Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genetic Diversity and Population Structure of Kerala Rice Landraces Using Genotyping-By-Sequencing

Maya Peringottillam
Central University of Kerala
Smitha Kunhiraman Vasumathy
Central University of Kerala
Hari Krishna Kumar
Central University of Kerala
Manickavelu Alagu (amanicks@cukerala.ac.in)
Central University of Kerala https://orcid.org/0000-0003-2875-2290

Research Article

Keywords: Rice landraces, genetic diversity, genotyping-by-sequencing, population structure.

Posted Date: June 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-516783/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Researchers stand at the vanguard of advancement and application of next-generation sequencing technology for creating opportunities to guide more realistic and applicable strategies for the sustainable management of genetically diverse rice resources. This study is a pioneering effort where GBS-SNP markers were employed to assess the tremendous genetic diversity and structure of rice landrace collections from northern Kerala. Kerala holds an immense diversity of rice landraces that encountered selection pressures of environmental heterogeneity, biotic and abiotic stresses, however competent rather provide good yields, whereby drawing the attention of the rice breeding sector. The population structure and diversity analyses separated the accessions into three distinct subpopulations with a huge amount of genetic variation within subpopulations. Nei's genetic distance analysis confirmed the existence of strong genetic differentiation among rice landrace populations. The values of F_{ST} and Nm established the farmers' effort to preserve the genetic purity of rice landraces despite the extensive seed exchange programs across the states of India. Moreover, this low level of gene flow among subpopulations could provide the opportunity for well-adapted combinations of genes to be established by natural selection. The clustering pattern based on SNP markers furnished sufficient knowledge in identifying rice genotypes that eliminates the likelihood of duplication among indigenous cultivars. Similar clustering patterns of genotypes revealed shared genetic characters among them. Collectively these analyses can be used to completely understand the population of rice landraces in Kerala while contributing insights toward the evolution and selective pressures underlying these unique landraces.

Introduction

Rice (*Oryza sativa* L.) is one of the most widely cultivated crops globally, covers an area of 162 million hectares and produces nearly 755 million tonnes in 2019 (http://www.fao.org/faostat). Considering about 13,500 political units, the present global average rate of annual yield increase of rice is 1%. However, doubling crop production by 2050 requires an average yield increase of 2.4% per annum, indicating the inadequacy of current yield trends (Ray et al. 2013). The present incompetence can only be vanquished by producing high-yielding rice varieties with minimum land and water resource utilization and resistance to biotic and abiotic stresses. The underpinning of the constant improvement of rice cultivars is the abundant phenotypic and genetic diversity that resides within domesticated populations and wild relatives. The classical approach of intentional interbreeding of closely related individuals to create new cultivars with desirable traits require great effort and long duration to produce several generations, which serves as a barrier to meet the increasing requirements of food needs and maintaining global food security (He et al. 2014; Liu et al. 2020). Therefore, an application of molecular plant breeding known as marker-assisted breeding, capable of discovering phenotypic variation by deploying markers linked to QTLs responsible for desirable traits, became popular. The mining of amount and distribution of genetic diversity from domesticated rice accessions preceding molecular plant breeding methods is of great value for guaranteeing a sufficient, safe, interminable supply of nutritious food to the world population.

Rice landraces are domesticated populations of rice that have been cultivated since ancient times along with the traditional farming practices that retain high genetic diversity and distinctive identity through local adaptation. It is reported that Kerala state had an estimated 2000 rice landraces (Devi et al. 2017). Though Kerala had such a significant number of genetic resources, reported research works are less comparatively. The area of rice research in Kerala is actively involved in characterization (Sreejayan et al. 2003; Joseph et al. 2007; Aiswariya and Thomas 2016; Manjunatha et al. 2018), simple documentation (Latha et al. 2013; Karunakaran 2014) and in the discussion of conservation of traditional rice varieties (Devi et al. 2017; Gopi and Manjula 2018; Blakeney et al. 2020). Population genetic research has been limited to genetic variability studies using molecular markers such as SSR (Vanaja et al. 2007; Vanaja et al. 2010; Thomas and Dominic 2016), RAPD (Thomas et al. 2001; Raj et al. 2010; Kumar et al. 2010; Skaria et al. 2011; Rekha et al. 2011; Rajani et al. 2013) and AFLP (Sreejayan et al. 2011). Collectively in all these studies, the existence of genetic differentiation between the cultivars of *Oryza sativa* and their relationships with geographical distance has been discussed. However, the genetic diversity contained within our germplasm holds significant potential for both Kerala and beyond. Therefore considerable work is required to characterize these samples completely.

Genotyping by sequencing (GBS) is a fast, powerful and cost-effective platform of next-generation sequencing (NGS) technology that uses enzyme-based complexity reduction to discover polymorphisms and genotypic information across the population of interest (Poland and Rife 2012). This method has been successfully used for assessing genetic diversity and population structure in several crops like large and complex genome of wheat (Manickavelu et al. 2014; Alipour et al. 2017; Eltaher et al. 2018; Yang et al. 2020), rice (Mgonja et al. 2017), maize (Leng et al. 2019; Yu et al. 2021), rye (Schreiber et al. 2019), pearl millet (Serba et al. 2019), finger millet (Kumar et al. 2016) etc., which demonstrated the power of GBS-SNP genotyping as an appropriate technology for high throughput genotyping in cereal crops. GBS is valued as a rapid and cost-effective high throughput genotyping to depict genomic diversity and population structure in non-referenced neglected orphan species (Peterson et al. 2014; Hu et al. 2015; D'Agostino et al. 2018). In addition to genetic diversity analyses, we have demonstrated the functional annotation of unmapped sequences and haplotype analysis of desirable genes as an application of this NGS technology (Vasumathy et al. 2020). GBS has been used to construct high-density linkage maps to identify the QTL for rice grain yield under various drought conditions (Yadav et al. 2019). It is also frequently used as an excellent genotyping approach for genotyping diverse set for genome-wide association studies (GWAS) (Nimmakayala et al. 2014; Begum et al. 2015; Manickavelu et al. 2016; Crowell et al. 2016; Descalsota et al. 2018; Girma et al. 2019; Hoang et al. 2019). GWAS has been used as a powerful strategy for understanding the genetic background underlying complex traits. Undetected or cryptic population structure in populations is a significant problem that causes false positives in association analyses (Pritchard et al. 2000). Therefore, it is imperious to appreciate the genetic structure, not only to acquire historical insights but also to apply a proper strategy of association studies to escape from the population stratification on GWAS.

This study investigates genetic diversity and population structure of rice landrace populations collected from the north Malabar region of Kerala, employing SNPs discovered from genotyping by sequencing.

Detailed analysis on population structure and genetic diversity using the GBS-SNP platform in rice landraces of Kerala is still lacking. Towards our understanding, the present work is the first and foremost reported study of its kind in rice landraces of Kerala.

