Environment and Ecology 41 (3C) : 1883—1890, July—September 2023
Article DOI: https://doi.org/10.60151/envec/JWHX9999
ISSN 0970-0420

Molecular Diversity and Population Structure Analysis in Rice Genotype using SSR Markers

Atul Kumar Pachauri, Deepak Gauraha, Sant Ram Sahu, Praveen Pujari, Deepak Saran

Received 9 September 2022, Accepted 20 June 2023, Published on 4 September 2023

ABSTRACT

The use of advanced molecular technologies is one possible approach to understand their diversity. A study was undertaken to evaluate forty two rice genotype including six checks during *kharif* 2019. The result of the investigation was allele number was recorded for the five markers with an average of 2.6 allele per locus. Out of twenty five markers 12 markers showed highly PIC value. However, UPGMA cluster diagram mean of 67% level of similarity showed forty two accessions into 10 distinct clusters. Maximum cluster II had 13 genotypes followed by cluster IV, X, III and IX consisting of eleven, six three and two genotype respectively. Population structure analysis population inferred ancestry, 10 pure accessions were assigned to subgroup SG1 whereas 30 pure accessions were assigned to subgroup SG2 and two lines were assigned to admixture (AD). It was observed that the genotypes of subgroup 1 have brown planthopper resistant, whereas the genotypes of subgroup 2 have high yielding.

Keywords Genetic diversity, Rice genotypes, SSR marker, Marker trait association.

Atul Kumar Pachauri[1]*, Deepak Gauraha[2], Sant Ram Sahu[3], Praveen Pujari[4], Deepak Saran[5]

[1]Senior Research Fellow, [2] Scientist
Genetics and Plant Breeding
[3]Senior Research Fellow
Plant Pathology

[1,2,3]Indira Gandhi Krishi Vishwavidyalaya, Raipur 492012, Chhattisgarh, India

[4]FA (AICRP-Rice), [5]PhD Scholar
Genetics and Plant Breeding

[5,6]Indira Gandhi Krishi Vishwavidyalaya, Raipur 492012, Chhattisgarh, India

Email: pachauriatul@yahoo.in
*Corresponding author

INTRODUCTION

Rice an vital food for about half of the world's population and 80% of it is being produced and consumed in Asia and share maximum in grain production. India is one of the centers for rice diversity (Allhgolipour *et al.* 2014). The rice accessions are a rich reservoir of valuable genes that rice breeder can harness for rice improvement program and the genetic variability exists among rice accessions leaving a wide scope for crop improvements (Singh *et al.* 2016). Genetic diversity is necessary for any crop improvement program as it helps in analyzing and establishing genetic relationship in accessions collection, its monitoring, identification of diverse parental combinations to create segregating progenies with high genetic variability and to obtain potential recombinations for further selection and introgression of desirable genes from these diverse accessions (Thomson *et al.* 2020 and Karimah *et al.* 2021). Since been a long time a major

goal in evolutionary biology is to characterize and quantify the genetic diversity. Determination of genetic diversity can be done by assessing morphological or molecular data. The use of advanced molecular technologies is one possible approach to understand their diversity. Evaluation of genetic diversity using DNA marker technology is non-destructive, not affected by environmental factors, requires small number of samples, and does not require large experimental setup and equipments for measuring physiological parameters (Akter *et al.* 2022).

Simple sequence repeat (SSR) is an important tool for genetic variation identification of accessions. SSR marker are highly informative, mostly monolocus, codominant, easily analyzed and cost effective and able to detect high level of allelic diversity, thus being widely applied in genetic diversity analysis, molecular map construction and gene mapping (Ma *et al.* 2011) and analysis of germplasm diversity (Jin *et al.* 2010). SSR markers even in less number can give a better genetic diversity spectrum due to their multi-allelic and highly polymorphic nature (Singh *et al.* 2016). Reports suggest that genetic diversity in crop varieties released over the years fluctuates in successive time periods. Over the last few centuries, rice has faced diversity loss especially, after the green revolution due to replacement of native varieties with high yielding varieties.

