Indian J. Genet. Plant Breed., (2024); 84(4): 587-601
https://doi.org/10.31742/ISGPB.84.4.9
ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Genetic diversity and marker-trait association analysis for grain quality, yield and yield attributes in hill rice (*Oryza sativa* L.)

Harendra Verma*, Sakuonuo Theunuo, S. P. Das[1], Amit Kumar[2], C. Aochen, Lamalakshmi Devi[3], K Sarika[4], L. K. Baishya, D. J. Rajkhowa, H. Kalita, V.K. Mishra[2] , Parthaa Pratim Behera[5] and R. N. Sarma[5]

Abstract

The studies on genetic diversity and marker-trait association analysis of north-eastern hill rice for grain quality and yield attributes are of enormous importance for the identification of promising rice genotypes with better grain and cooking quality. Genetic diversity among 130 rice genotypes of North East India and marker-trait association was assessed using genotypic data of 90 SSR markers and phenotypic data. The phenotypic diversity analysis showed considerable variation across genotypes for grain quality and yield-attributing traits. Population structure analysis, distance-based neighbor joining cluster and principal coordinate analysis using genotypic data grouped the genotypes into two sub-populations. Analysis of molecular variance and pairwise FST values showed significant differentiation among all the pairs of sub-populations. Marker trait association analysis revealed a total of six associations at $p < 0.0001$ for grain quality and yield attributing traits with R^2 ranging from 3.55 to 11.91% under the upland ecosystem. The present study validated the association of the RM240 marker with Gel consistency and plant height gene on chromosome 3. The study indicated the presence of novel QTLs for days to maturity with RM210 and RM105 and days to flowering with RM101.

Keywords: Hill rice, molecular markers, diversity analysis, marker-trait association.

Introduction

Rice is consumed by 3.5 billion people, especially in Asia. This food source offers 31% of human caloric needs. India ranked second globally in rice output, producing 127.4 million tons on 44.6 million hectares (FAOSTAT 2022). Improving rice grain productivity by 70% is crucial to feed a growing population of 10 billion by 2050 despite ongoing biotic and abiotic challenges (FAOSTAT 2018). The demand for high-quality rice has increased due to improved life standards and, increased awareness of rice's nutritional value and changing eating habits. This necessitates plant breeders to develop rice varieties with enhanced yields and quality (Hori and Sun 2022).

The price of rice depends on quality, which is influenced by complex quantitative trait genetics such as grain color, size, shape, gel consistency, cooking and eating quality, nutritional content, and fragrance. Cooking and eating quality are also affected by amylose concentration (AC), gel consistency (GC), and gelatinization temperature (GT). Historical, geographical, and social factors influence rice grain quality preferences (Custodio et al. 2019). Southeast Asians favor soft gel consistency, long grains, and intermediate AC, while South Asians choose firm gel, long grains, and high AC (Sultana et al. 2022). Consumers of north China prefer

ICAR Research Complex for NEH Region, Nagaland Centre, Medziphema 797 106, Nagaland, India.

[1]ICAR Research Complex for NEH Region, Tripura Centre, Lembucherra 799 210, Tripura, India.

[2]ICAR Research Complex for NEH Region, Umroi Road, Umiam 793 013, Meghalaya, India.

[3]ICAR Research Complex for NEH Region, Sikkim Centre, Tadong 737 102, Sikkim, India.

[4]ICAR Research Complex for NEH Region, Manipur Centre, Imphal 795 004, Manipur, India.

[5]Department of Plant Breeding and Genetics, Assam Agricultural University, Jorhat 785 013, Assam, India.

***Corresponding Author:** Harendra Verma, ICAR Research Complex for NEH Region, Nagaland Centre, Medziphema 797 106, Nagaland, India, E-Mail: harendraicar@gmail.com

How to cite this article: Verma H., Theunuo S., Das S.P., Kumar A., Aochen C., Devi L., Sarika K., Baishya L.K., Rajkhowa D.J., Kalita H., Mishra V.K., Behera P.P. and Sarma R.N. 2024. Genetic diversity and marker-trait association analysis for grain quality, yield and yield attributes in hill rice (*Oryza sativa* L.). Indian J. Genet. Plnt Breed., **84**(4): 587-601.