Materials And Methods

Plant Material and DNA Extraction

A diversity panel of 96 *Oryza sativa* accessions was collected from different regions of northern Kerala (Supplementary Material 1). Included among them, 78 are rice landraces from Kerala, 15 are from other states and three elite varieties. Rice landraces from other states of India are collected by farmers through seed exchange programs of indigenous rice cultivars and have been cultivating in Kerala for a long time. Collections of germplasm were cultivated at a field in Kerala (12.186687°N latitude and 75.222666°E longitude). Approximately 50 mg of leaf tissue from seedlings for each accession were collected after 15 days of emergence. Total genomic DNA from leaf tissues of each sample was isolated using CTAB and Phenol: Chloroform DNA extraction method followed by RNase A treatment and purification (Doyle and Doyle 1990). The quantity of the isolated DNA was estimated using a Nanodrop Spectrophotometer. Furthermore, purity was substantiated by running the DNA in 0.8% agarose gel electrophoresis.

GBS library preparation and sequencing

GBS libraries were prepared using the protocol adapted from Poland and Rife (2012). For GBS library preparation, 10-20 µg of genomic DNA was digested using restriction enzymes *ApK* 1 and *Pst* 1, followed by ligation of barcoded adapters to the sticky ends of digested DNA. Adaptor ligated products were then pooled and PCR enriched with specific primers enabling them to hybridize Illumina flow cell primers and priming subsequent DNA sequencing reactions. The QC of prepared libraries was then checked with Agilent Tape Station. These libraries were sequenced on Illumina NextSeq 500 platform with 2x 150 bp v2 chemistry. The adaptors and barcodes of the sequenced reads were trimmed using trimmomatic v 0.36 and checked the sequence quality using FASTQC using Phred quality score >= 30. These high-quality tag sequences were aligned to the reference genome *Oryza sativa* ssp*. indica* (BioProject: PRJNA361) using Burrows-Wheeler Alignment (BWA) tool (Li and Durbin 2009). The SAM files were piled up and converted to BAM files using SAM tools. Individual Single Nucleotide Polymorphisms (SNPs) were detected using SAM tools with the following parameter: ‘mpileup -m 2 -F 0.002 -d 2000’. For reducing the error rate in SNP detection, results were filtered with the following criteria: (1) The number of support reads for each SNP should be more than 4. (2) The mapping quality (MQ) of each SNP should be higher than 20. (3) The SNPs with minor allele frequency (MAF) MAF > 0.05 were retained. (4) SNPs having more than 20% missing information were removed using vcftools.

Genetic Properties of Markers

The number of alleles and allele frequencies for the selected SNPs were calculated using a custom-made Perl script. The GenAlEx v6.503 (Peakall and Smouse 2012) was used to convert various file formats for

different analyses and to calculate genetic diversity parameters, including observed heterozygosity (*Ho*) and expected heterozygosity (*He*). The polymorphism information content (PIC) values were calculated using the following formula (Botstein et al. 1980), where P_i and P_j are the frequencies of i^{th} and j^{th} alleles for the selected marker, respectively.

$$PIC = 1 - \sum_{i=1}^{n} P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 P_j^2$$

Analysis of population structure and genetic relationships

The population genetic structure of the 96 genotypes was estimated using a Bayesian Markov Chain Monte Carlo (MCMC) algorithm implemented in STRUCTURE v2.3.4 (Pritchard et al. 2000) software. The population structure was analyzed employing K-values (an assumed fixed number of subpopulations) from 1 to 10 in the population. Five independent analyses were used for each K-value, and the program was set on 100,000 as burn-in iteration, followed by 100,000 Markov chain Monte Carlo (MCMC) replications (Chen et al. 2012; Zori´c et al. 2012). The best number of K was determined by STRUCTURE HARVESTER (Earl and Vonholdt 2012), applying the log probability of the data [LnP(D)] and delta K (ΔK) based on the rate of change in [LnP(D)] between consecutive K-values. For the optimal K-value, membership coefficient matrices of five replicates from STRUCTURE were fed to CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) to generate a single Q-matrix. Accessions with membership probabilities greater than 0.5 were assigned to a subgroup, while those with lower probabilities (< 0.50) were assigned to the “mixed” subgroup. Genetic distances between pairs of accessions were calculated using GenAlEx v6.503, from which a principal coordinate analysis (PCoA) was conducted. An unrooted neighbor-joining phylogenetic tree without the assumption of an evolutionary hierarchy was then constructed using the MEGA X program v10.2.2 (Kumar et al. 2018) based on the obtained distance matrix, with 1,000 bootstrap replicates. The number of subpopulations determined with STRUCTURE was used for AMOVA to test the structure of the genetic diversity of the genotypes. The AMOVA and calculation of genetic diversity indices were performed in GenAlEx v6.503. From AMOVA, the fixation index (F_{ST}) and Nm (haploid number of migrants) within the population were obtained.

Results

Distribution of SNPs in the *Oryza sativa* genome and genetic properties of markers

A total of 96 rice accessions were genotyped using GBS. After sequencing, data processing and SNP filtering, a total of 5,993 SNPs were identified. A sum of 5,856 high-quality SNPs was physically mapped across 12 chromosomes with an average marker density of 64.86 kb per chromosome, and the remaining 137 SNPs were identified as present in unclassified chromosomes. A genome-wide SNP density plot (Fig.

1) revealed that the highest number of SNPs were physically mapped to chromosome 1 (12.89%, 755 SNPs), and the lowest number of SNPs were mapped to chromosome 9 (4.76%, 279 SNPs). The highest and lower marker densities were observed on chromosome 2 (103.32 kb) and chromosome 12 (35.11 kb), respectively (Fig. 1, Table 1). Transition SNPs (4120, 68.75%) were more frequent than transversions (1873, 31.25%), with a ratio of 2.20. The frequencies of A/G and C/T transitions were almost similar (i.e., A/G 34.47% and C/T 34.27%). The A/G transitions (34.47%) accounted for the highest frequency, while G/C transversions (5.67%) occurred at the lowest frequency among all six SNP scenarios. The frequencies of the four transversion types ranked as follows: A/T (9.14%), A/C (8.28%), G/T (8.16%) and G/C (5.67%) (Table 2). The expected heterozygosity (He) in the population varied from 0.1 (93 SNPs) to 0.5 (161 SNPs) with an average of 0.25. The Polymorphic Information Content (PIC) values varied from 0.1 (154 SNPs) to 0.4 (822 SNPs) with an average of 0.22 (Fig. 2A). A total of 4238 (70.7%) SNPs had a minor allele frequency less than 0.2 (Fig. 2B).