Therefore the present study was undertaken with the aim to assess the trend in genetic diversity in twenty five accessions of rice using SSR markers (Nachimuthu *et al*. 2015). The generated information will enable maximized selection of diverse parents and selecting appropriate parental genotypes in breeding program.

MATERIALS AND METHODS

The experiment was carried out during *kharif* season 2019 at Instructional Farm, Indira Gandhi Agricultural University Raipur, Chhattisgarh. The experimental seed material comprises of 42 diverse rice germplasm accessions along with 6 checks (Table 1). The 21 days old seedling was transplanted in a Randomized Block Design (RBD) with three replications by maintaining row to row and plant to plant spacing 20 × 15 cm, respectively. Rice germplasm accessions list as follow.

Morphological assessment

Observations were recorded on eleven quantitative traits viz., days to 50% flowering (DF), number of effective tillers per plant (NET), plant height (PH), panicle length (PL), 100 grain weight (GW) and grain yield per plant (GYP).

Genomic DNA extraction

Young leaves of 15-20 days old seedlings from sixty five rice genotypes were clipped and stored in ice-box to carry it to the lab which is then stored in -800C till DNA extraction. Genomic DNA was then extracted using CTAB method. DNA samples were diluted to 10 ng/ µl. The DNA was quantified spectro photometrically (Perkin Elmer, Singapore) by measuring A260/A280 and DNA quality was checked by electrophoresis in 0.8% agarose gel.

Table 1. List of rice accessions used under study.

Sl. No.	Accession No.	Sl. No.	Accession No.	Sl. No.	Accession No.	Sl. No.	Accession No.
1	IC453695	11	IC459797	21	IC124743	31	IC207184
2	IC454040	12	IC460308	22	IC124763	32	IC207311
3	IC454377	13	IC114326	23	IC124989	33	IC214133
4	IC454223X	14	IC114796	24	IC126472	34	IC124028
5	IC460174X	15	IC461616	25	IC126596	35	IC301736
6	IC459148	16	IC124536	26	IC301732	36	IC386429
7	IC459199	17	IC267449	27	IC301734	CH 1	Annada
8	IC125139	18	IC300226	28	IC206699	CH 2	Pusa Basmati1
9	IC86313	19	IC124546	29	IC135425	CH 3	Jaya
10	IR79538	20	IC124735	30	IC206606	CH 4	NDR 97
CH5	IR 64	CH6	Swarna				

SSR markers and PCR amplification

A total of twenty rice SSR markers were used for molecular diversity (Table 2). The PCR amplification was carried out in 15µl of reaction mixture containing 30 ng genomic DNA, 1.5 mM PCR buffer (MBI Fermentas, USA), 400 µMdNTPs (MBI Fermentas), 1 U Taq DNA polymerase (MBI Fermentas) and 0.4 µM primer using a thermal cycler (Master cycler gradient, Eppendorf). Thermal cycling program involved an initial denaturation at 94° for 45 sec, annealing at 2° bellow Tm of respective primers for 30 sec, primer extension at 72° for 30 sec, followed by a final extension at 72° for 7 min. Electrophoresis separation and visualization of amplified products. The amplified PCR products along with a 50 bp DNA marker ladder (MBI Fermentas) were size fractioned by electrophoresis in 2.5% agarose gel prepared in TAE buffer and visualized by staining with ethidium bromide (0.5 µg/ml) in a gel documentation system (BIO-RAD, USA). The reproducibility of amplification products was compared twice for each primer.

Statistical analysis

The morphological data were managed using Excel 2007 and the descriptive statistics and variance analysis were produced using SPSS 16.0 (IBM SPSS). Broad sense heritability was calculated by using OPSTAT. Informativeness of molecular markers and their diversity parameters viz., average number of alleles per locus, gene diversity and the Polymorphic information content (PIC) values were calculated using Power Marker v3.25.