Source of support: Nil

Conflict of interest: None.

Received: Aug. 2024 **Revised:** Oct. 2024 **Accepted:** Nov. 2024

whole-grain milled rice with lower GT, softer GC and lower AC (Wang et al. 2019). Southern China, South and Southeast Asia, and the US prefer skinny rice, while Northern China, Korea, and Japan prefer round rice. Hong Kong likes gel-like long-grain, low-amylose rice, while Laos, Korea, and Japan appreciate sticky rice (Juliano and Villareal, 1993). Whereas in the north-eastern hill (NEH) region of India, people of Nagaland, Manipur, Mizoram, Meghalaya, and Arunachal Pradesh states prefer Mizoram, Arunachal, and Meghalaya prefer a range of amylose types of rice from waxy to vary low amylose content. Rice with low amylose content ranging from 10-16% is used for daily consumption, while waxy rice is used for the preparation of savory dishes, dessert, porridge, gum rice roties, cakes, biscuits, pancakes, puffed rice, flakes and most popularly as rice beer, whereas Assam, Tripura, and Sikkim prefer intermediate amylose-containing rice (Verma et al. 2021; Kapoor et al. 2019). Short-grain glutinous rice is preferred in Japan and Taiwan, whereas medium-grain rice with intermediate amylose content is preferred in Indonesia and Myanmar. Meanwhile, Malaysian, Philippines and Iranian consumers prefer aromatic long-grain rice with intermediate amylose content (Aznan et al. 2003). Meeting market expectations for rice with location-specific desirable qualities requires grain quality parameter analysis.

To understand species diversity and their genetic makeup, genetic diversity studies are essential to discover and conserve ideal parental genotypes and establish genetic linkages (Roy and Shil 2020). Understanding genetic diversity is crucial for preservation methods and expanding the genetic foundation to meet global and local food demand. Environmental factors greatly affect grain quality, yield, and yield-related attributes (Berdugo et al. 2017). The North-Eastern Himalayan (NEH) Region is home to a variety of tribes and ethnic groups. Different tribes produce their favored rice landraces based on taste, quality, and other sociocultural aspects. A wide variety of rice landrace are found in North East India. The region of Northeast India, which is geographically close to the eastern Himalayan range, has a rich history of cultivating various indigenous rice varieties. These varieties are well adapted to the diverse agro-climatic conditions of the region. The Eastern Himalayan states of India, including Assam, Arunachal Pradesh, Nagaland, Manipur, Mizoram, Meghalaya, and Tripura, are known for their unique rice germplasm. These rice germplasm exhibits remarkable diversity, with hundreds of landraces adapted to the challenging high-altitude and often monsoon-dependent conditions. The origin of rice in this region is either independent domestication from wild cousins like *Oryza rufipogon* or migration from China with adaptation and diversification. The savory, fragrant, and superfine Joha rice from Assam, the Kalikhasa rice from Tripura, and the GI-tagged exceptional glutinous black scented chakhao (Chak - rice; ahaoba - wonderful) rice from Manipur, which is strong in antioxidants, are some examples of Northeast India's renowned quality rice. However, there are very few studies on the grain quality and yield characteristics of rice from the North East Hill region. Therefore, to create better rice varieties, these landraces must be preserved and utilized by wisely choosing among local cultivars based on genetic diversity and structure (Tarang et al. 2020).

Molecular marker-based diversity analysis accomplishes genotype selection faster and more accurately than breeding. Simple sequence repeats (SSRs) are commonly used PCR-based molecular markers due to their co-dominance, high repetition, polymorphism, and durability. Allelic count, which measures genetic diversity and evolutionary potential in a population, is less commonly used than observed (Ho) and predicted (He) heterozygosity estimations due to sample size (Pattanayak et al. 2018). While both Ho and He are mostly influenced by allele frequencies, a decrease in allelic richness, which measures allele diversity rather than frequency, may hinder a population's ability to adapt to different environments. A lot of allelic variability can be detected using SSRs (Singh et al. 2013). Population structure and genetic diversity studies using association mapping (AM) can also reveal genes responsible for critical agronomic features (Wang et al. 2023). This method permits functional variation research over more germplasm. Therefore, the present study aimed to explore marker-mediated genetic diversity and genetic structure in rice of the North-Eastern Himalayan Region and identifying markers associated with grain quality, yield, and yield related attributes through association analysis.

