. The state has a population density of 17 persons per sq. km, and around 68.8% of the population is tribal (Govt. of India, Statistics, 2015). The tribal communities of Arunachal Pradesh are of Tibeto-Burman linguistic origin, and there are as many as 21 tribes and 50 sub-tribes. The topography of the

**Author for Correspondence: Email- rakesh.singh2@icar.gov.in*

state is mostly hilly terrain and the agro climate varies from tropical to alpine (Roy *et al.,* 2016). The farmers majorly follow traditional shifting cultivation (locally known as jhum) and with little sedentary agriculture. Shifting or 'slash-and-burn' cultivation is the earliest form of agriculture and is still practiced in vast areas of the state. About 76 % of the total cropped area in the state is under jhum cultivation. However rice productivity in this state is low, 2065kg/ha, as rice is cultivated in high altitudes, plateaus, terraces and river beds. Despite having low yield potential, rice landraces grown in the mountains under jhum cultivation system possess many important biotic, abiotic and quality traits which can be important source for crop improvement programmes. Thus, keeping in view the above points, the genetic diversity of rice germplasm of Arunachal Pradesh was assessed using SNP markers, the most advanced set of markers reported till now.

Materials and Methods

Plant Materials

Six hundred and sixty two seed samples of Arunachal Pradesh were procured from Indian National Genebank, National Bureau of Plant Genetic Resources (NBPGR),New Delhi. The details of each accession alongwith passport data (national ID i.e. Indigenous Collection (IC) number) were identified from the online available data of NBPGR.

DNA Extraction from Rice Seed

Total Six to nine seeds of each accession were dehusked and used for DNA isolation using QIAGEN DNeasy plant mini kit. Kernels were ground into fine powder using tissue lyser (Tissue lyser II Retsch, Germany) with a tissue lyser adapter set (QIAGENq). DNA extraction procedure was as per manufacturer's protocol.

Genotyping of Rice Accessions using SNP Markers

Genomic DNA of all the 662 accessions was diluted to prepare working stocks of 10 ng/µl. The Sequenom MassARRAY system uses matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometer for accurate detection of SNPs in a high-throughput manner (www.sequenom.com). Sequenom MassARRAY multiplex assays were designed for 36 SNPs (iPLEX gold chemistry), representing conserved single-copy rice genes (Singh *et al.,* 2007), taking three genes per rice chromosome. The unlinked SNP markers located on short arm, centromeric region and long arm of all 12 rice chromosomes were developed and used for diversity analysis (Table 1). The 36-plex assays were designed and validated by Sequenom Corporation (San Diego). The 30-mer pre-amplification primers and variable length genotyping primers generated by the Assay Design 3.1 software were procured and used for the validation of SNPs according to the Sequenom user manual. MassARRAY Typer 3.4 Software was used for the visualization of SNPs and allele calling.

Genetic Diversity and Phylogenetic Analyses

The major allele frequency, gene diversity, heterozygosity and PIC for each locus were calculated for SNP markers using Power Marker 3.5 (Liu and Muse, 2005). In addition, principal component analyses, genetic distances (Nei *et al*., 1983) across the genotypes and neighbour-joining (NJ) tree were calculated using Power Marker 3.5 (Liu and Muse, 2005). The dissimilarity matrix generated by Power Marker was used to construct un-weighted neighbour joining tree using DARwin software 5.0.158 (Perrier and Jacquemoud-Collet, 2006). Software STRUCTURE V2.3.1 was applied to infer historical lineages that show clusters of similar genotypes (Pritchard *et al*., 2000). The membership of each genotype was run for range of genetic clusters from value of K= 1 to 20 with the admixture model and correlated allele frequency. For each K it was replicated 3 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 and then plotted to find the plateau of the ΔK values (Evanno *et al.* 2005). To calculate final population structure online available programme "structure harvester" (Earl and vonHoldt, 2012) was used (http: //taylor0. biology.ucla.edu). The proportion of the genome of an individual that belongs to each inferred population (admixture) was also estimated.

Results and Discussion

Statistical Analysis

Genetic diversity of Arunachal Pradesh rice germplasm collection (662 accessions), available at National Gene Bank, NBPGR, New Delhi was estimated using 36 SNP markers. These unlinked SNP markers located on the short arm, centromeric region and long arm of all 12 rice chromosomes were developed and used for diversity analysis.SNP is bi-allelic in nature therefore; total 72 alleles were amplified with an average of 2 alleles per

Table 1. List of primers used for characterization of 662 rice germplasm of Arunachal Pradesh along with their physical position on chromosome and primer sequences