The population structure analysis was performed in STRUCTURE v2.3 using a burn-in period of 100,000 and MCMC replicates of 100,000. The number of groups (K) was set from 1 to 10, with five iterations. Probable structure number (K) was estimated by Evanno's method at Structure Harvester software (Evanno *et al.* 2005). The rice lines with membership probability of 0.80 or above were considered in a specific group whereas genotypes with membership probability of Q<0.80 were considered as Admixture type (AD). TASSEL v4.3 was also used to calculate Jaccard's similarity-based Neighbour-Joining tree. XLSTAT was used for PCoA analysis based on the molecular markers.

The association mapping analysis was performed by TASSEL v2.3 using the mixed linear model (MLM). The model MLM_Q+K, using kinship matrix and Q-matrix as the concomitant variable, was used to identify the marker-trait association. Marker trait association at p<0.01 were considered significant.

Genetic dissimilarity and cluster analysis based on UPGMA

The binary data matrix generated by polymorphic SSR markers were subjected to further analysis using NTSYS-pc version 2.11W. The SIMQUAL program was used to calculate the Jackard' dissimilarity coefficient. The dissimilarity matrix was used as an input for analysis of clusters. UPGMA-based clustering was done using SAHN module of NTSYSpc for dendrogram construction. In unweighted pair-group average (UPGMA) clusters are joined based on the average distance between all members in the two groups.

Polymorphic information content (PIC) and principal component analysis (PCA)

PIC for SSR markers was calculated as per the formula:

$$PIC = 1 - Pij^2$$

Where, PICi is the polymorphic information content of a marker i and the summation extends over n patterns. PCA was also done to check the result of UPGMA base clustering using EIGEN module of NTSYSpc. In principal component analysis (PCA), the total variance of original characters is divided into a limited number of uncorrelated new variables known as principal components (PCs). The first step in PCA is to calculate eigen values, which define the amount of total variation that is displayed on PC axes. The first PC summarizes most of the variability not summarized by, and uncorrelated with, the first PC, and so on. PCs were used for 2-dimentional (2-D) and 3-dimentional (3- D) plotting, respectively, against each other using module PROJ and MXPLOT of NTSYSp.

Table 2. List of twenty five microsatellite markers with their chromosome locations, number of alleles, allele size and PIC value found among forty two rice accessions.

Sl. No.	Marker	Chr	Position cM	Amplicon size (bp)	Number of alleles	PIC value
1	RM237	1	115.2	105-153	4	0.699
2	RM171	10	73	307-347	2	0.375
3	RM283	1	31.4	130-176	4	0.693
4	RM431	1	178.3	233-267	3	0.592
5	RM 44	8	60.9	82-132	2	0.374
6	RM484	10	97.3	286-298	2	0.375
7	RM307	4	0	116-9111	4	0.694
8	RM124	4	150.1	257-289	4	0.702
9	RM447	8	124.6	95-146	2	0.374
10	RM433	8	116	216-248	2	0.375
11	OSR13	3	53.1	85-122	4	0.699
12	RM105	9	32.1	100-141	2	0.375
13	RM316	9	1.8	194-216	2	0.375
14	RM287	11	68.6	82-118	3	0.375
15	RM536	11	55.1	223-247	3	0.59
16	RM489	3	29.2	223-289	2	0.592
17	RM11	7	47	118-151	2	0.375
18	RM161	5	96.9	154-187	2	0.374
19	RM413	5	26.7	71-114	2	0.375
20	RM271	10	59.4	80-120	3	0.590
21	RM507	5	0	234-257	3	0.702
22	RM338	3	108.4	178-184	1	0
23	RM55	3	168.2	216-247	3	0.592
24	RM510	6	20.8	99-127	2	0.373
25	RM495	1	2.8	148	3	0.592