Materials and methods

Plant materials

The experimental material consisted of 130 hilly rice landraces collected from different parts of Nagaland and Manipur (Supplementary Table S1). The genotypes were directly sown in moist soil in a randomized complete block design (RCBD) with three replications at a row-to-row spacing of 30 and 10 cm plant-to-plant distance in the upland ecosystem. The material was evaluated for yield and grain quality traits during two consecutive *kharif* seasons of 2019 and 2020 at ICAR Research Complex for NEH Region, Nagaland Centre. This region is located at an altitude of 295m above mean sea level and lies at a latitude of $25^0$45'24" N and a longitude of $93^0$50'26» E. The recommended agronomic practices, including nutrient, pest and disease management practices, were followed for the cultivation of these landraces. The recommended dose of N, P and K @ 120, 60 and 60 kg/ha was applied in the form of urea, SSP and MOP. A half dose of N and a full dose of P and K were applied to the field as basal doses before the last plowing. The remaining half dose of N was applied at the tillering stage.

Data recording

Mean performance was recorded for yield traits such as days to flowering (DF), days to maturity (DM), plant height (PH), effective tiller number (ETN), panicle length (PL), grains per panicle (GPP) spikelet fertility (SF), after panicle maturation, whereas, yield per plant (YPP) was recorded from two months air-dried harvested crop (Supplementary Table S2). Similarly, grain quality traits such as grain length (GL), grain width (GW), decorticated grain length (DGL), decorticated grain width (DGW), gelatinization temperature (GT), gel consistency (GC), amylose content (AC) and aroma were estimated from 3 months stored harvested crop (Supplementary Table S2).

Amylose content was calculated using Juliano (1979) and classified as extremely low (3–9%), low (10–19%), medium (20–25%), high (26–30%), and very high (>30%). GT was assessed indirectly through the alkali spreading the value of the hulled kernel (Little et al. 1958) and based on the alkali spreading value, GT was categorized as low (6–7), medium (4–5), high, medium GT (3) and high GT (1–2). Similar to Zhang et al. (2020a) GC was measured. The landraces were classified by gel length as very hard (<25 mm), hard (26–40 mm), medium (41–60 mm), and soft (61–100 mm). Cooked rice with a hard gel consistency hardens after cooling, while rice with a soft gel consistency stays soft. A vernier caliper measured the length and breadth of fifteen decorated rice in each replication, and the average was calculated from three replications. Grain shape was determined as per the guidelines of the PPV&FRA (PPV&FRA 2007) classification.

SSR genotyping

The total genomic DNA of each genotype was extracted from healthy leaf tissue of 10-day-old seedlings following the Dellaporta method (Dellaporta et al. 1983) with slight modification. The DNA samples were stored in 1X TE buffer. The PCR amplification conditions were based on the procedure of Panaud et al. (1996). PCR was carried out for amplification of the genetic markers or genomic region with a 10 µL reaction volume consisting of 5 µL of PCR master mix (TaKaRa EmeraldAmp GT PCR Master Mix®), 0.5 µL each of forward and reverse primer and 1-µL of the DNA sample and 3 µL of double distilled water. The reagents were mixed thoroughly and then placed in a Thermal Cycler (PCR Gene AMP® 2400, Applied Biosystems, USA) for cyclic amplification. The amplified products were separated by size in a 3% agarose gel. The image of ethidium bromide-stained gel was digitally documented in the Gel Documentation System. Comparing the band size of the 100-bp ladder (Promega, WI, USA) with IR-36 as a molecular weight reference showed the base pair size of separate amplified bands.