Population Structure and Genetic Relationships

The STRUCTURE v2.3.4 was used to study the population structure and genetic relations among 96 rice accessions. K- value was employed to estimate the number of clusters in the rice landrace germplasm based on genotypic data. The optimal K-value can be determined by plotting the number of clusters (K) against ΔK, which showed a sharp peak at K=3 (Fig. 3). The optimal K value indicates that three subpopulations (SP1, SP2, SP3) showed the highest probability for population clustering, and these three subpopulations consisted of 19, 18 and 57 genotypes, respectively (Table 3). Among the accessions, two landraces, ‘Garudachampa’ (KS 72) and ‘Kavungin poothada’ (KS 107), were classified as admixed (Supplementary Material 2). Estimating the fixation index (F_{ST}) for each of the subpopulations from the STRUCTURE results suggested a significant divergence within the three populations (Table 3). An F_{ST} value of 0.52706, 0.43268 and 0.55714 was obtained for subpopulation 1 (SP1), subpopulation 2 (SP2) and subpopulation 3 (SP3) (Table 3). The principal coordinates analysis (PCoA) was performed based on the pairwise genetic distance matrix to investigate the population cluster among all the 96 *O. sativa* accessions. The first two components from PCoA contributed 29.74% and 48.29% variation, respectively. Following the STRUCTURE results, PCoA also showed three groups clustered separately (Fig. 4). Moreover, the distribution of three subpopulations was relatively concentrated.

A neighbor-joining phylogenetic tree was constructed to represent the genetic distances among the population (Fig. 5). ‘Gandhakasala’ and ‘Jeerakasala’, the two well-known aromatic rice landraces of Kerala, were grouped in subpopulation 1 (SP1). The landraces from other states such as Assam black, Basmati, Burma black, Ramla, Manipur black, White jasmine, Black jasmine and Red jasmine were clustered together subpopulation 2 (SP2). The two elite varieties Jaya and Athira, were grouped separately, as expected, in subpopulation 2. The other elite one, IR 20 (KS 120), had a different grouping in subpopulation 3. The phylogenetic tree showed IR 20 with Ponnaryan (KS 2) and Poonaran (KS 21). Photosensitive landraces like Velleriyan, Arikarai, Chitteni, Punchakazhama were clustered in subpopulation 3 (SP3). The landraces such as Allikkannan, Ponnaryan, Malakkaran, Rajakazhama (Rajkayama), Poonaran, Palliyaral, Thonnuran, Undakkazhama, Vachan, Kothambarikazhama were

grouped in SP3. Medicinal rice genotypes of Kerala such as Chennellu (KS 49), Chennellu 2 (KS 99), Kannichennellu (KS 29), Valiyachennellu (KS 41), Karachennellu (KS 126), Rakthasali (KS 35), Chuvanna Navara (KS 16), Manja Navara (KS 8), Karutha Navara (KS 20) were clustered together in SP3. However, Navara (KS 70) and Kunjinellu (KS 67) clustered in Subpopulation 1. The germplasm Rakthasali (KS 35) and Choman (KS 33) were shown the closest genetic relationship. Similarly, Koyyala (KS 32) and Karinchan (KS 78) showed the closest genetic relationship.

Genetic Differentiation of Populations

For identifying the distribution of genetic diversity between and within the populations of rice landraces, an Analysis of Molecular Variance (AMOVA) was performed. The three subpopulations identified in STRUCTURE were then applied in GenAlEx v6.503 to calculate the AMOVA. The AMOVA, F_{ST} and Nm are provided in Table 4. The results from AMOVA corroborated the existence of population genetic structure revealed by STRUCTURE. The AMOVA revealed that 27% of the total variation was found among the three subpopulations, while the remaining 73% was within subpopulations. Among the 73% variance within subpopulations, 36% was recorded among individuals and 37% variance within individuals. In addition, a high F_{ST} (0.266) and a low Nm (0.692) were obtained according to Nei's genetic distance analysis which confirmed the existence of strong genetic differentiation among rice landrace populations.

Discussion

Kerala is a major part of the Western Ghats-Sri Lanka biodiversity hotspot, enriched with diverse flora and fauna (Pradheep et al. 2021). The versatility in climate, topography, soil, farming methods, cultural and culinary preferences has made the state a source of diversity in rice. Inspite of introduction of many high yielding elite rice varieties, landraces are still popular in some farmers' fields due to their deliberate seed exchange programs of indigenous rice cultivars and awareness of conservation. Rice landraces included in this study such as Chennellu, Kannichennellu, Valiyachennellu, Karachennellu, Rakthasali, Chuvanna Navara, Manja Navara, Karutha Navara possess valuable medicinal properties. Chitteni, Arikarai, Mundon, kalladiyaran are drought tolerant landraces. Landraces like Thavalakannan, Karutha allikkanan, Choman, Chomala can survive in fields saturated with water (Vasumathy et al., 2020). These are important reservoirs of beneficial characteristics and need special attention for future conservation. In this study, using Genotyping by sequencing technology, a diversity panel of 96 rice germplasm from northern Kerala was genotyped and obtained 79,273 genome-wide SNPs. A sum of 5,993 high-quality SNPs was retrieved when filtered by keeping only bi-allelic sites with 20% missing data. Subsequently, analysis of SNPs distribution in the genome, the genetic diversity and population structure of rice population was done, which may encourage future breeding efforts like GWAS, MAS, etc., in *Oryza sativa*.

A higher frequency of transitions than transversions was observed (Ts/Tv= 2.20). Similar transition/transversion bias has been published in other plant species (Manickavelu et al. 2014; Taranto et al. 2016; Alipour et al. 2017; Eltaher et al. 2018; Luo et al. 2019). Therefore, it suggests that transition is more conservative than transversions because it is less likely to bring amino acid substitutions (Stoltzfus

et al. 2016; Luo et al. 2019). The frequencies of A/G and C/T transitions were almost similar (i.e., A/G 34.47% and C/T 34.27%), which contradicts the increased frequency of C/T transition due to deamination of cytosine (Beletskii et al. 1996; Bashir et al. 2014).

Polymorphic information content provides information about the marker loci. When PIC > 0.5 the marker loci can be interpreted as highly informative, when PIC value lies in between 0.5 and 0.25 (0.5 > PIC > 0.25) the markers can be assumed as reasonably informative and when PIC value falls below 0.25 (PIC < 0.25) they are slightly informative. In the present study, most SNP markers (80%) have a PIC value between 0.2-0.3, suggesting that these SNPs were reasonably informative markers. Gene diversity estimates the average heterozygosity and genetic distance among individuals in a population. In our study, the overall GD value (0.25) was slightly greater than the PIC value. This agrees with Luo et al. (2019), who stated that the PIC value should be less than or near GD when there are more alleles or an increase in the uniformity of allele frequencies in non-identical heterozygote individuals. The genotype panel with minor allele frequency (MAF) > 0.1 is desirable for genome-wide association mapping (Sharma et al. 2018), which shows that our marker density is good enough to perform genome-wide association studies.