Primer ID	Physical Position	Amplification primer-1	Amplification primer-2
01-2916-1_C_156	25381654	ACGTTGGATGGGGTTTGCATGTTAATAGGG	ACGTTGGATGCCGAATCTCTATCAAGGAAG
01-608-4_2_375	3421011	ACGTTGGATGAGGACCATCTTCTTGCACTG	ACGTTGGATGCCATTTGCAAGGCCCATTTC
01-6351-1_C_202	40914292	ACGTTGGATGGTTGGAACACATGATTTCAC	ACGTTGGATGATCTCTTTGGACAGAGTCCC
02-267	1570149	ACGTTGGATGGTCAATCTTGCAGGAGTTGG	ACGTTGGATGTGGCTCCTCTTCTCCGGTCT
02-3029-1_C_474	18821156	ACGTTGGATGTGTCTGCAATAACTTGTGCC	ACGTTGGATGAAATCAGCTGCAGCATTACC
02-4333-1_2_293	28688819	ACGTTGGATGGGAATGTTTAGTTTTGAGG	ACGTTGGATGTGTAGGTGCTACTTGCTTCC
03-1691-1_C_373	10849512	ACGTTGGATGAACAACGCCAGGAACATCAC	ACGTTGGATGAAGCGGCTCAAGGTACAATC
03-3478-1_C_206	22815422	ACGTTGGATGCCTGCAGCAAACGCCAATTT	ACGTTGGATGTCAGGTAACCGATCGATTTG
03-4660-1_2_355	31020366	ACGTTGGATGCTCCCATCCTAGTATCCATC	ACGTTGGATGTGCCTTCTCTTACAGGTTCC
04-1801-20_2_428	11859836	ACGTTGGATGCCCTCAAAAAAAAGTTGTAAG	ACGTTGGATGCAGTAAATTTCCAGGGAGATA
04-19-4_C_240	225838	ACGTTGGATGTCTACACATTAGCTCGCTGG	ACGTTGGATGACAGTAACCACAATATGCCG
04-3787-3_1_358	25211800	ACGTTGGATGTTATCTCTGCTTGCTCGCTC	ACGTTGGATGAAGTATCTGCCCCAAGTGAC
05-2692-1_2_109	18783426	ACGTTGGATGGAACTTTACTCTCAGTACA	ACGTTGGATGTGGTTTGATGAGTCGTTTGC
05-4192-1_C_280	28065769	ACGTTGGATGAGTTTGTTGACAGCAGAACC	ACGTTGGATGTAGCTTACTAGTTCATGTG
05-48-1_C_279	287362	ACGTTGGATGCAGAGATGTCTGTTGTTAGC	ACGTTGGATGCAACCAGGGATACAATATGAC
06-1256-1_C_147	7573979	ACGTTGGATGCACGTGCCTATGATTAGCAG	ACGTTGGATGGATCGTTTACTTCTTTGCCC
06-1776-1_1_501	11093772	ACGTTGGATGGGGCCAATTTGCTTAGTGC	ACGTTGGATGAGCATAAGGTATTAAAGTC
06-2509-1_C_497	15737387	ACGTTGGATGCCTTCGCGCTTGCAATTTGG	ACGTTGGATGAAATCAGCACGCGTCAACAC
07-2904-39_C_299	19160255	ACGTTGGATGAATGGTGGTGTATCTTGAGC	ACGTTGGATGGGTGTGACTTCTCATGACAG
07-293-12_1_368	1859603	ACGTTGGATGCACTAATTCTTGGTATTATGG	ACGTTGGATGTCAATGTGTTCTCACAGACC
07-4304_new			
08-2765-2_C_360	18084851	ACGTTGGATGTCCCTCCATGTTGTGAGTTC	ACGTTGGATGCTTGCAAGAGACATCCAAGA
08-4218-5_C_129	27692470	ACGTTGGATGGGTGGACAAAGATAAGGAAG	ACGTTGGATGGACTGGAAATATACTCCCTC
08-847-6_C_113	5399913	ACGTTGGATGCCCAACGTATTAATGGCAAC	ACGTTGGATGGCTGTGTAGTAATTTGCCTG
09-209	1297966	ACGTTGGATGGAGGCAAAAGGCAAACCGAC	ACGTTGGATGGACTTGAGCGAGTCGATGTC
09-2107-5_C_145	13705487	ACGTTGGATGTGACCACACCACACAAACAC	ACGTTGGATGGGGATTTGCGGTTTTTGGAC
09-2716-4_C_457	19541336	ACGTTGGATGTGAGCCACAGATTCCCTTTC	ACGTTGGATGCTCGAGTAATTCAAAACCAC
10-1192-7_C_178	8122635	ACGTTGGATGCTTTGCTACGGATAAAATG	ACGTTGGATGTCATGCAAATACAGACATGG
10-188-1.	1218215	ACGTTGGATGGCGCCAGTGTATGGAAAAAG	ACGTTGGATGGTCCATAACATCATGGACTC
10—2723	20696970	ACGTTGGATGCCCACAATGAGATGCAGATG	ACGTTGGATGAGACAAAATGCAACACTCCG
11-1849.	11974790	ACGTTGGATGCGCCACTCTTCCTGATTTAG	ACGTTGGATGACAGATACGGGAGGCATTTC
11-3935.	28434679	ACGTTGGATGATCCCTGAGACTTTGGATGG	ACGTTGGATGCCAACTTGAATGTCCATTCC
11-522-1_C_214.	3033366	ACGTTGGATGCTACATGGTATCAGATACCG	ACGTTGGATGAGAAGCGAACGCGGAAAAAG
12-1794	11215946	ACGTTGGATGGTGAGCCCCAAAAGTTGGTG	ACGTTGGATGTAAGGTCCAGTTTGCTTGGT
12-3200-2_C_389	21396181	ACGTTGGATGGCTCAAACCTAGCAATAACTG	ACGTTGGATGCCTCCTTCCTACAAGTTTAA
12-400	2160546	ACGTTGGATGCCAATAGAGTCCATCTCAGC	ACGTTGGATGGCACGAGGATTTAAGACAGC