RESULTS AND DISCUSSION

Microsatellite marker informativeness

A total 25 microsatellite markers was used for measuring allelic diversity of 42 diverse penal. The marker diversity paraquotes such as allele per locus major allele for polymorphic information content value. Each marker were scored on the base of expected product size as available on graminea database with most intensely amplified band using standard weight size marker (50bp ladder).The total 66 allele observed for all the marker of which 65 by highly variable SSR markers (Table 2). The maximum allele number of was recorded for the marker RM237, RM283, RM307, RM124 and OSR13 with an average of 2.6 allele per locus (Fig.1). Result indicated that allelic diversity found in the genotype studies. The range of PIC value varied from 0.37 (RM161) to 0.70 (RM124) with the mean of 0.50. Similar result was supported by Nagy *et al.* (2012), Pritchard *et al.* (2016).

Dandogram analysis

A dendrogram based on Jackard's dissimilarity coeffi-

Fig. 1. Polymorphic information content value.

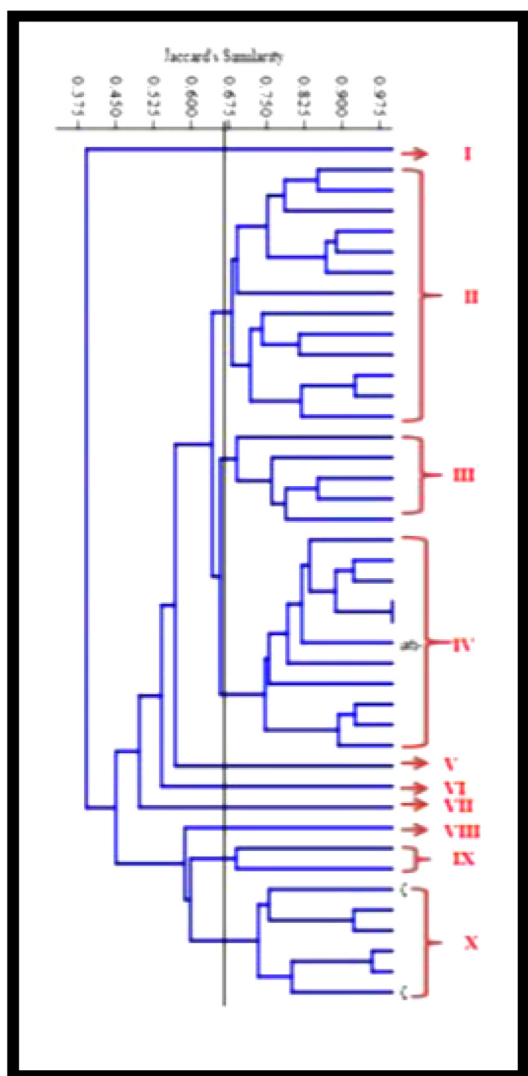

Fig. 2. Clustering 42 genotypes based on the UPGMA methods with Jaccard's coefficient,

cient was constructed using UPGMA (Jaccard 1908). The genetic divergence was studied based on D^2 statistics (Fig. 2). The forty two rice accessions were grouped into ten clusters. In the dendrogram cluster II maximum number of 13 genotypes followed by cluster IV, X, III and IX consisting of eleven, six three and two genotype respectively. Whereas the cluster I, V, VI, VII and VIII had single genotype in each (Fig. 2). In this context parallel result was announced by Yadav *et al*. (2013), Akter *et al.* (2022).

Table 3. Grouping of forty two rice accession into diverse clusters.

Cluster	Name of genotypes	Number of genotypes
I	IC386986	1
II	IC206606, IC125139, IC454040, IC214133, IC267449, IC124028, IC389453, IC860313, IC114326, Annada, IC124536, TN1 and IC 124546	14
III	IC 459199, IC114796, IC126596, IC207184 and IC 207311	5
IV	IR64, IC388153, IC124763, IC459797, IC 461616, Pusa Basmati1, Jaya, Swarna, IC 124989, and IC126472	11
V	IC300226	1
VI	PTB33	1
VII	IC206699	1
VIII	IR79538	1
IX	IC460308 and IC459148	2
X	IC460174X, IC301732, IC301736, IC301734,IC454377and IC454223X	6

Cluster II consisted 13 genotypes viz., C206606, IC125139, IC454040, IC214133, IC267449, IC124028, IC89453, IC860313, IC114326, Annada, IC124536, TN1 and IC124546 cluster I,V,VI,VII and VIII had a single genotype IC386986, IC300226,PTB33, IC206699 and IC459220 respectively (Fig.2) and (Table 3). Cluster I found the high dissimilarities at the level (37%) of genotype IC386986 (Fig. 2).