Phenotypic diversity analysis

The agronomic yield and grain quality data were subjected to analysis of variance (ANOVA). The mean data of all the traits were used to determine the phenotypic diversity among the genotypes. The phenotypic diversity was illustrated by *K* mean cluster analysis and principal component analysis (PCA) using the R package FactoMineR (Le et al. 2008), Factoextra (Kassambara and Mundt 2020) and clustering analysis based on the algorithm of Ward's method (Galili 2015).

Factor analysis was used to group the 15 quantitative traits into statistical factors that explain much of their overall variability. Principal component analysis with Spearman correlation matrix extracted these factors. The data was normalized using a z-transformation. The factor pattern (factor loading) and eigenvalue indicated the factors. Factors with eigen values greater than 1 and commonalities greater than 0.5 for the variables were chosen for further analysis. To maximize interdependent variable relationships, these factors were varimax rotated with three factors. The adequacy of the sampling was assessed through the Kaiser-Mayer-Olkin measure (Kaiser 1974), and traits with values surpassing 0.5 were retained. Lastly, two biplots were generated to illustrate the distribution of accessions along the factor axis. For further examination of the distribution of genotypes among the factors, collection locations were divided into four altitude groups (Group 1: 700–100 m, Group 2: 1000–1400 m, Group 3: 1400–1600 m and Group 4 - > 1600m) mentioned in Supplementary Table S1. Four districts, Longleng, Mokokchung, and Wokha (all in Nagaland state; 42, 20, and 26 genotypes, respectively) and Imphal (in Manipur state; 16 genotypes) contributed about 80% of the genotypes studied. Therefore, district-wise distribution within the factors was also studied for these four districts.

Genetic diversity analysis

The Bayesian clustering approach and admixture model-based clustering approach in software STRUCTURE 2.3.4 were used to analyze the population structure of 130 rice genotypes utilizing 90 polymorphic SSR markers (Pritchard et al. 2000; Falush et al. 2007). To determine the true number of subpopulations, 10 runs were performed for every K value. Burn-in was 1,00,000 with a 5,00,000 MCMC (Markov Chain Monte Carlo) iteration. The mean estimate of the log posterior probability of the data [L (K)] was plotted against the specified K value to determine the K value. The delta K (ΔK) method from Evanno et al. (2005) was used to estimate K using Structure Harvester version 0.6 (Earl and VonHoldt 2012).

Genetic diversity among rice genotypes was determined by analyzing the number of alleles (Na), effective alleles (Allelic richness) (Ne), total expected heterozygosity (Ht), gene diversity (He), observed heterozygosity (Ho), inbreeding coefficient (FIS), fixation index (FST), inbreeding coefficient to total (FIT), gene flow (Nm), and fixation index (F) using GenAlEx 6.51b2 (Peakall and Smouse 2012). FST examines genetic variance based on Wright's F-statistics

(Wright 1965). The genetic distance between genotypes was determined using a basic matching dissimilarity coefficient of allelic data with a 1000-time bootstrap value, and genotypes were clustered using the neighbor-joining method without the assumption of an evolutionary hierarchy (Perrier and Jacquemoud 2006). Analysis of molecular variance (AMOVA) was used to estimate molecular variance components within the population, within individuals, and between populations using GeneAlEx 6.51b2 (Peakall and Smouse 2012) with 1000 permutations. Principle Coordinate Analysis (PCoA) was used to create a similarity matrix in DARwin 6.0.021. (Perrier and Jacquemoud 2006). The allelic counts for all hypothetical populations designated according to their collection sites were computed. To ensure that allelic counts were not affected by sample size, the permutation method, as outlined by Fu et al. (2003) was done using FP Test R software, with 10,000 repetitions of allele permutations.

After taking into consideration the gross level population structure (Q) in the GLM analysis, association analysis between marker loci and phenotypic characteristics was carried out in all trials using TASSEL (Trait Analysis by association, Evolution and Linkage) software version 4.0 (Bradbury et al.2007). The MLM analysis with the P3D algorithm uses the Q+K (kinship) model (Zhang et al. 2010). According to reports, the Q + K model lessens false positive connections (Rincent et al. 2014). As a result, the study reports the outcomes of MLM using the Q+K model. Each marker-trait connection was evaluated for significance using the marker P value (0.001).