locus in all tested rice accessions. The Polymorphic Information content (PIC) ranged from 0.006 for marker 04-19-4_C_240 to 0.375 for marker 01-608-4_2_375 (Table 2). Major allele frequency ranged from 0.51 for marker 01-608-4_2_375 to 0.99 for 04-19-4_C_240. Similarly, heterozygosity and gene diversity ranged from 0.003(04-19-4_C_240.) to 0.162 (10-1192-7_C_178) and 0.006 (04-19-4_C_240.) to 0.49 (01-608-4_2_375.) with mean values 0.07 and 0.29, respectively (Table 2).The amount of genetic diversity and the population structure for the popular rice genotypes of India are extremely important to know in terms of the extent of genetic variability. This is also useful in order to develop effective breeding strategies for broadening the genetic base of commercial varieties, to identify molecular tags, and for germplasm conservation (Upadhyay *et al.*, 2012). Average PIC has been reported to 0.23 in this study which is lower to those compared to studies done by Lu *et al.*, 2005; Pervaiz *et al.*, 2009 and Lin *et al.*, 2012. This could be due to SNP markers used in the present study. Mean PIC was reported to be high (>0.5) by Anupam *et al.* (2017) for the state of Tripurawith trait linked markers. The value of gene diversity in the present study is reported to be 0.29 which is low as compared to the results obtained by Anupam *et al.*, 2017

Table 2. List of SNP primers used for genotyping of 662 rice germplasm of Arunachal Pradesh along with major allele frequency, gene diversity, heterozygosity and PIC

S. No.	Primer ID	Major allele frequency	Gene diversity	Heterozygosity	PIC
1	01-2916-1_C_156	0.5383	0.4971	0.0781	0.3735
2	01-608-4_2_375	0.5116	0.4997	0.1435	0.3749
3	01-6351-1_C_202	0.8803	0.2107	0.0534	0.1885
4	02-267	0.6859	0.4309	0.1313	0.3380
5	02-3029-1_C_474	0.6269	0.4678	0.1419	0.3584
6	02-4333-1_2_293	0.7952	0.3257	0.0383	0.2726
7	03-1691-1_C_373	0.7374	0.3873	0.0689	0.3123
8	03-3478-1_C_206	0.5340	0.4977	0.1422	0.3738
9	03-4660-1_2_355	0.7926	0.3287	0.0736	0.2747
10	04-1801-20_2_428	0.8112	0.3063	0.1373	0.2594
11	04-19-4_C_240	0.9968	0.0064	0.0032	0.0064
12	04-3787-3_1_358	0.8793	0.2122	0.0575	0.1897
13	05-2692-1_2_109	0.9040	0.1736	0.0323	0.1585
14	05-4192-1_C_280	0.7090	0.4127	0.0879	0.3275
15	05-48-1_C_279	0.5646	0.4916	0.1417	0.3708
16	06-1256-1_C_147	0.9916	0.0166	0.0076	0.0165
17	06-1776-1_1_501	0.6209	0.4708	0.1404	0.3600
18	06-2509-1_C_497	0.9107	0.1627	0.0376	0.1495
19	07-2904-39_C_299	0.9084	0.1664	0.0275	0.1526
20	07-293-12_1_368	0.8734	0.2211	0.0253	0.1967
21	07-4304_new	0.5190	0.4993	0.0471	0.3746
22	08-2765-2_C_360	0.9164	0.1532	0.0175	0.1415
23	08-4218-5_C_129	0.5180	0.4994	0.1095	0.3747
24	08-847-6_C_113	0.6868	0.4302	0.1146	0.3377
25	09-209	0.8683	0.2288	0.0416	0.2026
26	09-2107-5_C_145	0.6907	0.4272	0.0879	0.3360
27	09-2716-4_C_457	0.9525	0.0906	0.0245	0.0865
28	10-1192-7_C_178	0.6396	0.4610	0.1625	0.3548
29	10-188-1.	0.7936	0.3276	0.1040	0.2740
30	10—2723	0.9245	0.1397	0.0826	0.1299
31	11-1849.	0.8152	0.3013	0.1102	0.2559
32	11-3935.	0.9375	0.1172	0.0324	0.1103
33	11-522-1_C_214.	0.9127	0.1593	0.0379	0.1466
34	12-1794	0.8950	0.1880	0.0396	0.1703
35	12-3200-2_C_389	0.9799	0.0393	0.0048	0.0386
36	12-400	0.8386	0.2707	0.0779	0.2340
	Mean	0.7822	0.2950	0.0740	0.2395

(0.7) where they have used PCR based allele specific markers for generation of genetic diversity.