Cluster IV consisted of 11 genotype IR64, IC388153, IC124753, IC459797, IC461616, Pusa Basmati-1,Jaya, Swarna, IC124989, NDR 97 and IC126472 (Table 3), while genotype IC459797

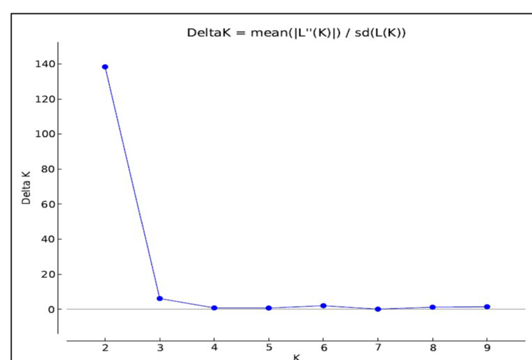

Fig. 3. This indicated that 42 rice germplasm had a genetic structure of two sub populations as shown in bar plot.

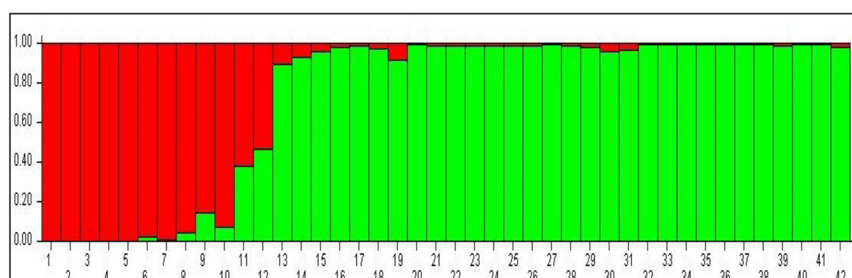

Fig. 4. Bar plot showing the population structure of 42 rice landraces based on SSR markers at K=2.

shown the highest genetic similarities with IC461616 at the level of 98%. Although genotypes IC46174X, IC301732, IC301736, IC454377 and IC454223X were present under cluster X (Table 3). Cluster III had 5 genotypes IC 459198, IC114796, IC126596, IC207184 and IC 207311 (Table 3). The result was supported by Hossain *et al.* (2012), Singh *et al*. (2016), Sahu *et al.* (2017).

Population structure analysis

Population structure of the 42 rice germplasm lines based on 25 molecular markers was analyzed by Bayesian model based approach. By comparing LnP (D) and Evanno's ΔK values by increasing K from 1 to 7. The log likelihood revealed by structure showed the optimum value as 2 (K=2). Similarly the maximum of adhoc measure ΔK was found to be K=2 (Fig. 4), which indicated that the entire population can be grouped into two subgroups (SG1 and SG2). Based on the membership fractions, the accessions with the probability of≥80 % were assigned to corresponding subgroups with others categorized as admixture (Fig. 4).

Estimation of K and population structure. Changes in ΔK value with the number of subpopulations than 0.80 score as an admixture. With population inferred ancestry, 10 pure accessions were assigned to subgroup SG1 whereas 30pure accessions were assigned to subgroup SG2 and two lines were assigned to admixture (AD). It was observed that the genotypes of subgroup 1 have brown planthopper resistant, whereas the genotypes of subgroup 2 (Fig. 4) have high yielding. By measuring molecular markers easy to know evolutionary interaction between genetic diversity in germplasm collection. The result was supported by Babu *et al.* (2014), Courtois *et al.* (2012) and Thomson *et al.* 2020).