Results

Analysis of phenotypic traits

Analysis of variance showed genetic differences for all study traits (Table 1). The optimal K value for the population was three, according to the Elbow Method (Fig 1a.) and three clusters were generated in the K-means clustering (Fig 1b). Cluster III was the largest, with 52 genotypes, followed by cluster 1 with 45 genotypes and cluster 2, with 33 genotypes. The genotypes of cluster I have medium duration, semi-dwarf, long grain length and preferred grain quality traits. Cluster-II genotypes feature tall, early-maturity lengthy panicles, such as Chakha (30.33 cm) and Chachak Hou (29.3 cm). Cluster-III genotypes were tall, early maturity with high yield potential, such as the Tsuksa landrace of this cluster has the best production potential of all genotypes (4.77 q/ha). With regards to grain quality, yield, and yield attributes, PCA analysis showed that PC1, with an eigenvalue of 3.5, accounted for 23.7% of the contribution to the total variance and PC2, with an eigenvalue of 2.2 accounted for 14.67% of the contribution to the total variation (Fig. 2). These two major components include plant height, panicle length, spikelet fertility, decorticated grain length and width, days to maturity, days to 50% flowering, and grain yield.

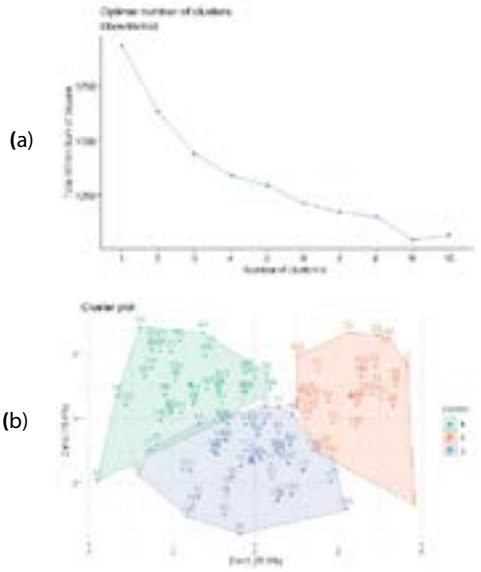

Fig. 1(a) Optimum number of clusters identified using Elbow method and (b) Grouping of genotypes based on the grain quality and yield related traits using K-means clustering. Note: Numbering indicates the genotypes serial number in Supplementary Table 1

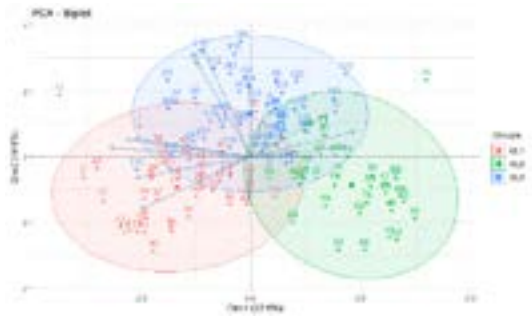

Fig. 2. Grouping of genotypes based on the first two principal components

Out of the ten factors, three (referred to as F1, F2, and F3) displayed desirable eigen values (3.149, 1.788, and 1.152, respectively; Kaiser-Mayer-Olkin measure (Kaiser 1974) and were retained for subsequent analysis. The first factor (F1) with the highest eigen value was designated as the 'duration and starch quality' factor, owing to its notably positive loadings for DF, DM, and GT. The second factor (F2), was designated as the 'grain boldness' factor, which exhibited prominent contributions from GW and DGW. The third factor (F3), termed the 'grain length' factor, displayed a greater involvement from GL and DGL. The communalities of the variables demonstrated that these three factors accounted for 56 to 69% of the variance in DF, DM and GT, over 67 to 75% for GW and DGW, and 75 to 90% of GL and DGL. Following the varimax rotation, these three factors jointly elucidated 40.6% of the total variability. The biplot illustrating the interrelation between Factor 1 and Factor 2 (D1:D2 in Fig. 3A) accounted for 28.06% of the variation and dispersed the accessions with respect to the

Table 1. ANOVA of 130 genotypes for grain quality and yield and yield attributing traits.