Hierarchical Cluster Analysis

All 36 SNP loci were scored across all 662 rice germplasm. Genetic distance was calculated and NJ tree was constructed using dissimilarity matrix. Unrootedtree as was generated to find genetic relationships among the rice germplasm. All germplasm were grouped into three major clusters. Cluster 1 having maximum genotypes got grouped into subclusters. Cluster 3 is the smallest comprising of only four genotypes (Fig. 1.). Garris *et al.* (2005) observed clear distinction of wild, *indica*, *japonica* and basmati types showing their independent evolution by genetic relationships. Similarly, Travis *et al.* (2015) studied genetic diversity among 511 cultivars from Bangladesh and North-East India using a 384-SNP microarray assay and identified 191, 229 and 142 SNPs clearly distinguish *indica*, *japonica* and *aus* accessions, respectively, and the *aus* group has been further resolved into two subpopulations *aus*1and *aus*2. However, our results do not correspond to the study conducted by

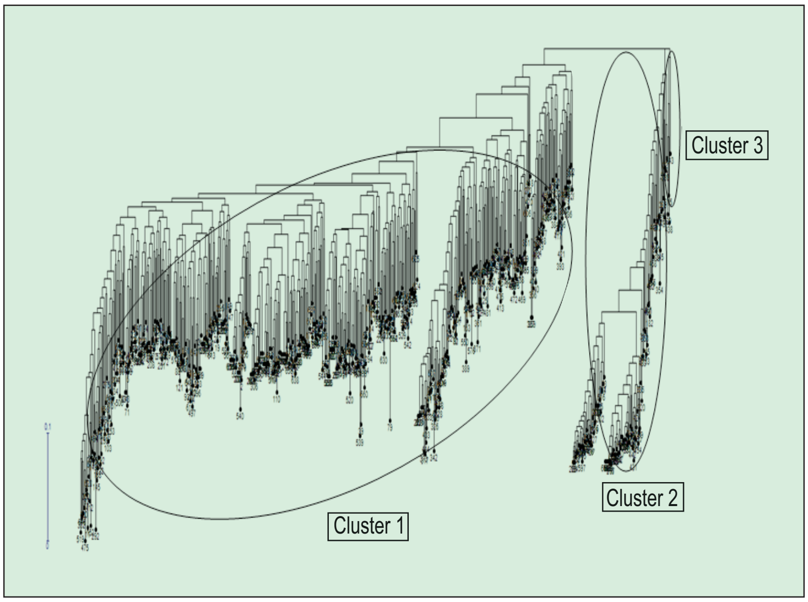

Fig. 1. NJ tree constructed for Arunachal Pradesh rice germplasm based on SNP data

Garris *et al.* (2005) and Travis *et al.* (2015). But our study provides a useful pool of SNP markers with uniform coverage throughout the rice genome. We have characterized 36 SNPs that were genotyped in 662 rice accessions. Our genomic diversity and structure analyses have included old and modern local cultivars, germplasm attempting to uncover a maximum spectrum of variability and occurrence of three different gene pools identified by structure analysis. This allowed us to reconstruct the genetic relationships and genetic diversity of the rice germplasm of Aunachal Pradesh.

Model Based Population Structure

To study the population structure, a model based programme STRUCTURE was used to determine genetic relationship among individual rice germplasm. This model assumes that the number of populations was k and the loci were independent and at Hardy–Weinberg equilibrium. Ln(PD) derived Δk was plotted against the K to determine the number of populations using a software "Structure harvester" available online. At K=3 maximum Δk was found, hence, this was considered as the number of population for the state of Arunachal Pradesh. Population structure divided 662 accessions into 3 populations (Fig. 2). Population1 (pop1), population2 (pop2) and population3 (pop3) contained 378, 203 and 81 germplasm, respectively. Further, based on the membership fractions, rice germplasm under different populations were categorized as pure or admixture. The accessions with the probability more than ≥0.80 score was considered as pure and less than 0.80 as an admixture. Pop1 showed 250 pure (66%) and 128 (34%) admixed individuals, pop2 showed 122 (60%) pure and 81 (40%) admixed individuals and pop3 showed 69 (85%) pure and 12 (15%) admixed individuals (Fig. 3). The relatively small value of alpha (α = 0.12) in present study reveals that, only few individuals were admixed. Alpha value approaching zero indicates that most individuals in the study are from separate populations (Li *et al.,* 2014) whereas; an alpha value greater than 1 indicates that most of accessions of populations are admixed (Ostrowski *et al.*, 2006).

AMOVA and PCoA of Clusters Obtained Using Hierarchical Approach

AMOVA for the 662 accessions was performed based on the three clusters obtained using hierarchical cluster analysis. The three populations showed 0% variance at population level, whereas, 79% variance was recorded among individuals and 21% variance within individuals (Fig. 4; Table 3). PCoA based on hierarchical clusters showed intermixing of the three groups across the coordinates (Fig. 5; Table 4). The first three axes explained 37.8% of cumulative variation.