The results of both STRUCTURE analysis and PCoA indicated that the 96 rice accessions could be clustered into three subgroups. The clustering of the rice landrace population into three subpopulations agrees with the genetic diversity analysis of rice landraces from Kerala using InDel markers (Vasumathy et al. 2020). The presence of population structure in our population may have arisen for several reasons: Although all genotypes have been cultivated in the same place for years, they were collected originally from different geographical regions. Therefore, it is impossible to establish whether the structure revealed through the present study is formed due to geographic distance. However, as rice landraces are domesticated rice populations, they might have undergone specific natural and artificial selection pressures, leading to population structure.

In the present study, the aromatic landraces ‘Gandhakasala’ and ‘Jeerakasala’ were clustered in the same group. This finding coincides with the earlier report of Rekha et al. (2011), which showed similar phylogenetic grouping of landraces using RAPD fingerprinting. In contrast, Vanaja et al. (2010) reported using SSR markers that these two landraces were clustered differently. The grouping of IR 20 with landraces indicates that it may have shared genetic characters with these landraces. The landraces such as ‘Allikkannan’, ‘Ponnaryan’, ‘Malakkaran’, ‘Rajakazhama’ (Rajkayama), ‘Poonaran’, ‘Palliyaral’, ‘Thonnuran’, ‘Undakkazhama’, ‘Vachan’, ‘Kothambarikazhama’ were grouped in SP3. A similar clustering pattern of landraces was observed in genetic diversity analysis using SSR markers (Vanaja et al. 2010). The germplasm ‘Rakthasali’ (KS 35) and ‘Choman’ (KS 33) were shown the closest genetic relationships when studied with molecular and morphological data. Farmers consider these two landraces different, but no evidence has been found in this study to differentiate them.

F_{ST} is the most relevant F-statistic used to study the amount of genetic differentiation between populations and within populations (Mohammadi and Prasanna 2003). An F_{ST} value equals 0 indicates

identical populations, above 0 implies a reduction in genetic exchange between populations, and a value of 1 indicates complete isolation, referring that the populations have different alleles that are fixed (Mohammadi and Prasanna 2003; Soorni et al. 2017). If the F_{ST} value is greater than 0.15, it can be considered significant in differentiating populations (Huang et al. 2010; Luo et al. 2019). An F_{ST} value of 0.266 was found between the three subpopulations, indicating moderate genetic differentiation between these three subpopulations. This result corresponded with AMOVA results, where 27% of the total variation was accounted for among subpopulations. Cheng et al. (2020) discussed three grades of Nm value and their biological significances. If the Nm (haploid) value is equal to or greater than one, high gene flow; if it is between 0.250-0.99, moderate gene flow; and between 0.0-0.249 signifies limited gene exchange between subpopulations. In the current study, the Nm value was 0.692, indicating modest gene flow among subpopulations. This result could prove the farmers' effort to maintain the genetic purity of rice landraces despite the extensive seed exchange programs across different states of India. Slatkin (1985) discussed significant evolutionary roles of gene flow in natural populations and reported that low levels of gene flow among subpopulations could furnish the opportunity for well-adapted combinations of genes to be established through natural selection.

The lack of genetic and genomic information of resources has been a major limiting factor for the genetic improvement of rice worldwide. This study is the first attempt wherein GBS-SNP markers have been used to assess the genetic potential of the untapped rice landrace gene pool in northern Kerala. The results of the contemporaneous study demonstrated that GBS is an effective tool for determining genetic diversity and the population structure of rice landraces. The clustering pattern based on SNP markers furnished sufficient knowledge in identifying and authenticating rice genotypes that eliminate the possibility of duplication among indigenous varieties. Since these rice landraces were originally gathered from different geographical regions, many of these genotypes should be precious reservoirs of genes to be used in breeding to address the challenges of biotic and abiotic stresses. Besides, the present study serves as a background for further progress in deciphering the genetics of nutritional and therapeutic properties of rice landraces that may lead the whole world to adopt a salubrious lifestyle.

Declarations

Funding

This research was supported by grants from Department of Science and Technology-Science and Engineering Research Board (DST-SERB), Government of India (Grant No. ECR/2016/001934).

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Code availability

Not applicable

Authors' contributions

AM conceived and supervised the research work; MP, SKV prepared genetic materials; AM acquired research grant for the research; HKKS, MP performed data analyses; MP wrote the manuscript; All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent to participate

Not applicable

Consent for publication

Not applicable

Acknowledgements

The authors acknowledge the Central University of Kerala for providing the research facilities. MP extends thanks to the University Grant Commission (UGC) for awarding financial aid.

References

Aiswariya K, Thomas G (2016) Characterization of five rice varieties using morphological traits and seed storage protein profiling. South Indian J Biol Sci 2(1):152-161

Alipour H, Bihamta MR, Mohammadi V, Peyghambari SA, Bai G, Zhang G (2017) Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. Front Plant Sci 8:1293

Bashir T, Sailer C, Gerber F, Loganathan N, Bhoopalan H, Eichenberger C, Grossniklaus U, Baskar R (2014) Hybridization alters spontaneous mutation rates in a parent-of-origin-dependent fashion in Arabidopsis. Plant Physiol 165(1):424-437

Begum H, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J, Virk P, Collard B, McCouch SR (2015) Genome-wide association mapping for yield and other agronomic traits in an elite breeding

population of tropical rice (Oryza sativa). PLoS One 10(3): e0119873

Beletskii A, Bhagwat AS (1996) Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. Proc Natl Acad Sci 93(24):13919-13924

Blakeney M, Krishnankutty J, Raju RK, Siddique KH (2020) Agricultural innovation and the protection of traditional rice varieties: Kerala a case study. Front Sustain Food Syst 3:116

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3):314–331

Chen X, Min D, Yasir TA, Hu YG (2012) Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. PLoS One 7(9): e44510

Cheng J, Kao H, Dong S (2020) Population genetic structure and gene flow of rare and endangered Tetraena mongolica Maxim. revealed by reduced representation sequencing. BMC Plant Biol 20(1):391

Crowell S, Korniliev P, Falcao A, Ismail A, Gregorio G, Mezey J, McCouch S (2016) Genome-wide association and high-resolution phenotyping link Oryza sativa panicle traits to numerous trait-specific QTL clusters. Nat Commun 7(1):1-14

D'Agostino N, Taranto F, Camposeo S, Mangini G, Fanelli V, Gadaleta S, Miazzi MM, Pavan S, Di Rienzo V, Sabetta W, Lombardo L, Zelasco S, Perri E, Lotti C, Ciani E, Montemurro C (2018) GBS-derived SNP catalogue unveiled wide genetic variability and geographical relationships of Italian olive cultivars. Sci Rep 8(1):1-13

Devi PI, Solomon SS, Narayanan M (2017) Conservation of traditional rice varieties for crop diversity in Kerala. Journal of Research ANGRAU 45:93-99