Source of Variation	Df	Mean square												
		FD	*DM*	*PH*	*ETN*	*PL*	*GPP*	*SF*	*YD*	*GL*	*GW*	*DGL*	*DGW*	*AC*	*GC*	*GT*
Env	1	0.62	0.79	1611.00	21.70	0.01	3338.50	1352.80	0.03	0.14	0.03	0.02	0.01	79.09	85.34	0.000
Rep (Env)	4	22.53	30.89	2008.00	36.08	4.88	302.40	64.93	0.13	0.09	0.03	0.03	0.01	33.54	102.33	0.001
Gen	129	294.04*	4781.18*	6594.00*	12.55*	39.94*	3420.60*	338.58*	5.11*	7.86*	1.05*	3.76*	0.68	263.34*	4093.40*	0.440*
Gen (Env)	129	1.05	1.53	695.00	1.14	1.09	1.9	43.96*	0.02	0.00	0.02	0.02	0	1.35	0.98	0.002
Residual	516	2.46	2.35	2195.00	0.55	1.02	28.60	7.63	0.03	0.01	0.00	0.03	0.00	1.51	13.49	0.002
Mean		88.52	119.48	144.80	6.54	25.70	128.34	85.12	3.09	7.64	2.90	5.70	2.56	12.27	56.94	0.37
CV		1.77	1.28	32.36	11.35	3.93	4.17	3.25	5.36	0.96	2.42	2.97	2.32	10.01	6.45	11.88
Pr (>F) Gen		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pr (>F) Gen x Env		0.999	0.992	1.00	0.00	0.353	1.00	0.00	0.95	1.00	0.00	0.95	0.00	0.737	1.00	0.600

DF = Days to flowering; DM = Days to maturity; PH = Plant height; ETN = Effective tiller number; PL = Panicle length; GPP = Grains per panicle; SF = Spikelet fertility; YPP = Yield per plant; GL Grain length; GW = Grain width; DGL = Decorticated grain length; DGW = Decorticated grain width; GT = Gelatinization temperature; GC = Gel consistency and AC = Amylose content

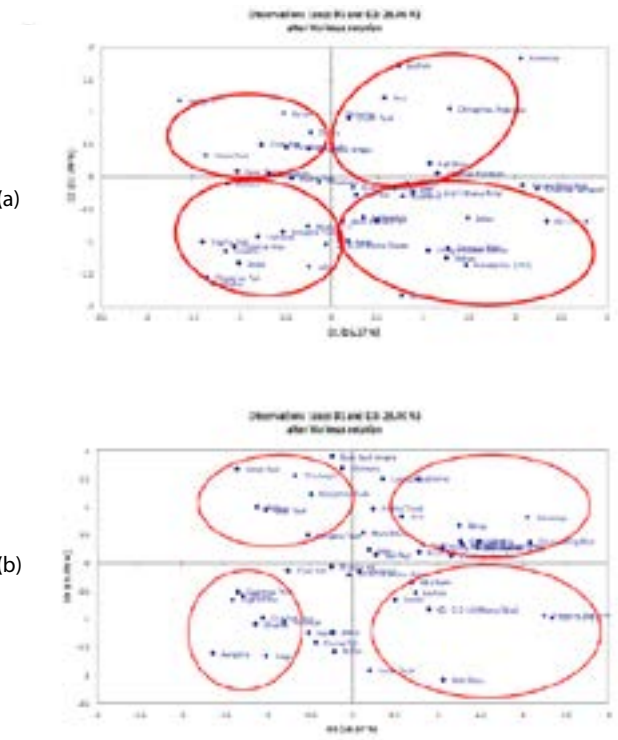

Fig. 3. (a) Grouping of genotypes based on the first two factors (b) Grouping of genotypes based on the first and third factors