AMOVA and PCoA of Populations Obtained Using Model Based Approach

AMOVA was performed on three populations obtained using a model based approach. Among three populations 31% variance was recorded, whereas, among individuals,

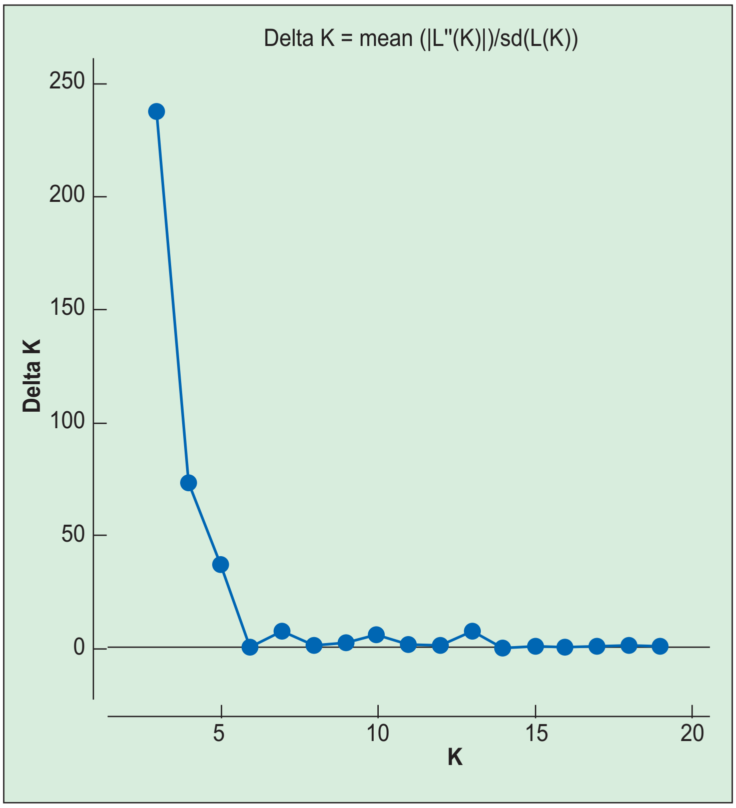

Fig. 2. Estimation of population using LnP(D) derived Δk for k from 1 to20 using SNPs marker data

51% variance and within individuals, 18% variance was found (Fig. 6; Table 5).PCoA revealed that large genetic diversity exists in rice germplasm of Arunachal pradesh. The first three axes explained 37.92% of cumulative variation (Fig. 7; Table 6). In PCoA, rice germplasm were labelled with three different colours which represent the three populations obtained from population structure. The pop1 and pop2 showed intermixing among themselves whereas pop3 occupied a distinct distribution between co-ordinates one and four. It showed very less intermixing with the other two populations. The above study shows that there is a lot of diversity prevailing in Arunachal Pradesh thus enhancing the belief that India has a rich resource of diverse germplasm.

Evaluation of accessions is important to discover useful sources and predict the genetic potential and breeding value of the currently underutilized materials in genebanks. Phenotypic evaluation has been successful in identifying beneficial materials for simple traits in genebanks. When an accession is identified as containing desirable traits, such as biotic or abiotic stress tolerance, the underlying alleles or haplotypes need to be transferred into elite germplasm for trait improvement with the help of molecular evaluation (Wang *et al.*, 2017). Hence, molecular characterization of germplasm can provide the raw material in this regard. Despite the advancement in genomics, the Indian rice collection remained uncharacterized at the molecular level, with respect to parameters such as genetic diversity and

Table 3. Summary of AMOVA table (Hierarchical Cluster based method)

Summary AMOVA Table					
Source	df	SS	MS	Est. Var.	%
Among Pops	2	26.564	13.282	0.013	0%
Among Indiv	659	7183.956	10.901	4.808	79%
Within Indiv	662	850.500	1.285	1.285	21%
Total	1323	8061.020		6.106	100%

Table 4. Percentage of variation explained by the first 3 axes using SNP markers in principal coordinate analysis (Hierarchial Cluster based method)