A total of 42 diverse set of rice germplasm accessions were used for genome wide scanning through 25 molecular markers. The mapping association was performed using phenotypic and genotypic data as source to analysis marker trait association using statistical model mixed linear model (MLM/Q+K model). Mixed linier model was estimated by taking into account of both population inferred ancestry Kinship matrix (K) using software TASSEL v2.1.The association obtained through MLM analysis were more effective and can be exploited further in smart breeding program. Hence, in the present study more emphasis was given to MLM result (Table 4) and their detailed descriptions. RM307 seven were identified as explaining 20% or more of the total variation (R^2) for grain yield (RM261, RM228), kernel length (RM284), kernel length-width ratio (RM7, RM228), and total kernel weight (RM440, RM122) as shown in Table 4.

Marker RM237 situated at chromosome 1 is significantly associated with BPH resistant (P value 0.043) along with 65.2% phenotypic variation explained. Similarly RM 307, RM271 and RM171 were also associated with BPH resistance with 76.3% 20% and 16% phenotypic variation explained respectively as presented in Table 4. Two markers viz., RM44 , RM507 and RM447 also significantly associated with four traits namely "days to 50% flowering, plant height, panicle length and net effective tillers (Table

Table 4. Significant marker-trait associations based on MLM model.

Sl. No.	Trait	Marker name	Chromo-some No.	Marker P value	Marker R^2 value
1	BPH	RM 237	1	0.043	0.652
2	BPH	RM 307	4	1.32E-04	0.7630
3	BPH	RM 271	10	0.0026	0.2008
4	BPH	RM171	10	0.049	0.1602
5	DTF 50%	RM44	8	1.64E-04	0.3008
6	DTF 50%	RM447	8	0.0359	0.1236
7	DTF 50%	RM507	5	0.0464	0.2416
8	DTF 50%	RM510	6	0.0324	0.999
9	PH	RM44	8	0.0086	0.1566
10	PH	RM447	8	0.0226	0.1203
11	PH	RM287	11	0.0250	0.2131
12	NET	RM484	10	0.0164	0.1403
13	NET	RM 307	4	0.0227	0.1643
14	NET	RM507	5	0.0031	0.2739
15	NET	RM 55	3	0.0357	0.1753
16	PL	RM44	8	0.0032	0.774
17	PL	RM316	9	0.0459	0.117
18	PL	RM 55	3	0.00243	0.2145
19	GYP	RM283	1	0.0301	0.2771
20	GYP	RM431	1	0.0546	0.2997
21	GYP	RM484	10	0.0415	0.0863
22	GYP	RM447	8	0.059	0.1708
23	GYP	OSR13	3	9.37E-04	0.3541
24	GYP	RM287	11	0.0156	0.3255

BPH :Brown planthopper, DTF 50% : Days to 50% flowering, PH: Plant height, NET : Net effective tiller, PL: Panicle length, GYP: Grain yield per plant.

4). The range of phenotypic variation explained by RM44 is 77.4% (Panicle length) to 30% (Days to 50% flowering) whereas RM507 was significantly associated with net effective tillers and plant height by way of Table 4. Moreover RM447 also associated with grain yield per plant and days to 50% flowering and plant height. Similar finding was supported by Devi *et al.* (2016), Huang *et al.* (2014b).

CONCLUSION

SSR markers used in the study were showing average PIC value 0.49 which indicated their effectiveness in polymorphism testing and diversity analysis. These markers can be effectively used in molecular breeding programs and QTL mapping studies since they exhibited very high level of polymorphism over other loci. Total 24 significant marker-trait associations were detected for four different traits. The results indicated that genome-wide association mapping may help to better understand the architecture of these traits and to discover novel trait related genes.

ACKNOWLEDGMENT

Authors articulate honest gratitude Indira Gandhi Krishi Vishwavidyalaya, Department of PMBB for provided that the pecuniary support to my PhD research work.