Descalsota GIL, Swamy BP, Zaw H, Inabangan-Asilo MA, Amparado A, Mauleon R, Chadha-Mohanty P, Arocena EC, Raghavan C, Leung H, Hernandez JE, Lalusin AB, Mendioro MS, Diaz MGQ, Reinke R (2018) Genome-wide association mapping in a rice MAGIC Plus population detects QTLs and genes useful for biofortification. Front Plant Sci 9:1347

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15

Earl DA, Vonholdt BM (2012) Structure harvester: a website and program for visualizing Structure output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KFM, Poland J, Baenziger PS (2018) Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. Front genet 9:76

Girma G, Nida H, Seyoum A, Mekonen M, Nega A, Lule D, Dessalegn K, Bekele A, Gebreyohannes A, Adeyanju A, Tirfessa A, Ayana G, Taddese T, Mekbib F, Belete K, Tesso T, Ejeta G, Mengiste T (2019) A large-scale genome-wide association analyses of ethiopian sorghum landrace collection reveal loci associated with important traits. Front Plant Sci 10:691

Gopi G, Manjula M (2018) Speciality rice biodiversity of Kerala: need for incentivising conservation in the era of changing climate. Curr Sci 114(5):997-1006

Grünwald NJ, Everhart SE, Knaus BJ, Kamvar ZN (2017) Best practices for population genetic analyses. Phytopathology 107(9):1000-1010

He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5:484

Hoang GT, Van Dinh L, Nguyen TT, Ta NK, Gathignol F, Mai CD, Jouannic S, Tran KD, Khuat TH, Do VN, Lebrun M (2019) Genome-wide association study of a panel of Vietnamese rice landraces reveals new QTLs for tolerance to water deficit during the vegetative phase. Rice 12(1):1-20

Hu Z, Mbacké B, Perumal R, Guèye MC, Sy O, Bouchet S, Prasad PV, Morris GP (2015) Population genomics of pearl millet (Pennisetum glaucum (L.) R. Br.): Comparative analysis of global accessions and Senegalese landraces. BMC Genomics 16:1048

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961-967

Jakobsson M, Rosenberg NA (2007) Clumpp: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806

Joseph J, Francies RM, Zachariah G, AV SK (2007) Characterisation of navara (Oryza sativa L.) a traditional medicinal rice of Kerala for qualitative traits. Indian J Agric Res 41(4):267-271

Karunakaran N (2014) Paddy cultivation in Kerala–Trends, determinants and effects on food security. Artha J Soc Sci 13(4):21-35

Kumar A, Sharma D, Tiwari A, Jaiswal JP, Singh NK, Sood S (2016) Genotyping-by-Sequencing Analysis for Determining Population Structure of Finger Millet Germplasm of Diverse Origins. The plant genome 9(2)

Kumar PS, Elsy CR, Nazeem PA, Augustin A (2010) Use of different marker systems to estimate genetic diversity in the traditional medicinal rice cultivar of Kerala. Intern J Plant Breed Genet 4(2):89-103

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35(6):1547-1549

Latha M, Abraham Z, Nair RA, Mani S, Dutta M (2013) Rice landraces of Kerala state of India: a documentation. Int J Biodivers Conserv *5*(4):250-263

Leng Y, Lv C, Li L (2019) Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). Mol Breed 39:38

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25:1754–1760

Liu JJ, Sniezko RA, Sissons R, Krakowski J, Alger G, Schoettle AW, Williams H, Zamany A, Zitomer RA, Kegley A (2020) Association mapping and development of marker-assisted selection tools for the resistance to white pine blister rust in the Alberta limber pine populations. Front Plant Sci 11:1404

Luo Z, Brock J, Dyer JM, Kutchan T, Schachtman D, Augustin M, Ge Y, Fahlgren N, Abdel-Haleem H (2019) Genetic diversity and population structure of a Camelina sativa spring panel. Front Plant Sci 10:184

Manickavelu A, Jighly A, Ban T (2014) Molecular evaluation of orphan Afghan common wheat (*Triticum aestivum* L.) landraces collected by Dr. Kihara using single nucleotide polymorphic markers. BMC Plant Biol 14:320

Manickavelu A, Joukhadar R, Jighly A, Lan C, Huerta-Espino J, Stanikzai AS, Kilian A, Singh RP, Ban T (2016) Genome-wide association mapping of stripe rust resistance in Afghan wheat landraces. Plant Science 252:222-229

Manjunatha GA, Elsy CR, Rajendran P, Joseph J, Francies RM, Krishnan S (2018) DUS Characterization of rice (Oryza sativa L.) landraces of wayanad, kerala. Electron J Plant Breed 9(2):617-630

Mgonja EM, Park CH, Kang H, Balimponya EG, Opiyo S, Bellizzi M, Mutiga SK, Rotich F, Ganeshan VD, Mabagala R, Sneller C, Correll J, Zhou B, Talbot NJ, Mitchell TK, Wang GL (2017) Genotyping-by-sequencing-based genetic analysis of African rice cultivars and association mapping of blast resistance genes against Magnaporthe oryzae populations in Africa. Phytopathology 107(9):1039-1046

Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci 43(4):1235–1248

Nimmakayala P, Levi A, Abburi L, Abburi VL, Tomason YR, Saminathan T, Vajja VG, Malkaram S, Reddy R, Wehner TC, Mitchell SE, Reddy UK (2014) Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. BMC Genomics 15(1):767

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28(19):2537–2539

Peterson GW, Dong Y, Horbach C, Fu YB (2014) Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity 6(4):665-680

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome 5(3)

Pradheep K, John KJ, Latha M, Suma A (2021) Status of crop plants of agricultural importance in Kerala state, India: an update. Genet Resour Crop Evol 68:1849-1873.

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

Raj RD, Kiran AG, Thomas G (2010) Population genetic structure and conservation priorities of Oryza rufipogon Griff. Populations in Kerala, India. Curr Sci 98(1):65-68

Rajani J, Deepu V, Nair GM, Nair AJ (2013) Molecular characterization of selected cultivars of rice, *Oryza sativa* L. using Random Amplified Polymorphic DNA (RAPD) markers. Intern Food Res J 20(2):919-923

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. PLoS One 8(6): e66428

Rekha T, Martin KP, Sreekumar VB, Madassery J (2011) Genetic diversity assessment of rarely cultivated traditional indica rice (Oryza sativa L.) varieties. Biotechnol Res Int

Schreiber M, Himmelbach A, Börner A, Mascher M (2019) Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. Evol Appl 12(1):66-77

Serba DD, Muleta KT, St Amand P, Bernardo A, Bai G, Perumal R, Bashir E (2019) Genetic diversity, population structure, and linkage disequilibrium of pearl millet. The Plant Genome 12(3):180091

Sharma SK, MacKenzie K, McLean K, Dale F, Daniels S, Bryan GJ (2018) Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. G3: Genes, Genomes, Genetics 8(10):3185-3202