Axis	1	2	3
%	22.89	8.74	6.25
Cum %	22.89	31.63	37.88

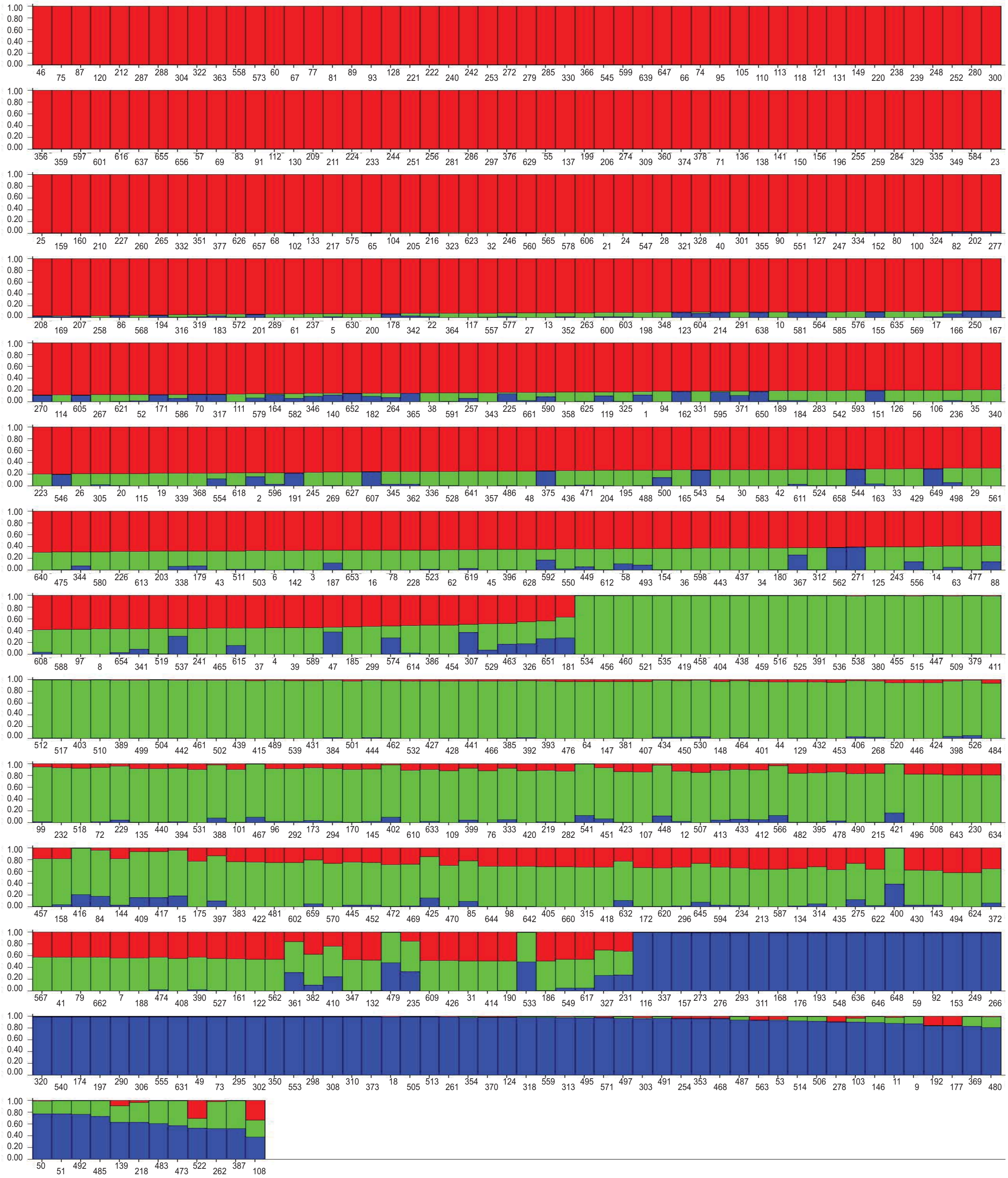

Fig. 3. Model based clustering of Arunachal Pradesh rice germplasm based on SNP (K=3) markers

population structure. This has been the major limiting factor in their utilization and development of improved cultivars.

Acknowledgements

Financial support was granted by Indian Council of Agricultural Research, New Delhi, India, under the NPTC project, is gratefully acknowledged. We are also thankful to Director, ICAR-National Bureau of Plant Genetic Resources, New Delhi, for providing research facilities.

References

Anupam A, J Imam, S Mohammad Quatadah, A Siddaiah, SP Das, M Variar and NP Mandal (2017) Genetic diversity analysis

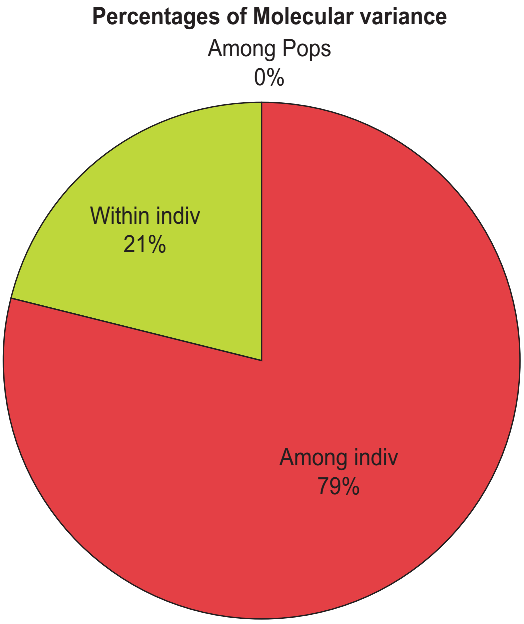

Fig. 4. Analysis of Molecular Variance (AMOVA) of 662 rice germplasm based on hierarchical cluster based method