REFERENCES

Akter MB, Mosab BA, Kamruzzaman M, Reflinur R, Nahar N, Rana MS, Hoque MI, Islam MS (2022) Morpho-molecular diversity study of rice Cultivars in Bangladesh. *Czech J Genet Pl Breed* 58: 64−72.

Allhgolipour M, Farshdfar E, Rabiei B (2014) Molecular characterization and genetic diversity analysis of different rice cultivars by microsatellite markers. *Genetika* 46(1): 187–198.

Babu BK, Meena V, Agarwal V, Agrawal PK (2014) Population structure and genetic diversity analysis of Indian and exotic rice (*Oryza sativa* L.) accessions using SSR markers. *Mol Biol Rep* 41(7): 4329–4339. DOI: 10.1007/s11033-014-3304-5

Courtois B, Frouin J, Greco R, Bruschi G, Droc G, Hamelin C, Ruiz M, Clément G, Evrard JC, Coppenole S (2012) Genetic diversity and population structure in a European collection of rice. *Crop Science* 52:1663–1675.

Devi KR, Parimala K, Venkanna V, Lingaiah N, Hari Y, Satish Chandra B (2016) Estimation of variability for grain yield and quality traits in rice (*Oryza sativa* L.). *Inter J Pure Appl Biol* 4(2): 250-255.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: A simulation study. *Molecular Ecology* 14:2611–2620.

Hossain MM, Islam MM, Hossain H, Ali MS, Jamie A, Komamine A, Prodhan SH (2012) Genetic diversity analysis of rice landraces (*Oryza sativa* L.) by microsatellite markers. *Genes, Genomes Genomics* (6) : 42-47.

Huang X, Han B (2014b) Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol* 65: 531–551.

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44:32–39.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bull Soc Vaud Sci Nat* 44: 223–270.

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

Karimah AZ, Siswoyo TA, Kim KM (2021) Genetic diversity of rice germplasm (*Oryza sativa* L.) of Java Island, Indonesia. *J Crop Sci Biotech* 24: 93–101.

Ma H, Yin Y, Guo ZF, Cheng LJ, Zhang L, Zhong M, Shao GJ (2011) Establishment of DNA fingerprinting of Liaojing series of japonica rice. Middle-East *J Sci Res* (8): 384-392.

Nachimuthu VV, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian BA, Ponniah G, Gunasekaran K, Swaminathan M, Suji KK, Sabariappan R (2015) Analysis of population structure and genetic diversity in rice germplasm using SSR markers: An initiative towards association mapping of agronomic traits in *Oryza sativa. Rice* 8: 30.

Nagy S, Poczai P, Cernak I, Gorji AM, Hegedus G. Taller J (2012) PIC calc: An Online Program to Calculate Polymorphic Information Content for Molecular Genetic Studies. *Biochemical Genetics* 50: 670–672.

Pritchard JK, Stephens M, Donnelly P (2016) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Sahu PK, Mondal S, Sharma D, Vishwakarma G, Kumar V, Das BK (2017) InDel marker based genetic differentiation and genetic diversity in traditional rice (*Oryza sativa* L.) landraces of Chhattisgarh, India. *PLoS ONE* 12 (11): 0188864

Singh N, Choudhury DR, Tiwari G, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Sharma AD, Singh NK, Singh R (2016) Genetic diversity trend in Indian rice varieties: An analysis using SSR markers. *BMC Genet* 17: 127.

Thomson MJ, Septiningsih EM, Suwardjo F, Santoso TJ, Silitonga TS, McCouch, SR (2020) Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. *Theor Appl* Genet. 114: 559–568.

Yadav S, Singh A, Singh MR, Goel N, Vinod KK, Mohapatra T, Singh AK (2013) Assessment of genetic diversity in Indian rice germplasm (*Oryza sativa* L.): use of random versus trait-linked microsatellite markers. *J Genet* 92(3):545–557.