Skaria R, Sen S, Muneer PMA (2011) Analysis of genetic variability in rice varieties (*Oryza sativa* L.) of Kerala using RAPD markers. J Genet Eng Biotechnol 1-24

Slatkin M (1985) Gene Flow in Natural Populations. Ann Rev Ecol Syst 16: 393-430

Soorni A, Fatahi R, Haak DC, Salami SA, Bombarely A (2017) Assessment of genetic diversity and population structure in Iranian cannabis germplasm. Sci Rep 7(1):1-10

Sreejayan, Kumar VR, Thomas G (2003) Collecting and morphological evaluation of 'Njavara', traditional medicinal rice (Oryza sativa L.) in Kerala. Plant Genet Resour Newsl 135:12-17

Sreejayan, Kumar US, Varghese G, Jacob TM, Thomas G (2011) Stratification and population structure of the genetic resources of ancient medicinal rice (Oryza sativa L.) landrace Njavara. Genet Resour Crop Evol 58:697–711

Stoltzfus A, Norris RW (2016) On the causes of evolutionary transition: transversion bias. Mol Biol Evol 33(3):595-602

Taranto F, D'Agostino N, Greco B, Cardi T, Tripodi P (2016) Genome-wide SNP discovery and population structure analysis in pepper (Capsicum annuum) using genotyping by sequencing. BMC genomics 17(1):943

Thomas J, Dominic VJ (2016) Assessment of genetic diversity and relationship of coastal salt tolerant rice accessions of Kerala (South India) using microsatellite markers. J Plant Mol Breed 4(1):35-42

Thomas G, Joseph L, Varghese G (2001) Discrimination between Oryza malampuzhaensis Krish. et Chand. and Oryza officinalis Wall ex Watt based on RAPD markers and morphological traits. Euphytica 122:181–189

Vanaja T, Randhawa GJ, Mammootty KP (2007) Pedigree evaluation and molecular diversity of some true breeding rice (Oryza sativa L.) genotypes of Kerala. J Trop Agric 44:42-47

Vanaja T, Singh R, Randhawa GJ (2010) Genetic relationships among a collection of Indica rice (Oryza sativa) genotypes of Kerala revealed by SSR markers. Indian J Agric Sci 80(3):191

Vasumathy SK, Peringottillam M, Sundaram KT, Kumar SHK, Alagu M (2020) Genome-wide structural and functional variant discovery of rice landraces using genotyping by sequencing. Mol Biol Rep 47:7391-7402

Yadav S, Sandhu N, Majumder RR, Dixit S, Kumar S, Singh SP, Mandal NP, Das SP, Yadaw RB, Singh VK, Sinha P (2019) Epistatic interactions of major effect drought QTLs with genetic background loci determine grain yield of rice under drought stress. Sci Rep 9(1):1-3

Yang X, Tan B, Liu H, Zhu W, Xu L, Wang Y, Fan X, Sha L, Zhang H, Zeng J, Wu D, Jiang Y, Hu X, Chen G, Zhou Y, Kang H (2020) Genetic Diversity and Population Structure of Asian and European Common Wheat Accessions Based on Genotyping-By-Sequencing. Front Genet 11:1157

Yu D, Wang H, Gu W, Qin T, Sun P, Lu Y, Shi B, Zheng H (2021) Genetic diversity and population structure of popcorn germplasm resources using genome-wide SNPs through genotyping-by-sequencing. Genet Resour Crop Evol 1-11

Zori´c M, Dodig D, Kobiljski B, Quarrie S, Barnes J (2012) Population structure in a wheat core collection and genomic loci associated with yield under contrasting environments. Genetica 140(4-6):259–275

Tables

Table 1 Genomic distribution of 5,856 SNPs mapped on 12 *Oryza sativa* chromosomes

Chromosomes	No. of SNPs	% SNPs	Start position	End position	Length (Mb)	Density (Kb)
Chr1	755	12.89	5,18,806	4,68,93,994	46.38	61.42
Chr2	361	6.16	1,36,547	3,74,36,783	37.30	103.32
Chr3	578	9.87	4,85,344	4,18,27,641	41.34	71.53
Chr4	599	10.23	3,43,975	3,36,28,723	33.28	55.57
Chr5	522	8.91	2,03,926	3,11,23,849	30.92	59.23
Chr6	455	7.77	5,65,885	3,21,29,220	31.56	69.37
Chr7	413	7.05	12,112	2,78,93,648	27.88	67.51
Chr8	416	7.10	44,611	3,01,75,628	30.13	72.43
Chr9	279	4.76	1,23,747	2,13,21,043	21.20	75.98
Chr10	415	7.09	1,29,896	2,21,34,477	22.00	53.02
Chr11	423	7.22	17,127	2,28,15,347	22.80	53.90
Chr12	640	10.93	44,851	2,25,13,172	22.47	35.11

Table 2 Percentage of transition and transversion SNPs across the *Oryza sativa* genome

SNP Type	Transitions		Transversions			
	A/G	C/T	A/T	A/C	G/T	G/C
Number of allelic sites	2,066	2,054	548	496	489	340
Frequencies	34.47%	34.27%	9.14%	8.28%	8.16%	5.67%
Total (Percentage)	4,120 (68.75%)		1,873 (31.25%)			

Table 3 The STRUCTURE results of 96 *Oryza sativa* accessions for the fixation index (F_{ST}), observed heterozygosity, expected heterozygosity and number of genotypes assigned to each subpopulation

Population	Inferred clusters	Mean F_{ST}	Obs. Het	Exp. Het	No. of Genotypes
SP1	0.277	0.52706	0.10081	0.21316	19
SP2	0.368	0.43268	0.13280	0.23408	18
SP3	0.355	0.55714	0.08240	0.18606	57

Table 4 Analysis of molecular variance (AMOVA) using 5,993 SNPs of the genetic variation among and within three subpopulations of 96 *Oryza sativa* accessions

Source	df	SS	MS	Est. Var.	%
Among Pops	2	26911.039	13455.520	239.094	27%
Among individuals	91	89819.902	987.032	325.660	36%
Within individuals	94	31557.000	335.713	335.713	37%
Total	187	148287.941		900.466	100%
Fixation index (F_{ST})	0.266				
Nm	0.692				