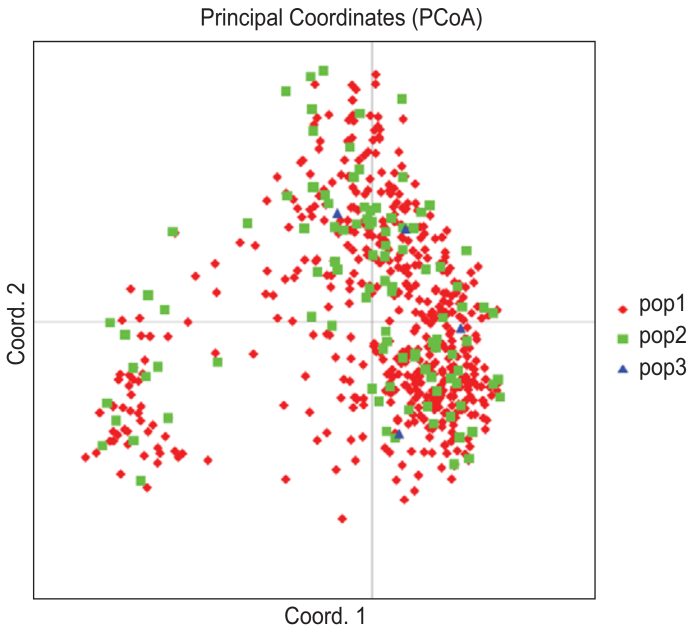

Fig. 5. Principal Coordinate Analysis (PCoA)of 662 rice germplasm based on hierarchical cluster based method

Table 5. Summary of AMOVA table (Model Based approach)

Source	df	SS	MS	Est. Var.	%
Among Pops	2	1667.178	833.589	2.206	31%
Among Indiv	659	5533.448	8.397	3.552	50%
Within Indiv	662	855.500	1.292	1.292	18%
Total	1323	8056.126		7.051	100%

Table 6. Percentage of variation explained by the first 3 axes using SNP markers in Principal Coordinate analysis(Model based approach)

Axis	1	2	3
%	22.91	8.76	6.25
Cum %	22.91	31.67	37.92

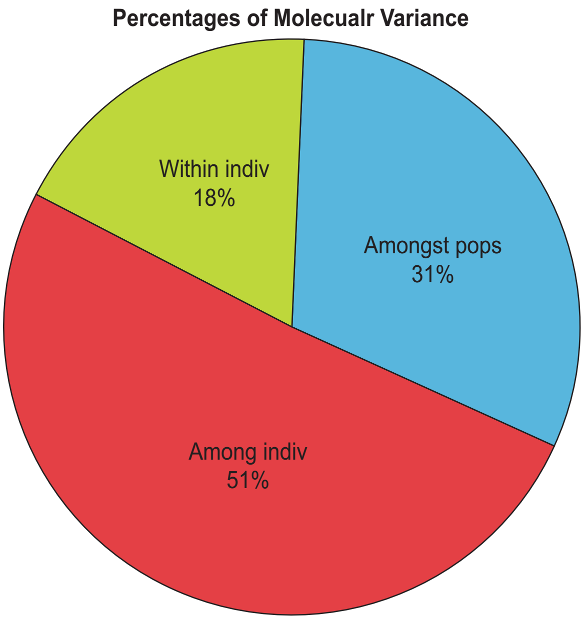

Fig. 6. Analysis of Molecular Variance (AMOVA) of 662 rice germplasm based on population structure based method

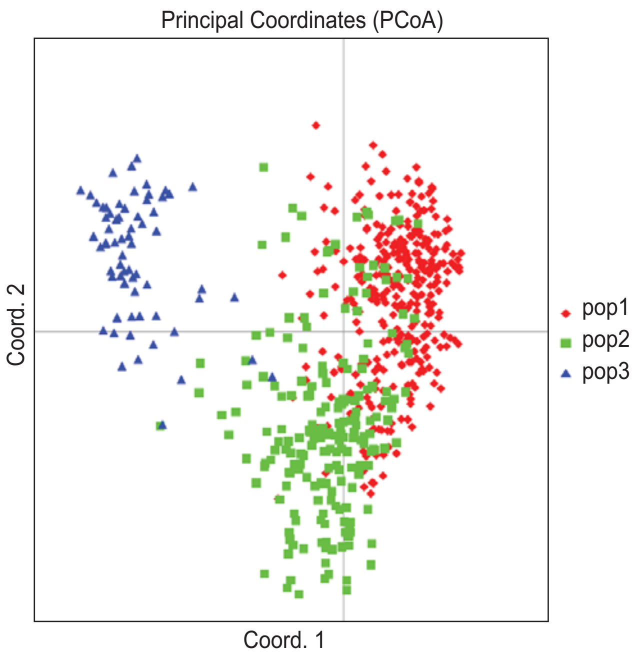

Fig. 7. Principal Coordinate Analysis (PCoA) of 662 rice germplasm based on population structure based method

of rice germplasm in Tripura state of northeast India using drought and blast linked markers. *Rice Sci.* **24**: 10-20.

Barry MB, JL Pham, JL Noyer, B Courtois, Billot and Ahmadi N (2007) Implications for in situ genetic resource conservation from the eco-geographical distribution of rice genetic diversity in Maritime Guinea*. Plant Genet. Resour.* **5**:45–54.

Chang TT (1976) The origin, evolution, cultivation, dissemination, and diversification of Asian and African rice. *Euphytica* **25**:425–441

Choudhury B, ML Khan and S Dayanandan (2013) Genetic structure and diversity of indigenous rice (*Oryza sativa*) varieties in the Eastern Himalayan region of Northeast India. *SpringerPlus* **2**:228. doi:10.1186/2193-1801-2-228.

Earl, DA and BM VonHoldt (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genetic. Resour.* **4**: 359-361. doi: 10.1007/s12686-011-9548-7.

Evano, G, S Regnuat, and J Goudet, (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**: 2611–2620.

Gao LZ, Z ChiHong, C LiPing, J JiZeng, Q ZongEn and YS Dong (2005) Microsatellite diversity within *Oryza sativa* with emphasis on *indica-japonica* divergence. *Genet. Res.* **85**:1–8

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

Government of India, Ministry of statistics and programme implementation (2015). *www.mospi.gov.in*.

Kennedy G and B Burlingame (2003) Analysis of food composition data on rice from a plant genetic resources perspective. *Food Chem*. **80**: 589–596.

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35**:25–34.

Li FP, YS Lee, SW Kwon, G Li and YJ Park (2014) Analysis of genetic diversity and trait correlations among Korean landrace rice (*Oryza sativa* L.). *Genet. Mol. Res*. **13**:6316–31.

Lin HY, YP Wu, AL Hour, SW Ho, FJ Wei, CYL Hsing and YR Lin (2012) Genetic diversity of rice germplasm used in Taiwan breeding programs. *Botanical Studies* **53**: 363-376.

Liu K and SV Muse (2005) Power Marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**: 2128–2129.

Londo JP, YC Chiang, KH Hung, TY Chiang and BA Schaal (2006) Phylo-geography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci. USA* **103**:9578–9583.

Lu H, MA Redus, JR Coburn, JN Rutger and SR McCouch (2005) Population structure and breeding patterns of 145 U.S. rice cultivars based on SSR marker analysis*. Crop Sci.* **45**:66–76.

Nei M, F Tajima, and Y Tateno (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Mol. Evol*. **19**: 153–170.

Ostrowski MF, J David, S Santoni, H McKhann, X Reboud, CV Le, C Camilleri, D Brunel, D Bouchez, B Faure and Bataillon T (2006) Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Mol Ecol.* **15**:1507–17.

Perrier X and JP Jacquemoud-Collet (2006) DARwin software http://darwin.cirad.fr/

Pervaiz ZH, MA Rabbani, SR Pearce and SA Malik (2009) Determination of genetic variability of Asian rice (*Oryza sativa* L.) varieties using microsatellite markers. *African J. of Biotech.* **8:** 5641-5651.

Pritchard JK, M Stephens and P Donnelly (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.

Roy S, BC Marndi, B Mawkhlieng, A Banerjee, RM Yadav, AK Misra and KC Bansal (2016) Genetic diversity and structure in hill rice (*Oryza sativa* L.) landraces from the North-Eastern Himalayas of India. *BMC Genet.* **17**: 107-121.

Singh NK, V Dalal, K Batra, BK Singh, G Chitra, A Singh, IA Ghazi, M Yadav, A Pandit, R Dixit, PK Singh, H Singh, KR Koundal, K Gaikwad, T Mahopatra and TR Sharma (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.

Thomson MJ, NR Polato, J Prasetiyono, KR Trijatmiko, TS Silitonga and SR McCouch (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.

Travis AJ, GJ Norton, S Datta, R Sarma, T Dasgupta, FL Savio, M Macaulay, PE Hedley, KL McNally, MR Islam and AH Price (2015) Assessing the genetic diversity of rice originating from Bangladesh, Assam and West Bengal. *Rice* **8**:35. doi: 10.1186/s12284-015-0068-z.

Upadhyay P, CN Neeraja, C Kole and VK Singh (2012) Population structure and genetic diversity in popular rice varieties of India as evidenced from SSR analysis*. Biochem. Genet*. **50**:770–83.

Wang C, S Hu, C Gardner, T Lubbersted (2017) Emerging Avenues for Utilization of Exotic Germplasm. *Trends Plant Sci.* DOI: http://dx.doi.org/10.1016/j.tplants.2017.04.002

Wing RA, JSS Ammiraju, MZX Luo, H Kim, YS Yu, D Kudrna, JL Goicoechea, W Wang, W Nelson, K Rao, D Brar, DJ Mackill, B Han, C Soderlund, L Stein, P SanMiguel and S Jackson (2005) The Oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol. Biol*. **59**:53–62.

Zhang H, J Sun, M Wang, D Liao, Y Zeng, S Shen, P Yu, P Mu, S Wang and Z Li (2007) Genetic structure and phylogeography of rice landraces in Yunnan, China, revealed by SSR. *Genome* **50**:72–83.

Zhang QF, JY Li, YB Xue, B Han and, XW Deng (2008) Rice 2020: a call for an international coordinated effort in rice functional genomics. *Mol Plant*. **1**:715–719.

Zhang P, L Jinquan, L Xiaoling, L Xiangdong, Z Xingjuan and L Yonggen (2011) Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR Markers, *PLoS ONE*. **6**: 1-13.