*P value 0.001

Figures

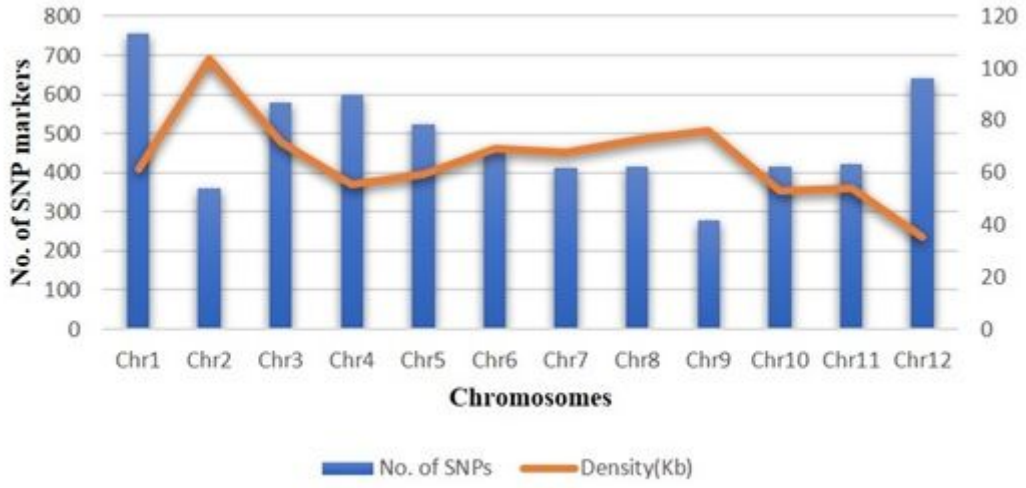

Figure 1

Genomic distributions of 5,856 SNPs mapped on 12 Oryza sativa chromosomes and the corresponding SNP density

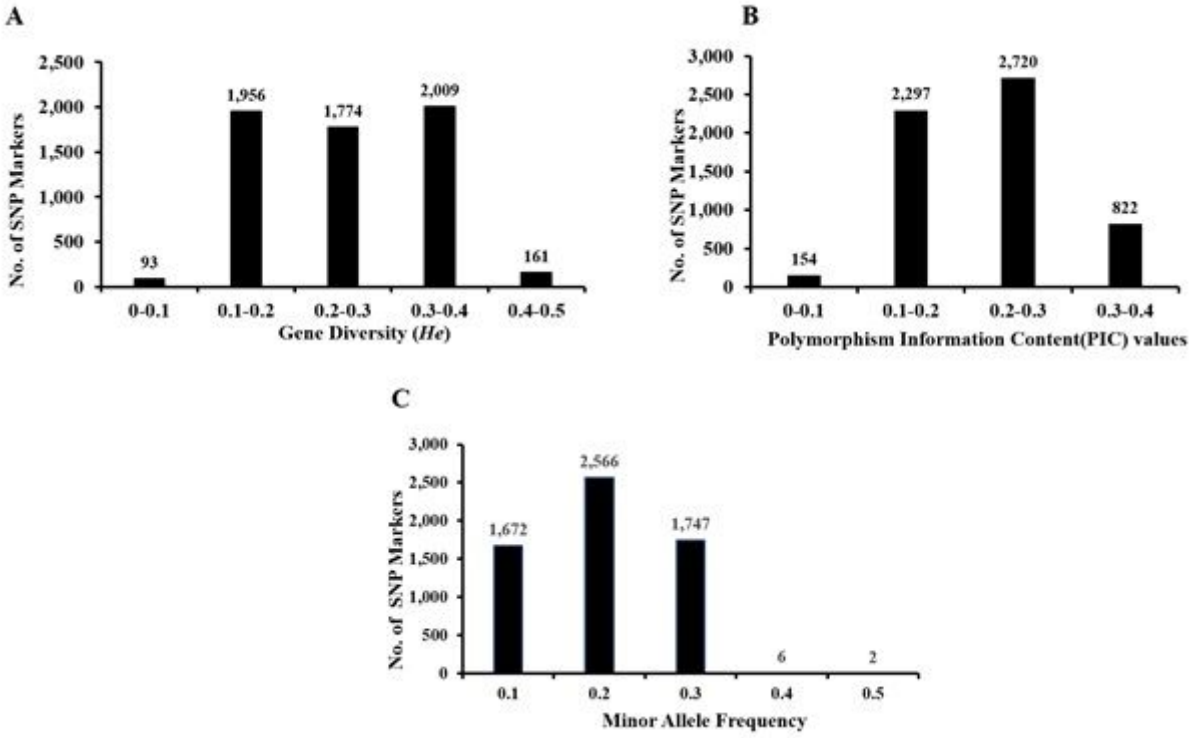

Figure 2

Distribution of genetic diversity for 5,993 SNP markers in 96 Oryza sativa accessions. (A) Gene diversity/ expected heterozygosity (He); (B) Polymorphic Information Content (PIC); (C) Minor Allele Frequency (MAF)

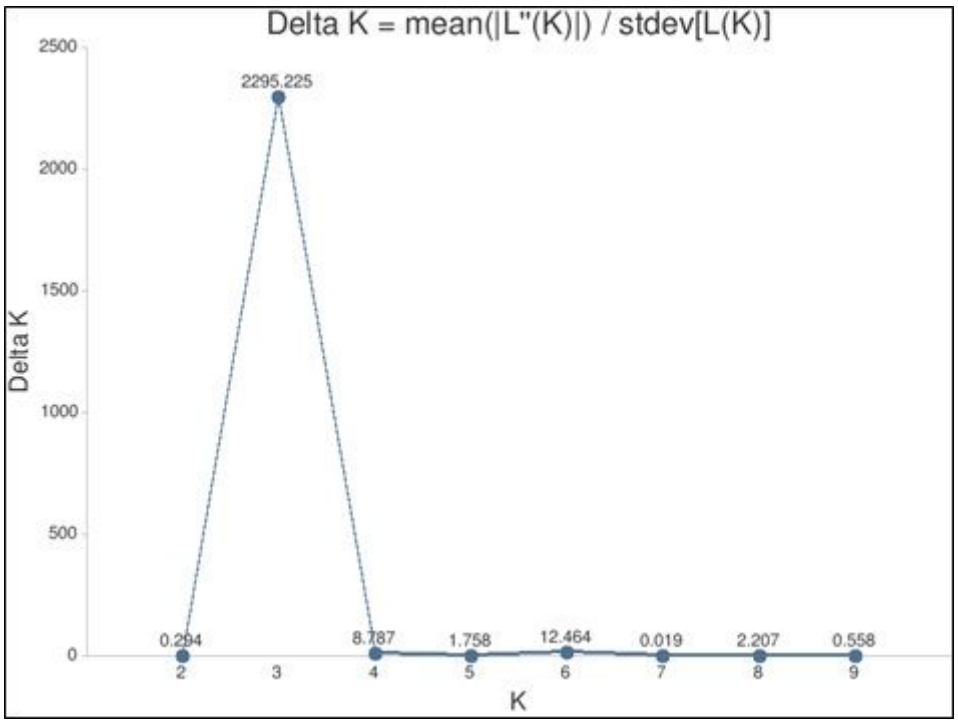

Figure 3

Delta K (ΔK) for different numbers of subpopulations (K)

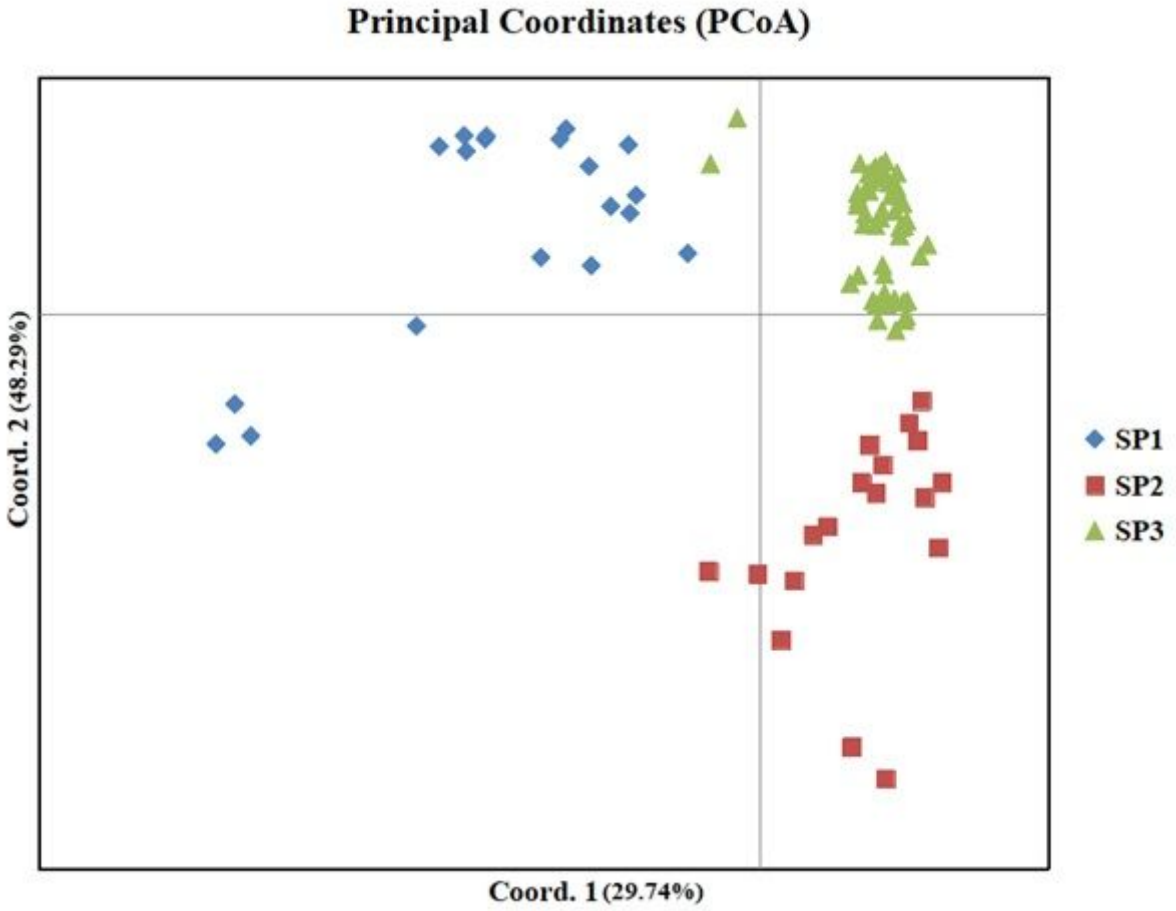

Figure 4

Principal Coordinates Analysis (PCoA) based on genetic distance showing three clustered subpopulations within studies Oryza sativa accessions

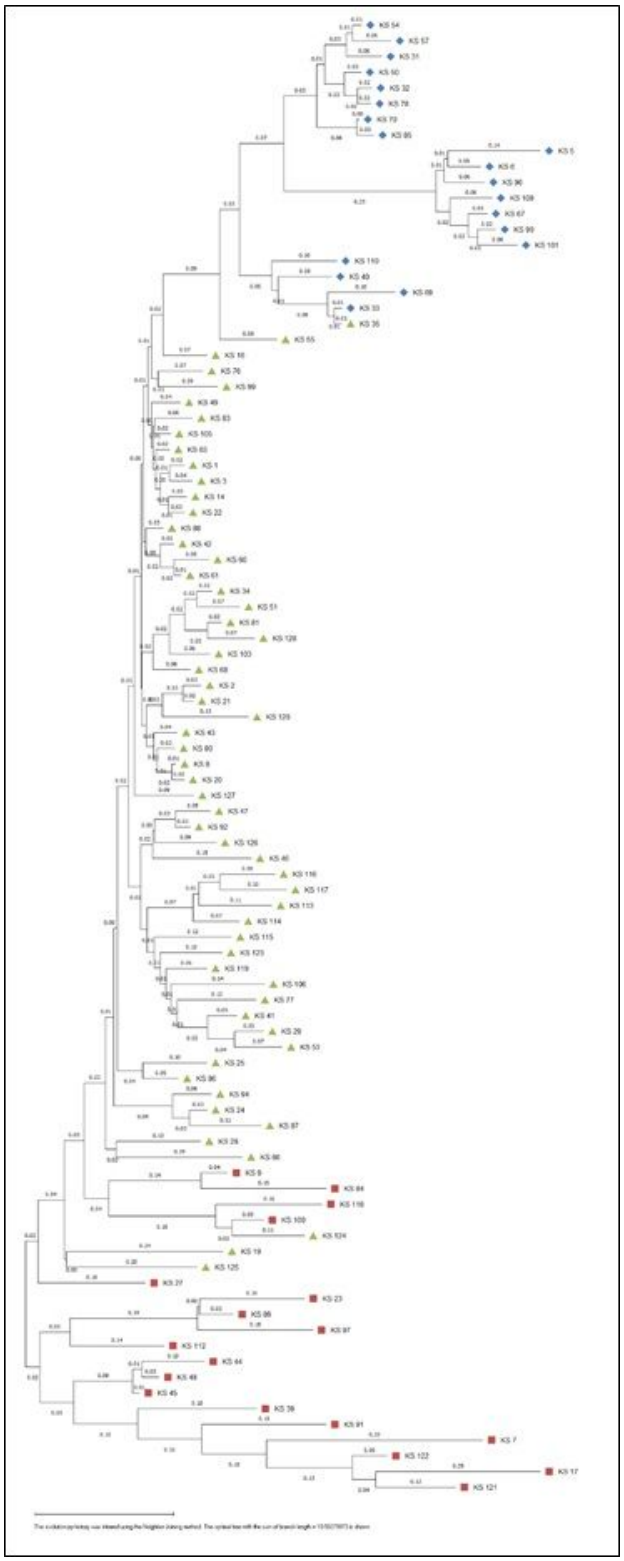

Figure 5

The neighbor-joining phylogenetic tree based on genetic distance matrix representing the grouping of 96 Oryza sativa accessions

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile1.pdf
- Supplementaryfile2.pdf