duration and, starch quality and grain boldness factors. This visualization also separated the accessions into four distinct groups (labeled A-D in Fig. 3A). Group 'A' encompassed accessions that exhibited moderate to high characteristics for the duration and starch quality factor (F1 or D1) but low to very low attributes for the grain boldness factor (F2 or D2). Group 'B' included accessions with moderate to very high values for both F1 and F2 factors. Group 'C' comprised accessions with moderate to very high attributes for the F2 factor and moderate to low characteristics for the F1 factor. Finally, Group 'D' encompassed accessions that displayed moderate to extremely low values for the F1 and F2 factors. Some unique genotypes were also seen in the biplot. The genotype Kemenya was highest for the F2 factor and very high for the F1 factor. On the other side genotype, Ereima was the lowest for the F2 factor and moderate for the F1 factor. The genotype Aongsho was very high for F2 but low for F1. Genotypes Chamnya Yoh and Dhaha were very low for F1 and F2. The biplot involving Factor 1 and Factor 3 (D1:D3 in Fig. 3B) also highlighted four groups and distinct accessions such as 'Hah Shou (lowest for F3), Epyo tsuk longsa (highest for F3), and Aongsho (lowest for both F1 and F3). An examination of the altitudinal distribution of the collected accessions unveiled that Group 2, with the highest number of accessions (92 out of 130), was distributed uniformly across the three factors (27, 32 and 33 for F1, F2 and F3, respectively), which indicated that in the mid-altitudes of the study area, a very diverse group of genotypes are grown.

Group 1 genotypes were mostly in the F1 factor followed by the F3 factor, indicating that the group consisted of long-duration genotypes with somewhat slender grains (a preference in NEH valleys). Group 3 genotypes also showed a similar distribution. There were only three genotypes from Group 4 and two of them belonged to Factor 3 (F3). A closer look at the district-wise distribution showed that Longleng genotypes contributed significantly and uniformly across all three factors but the contribution was mostly negative (78.5% genotypes). This indicated that the genotypes were not very long duration, low tillering and medium sticky type, which is expected because they are grown interchangeably in aerobic uplands and rainfed irrigated conditions and semi-sticky types are preferred by the population of that area. Mokokchung genotypes also contributed somewhat evenly across the three factors. About 60% of genotypes contributed negatively to Factor 1 and Factor 3 (medium duration semi-sticky), while 40% of genotypes contributed positively to Factor 2 (bold type). This distribution also matches with the cultivation practice similar to Longleng district. Wokha district showed a different picture. About 50% of genotypes were positive for Factor 2, indicating a preference for bold grains. The distribution of genotypes from the Imphal district of Manipur (valley area with rainfed irrigated cultivation) showed that 81% of the genotypes were positive for Factor 1 and Factor 3, which indicated that genotypes were mostly long duration and slender grain type, which matches with the growing condition and grain type preference of the population.

SSR polymorphism among rice genotypes

The PIC of the 90 SSR markers with information on allele number, size and allele frequency have been worked out. The PIC of a marker defines its ability to detect polymorphism among individuals in a population. The PIC values of the SSR markers range from 0.19 for RM 480 to 0.80 for RM 5891, with a mean of 0.518 for all the accessions under study. Markers with PIC values greater than 0.5 are considered highly informative and useful. This indicated the informativeness of the markers and the presence of high allelic diversity in the population for the markers used under study. This suggests the usefulness of markers for genetic diversity and QTL mapping studies. It is also a reflection of allelic diversity among varieties (Suvi et al. 2020). The highest PIC value of 0.80 was observed for RM 5891, followed by 0.68 for RM 1370 and 0.66 for RM 267. The highest allele frequency of 0.84 was detected for marker RM 480, whereas the lowest (0.35) was detected for marker RM 1370 with a mean of 0.56.

Population genetic diversity

The population genetic diversity of the rice genotypes based on SSR markers under study is given in Supplementary Table S3. Analysis of all the 90 SSR markers exhibited that all the SSR markers were polymorphic and highly effective in discriminating against all the accessions used in the present study. The markers detected a total of 237 alleles across all the genotypes and the number of alleles ranges from 2 to 4 per locus with an average of 2.6 alleles per locus, which indicates that rice accession used in the present study was diverse. The highest number of alleles (four alleles) was detected for RM430, RM340, RM440, RM515, RM1370 and RM1281. The marker RM 1370 showed the highest gene diversity or expected heterozygosity (He) of 0.795, followed by 0.679 for RM 15078 and 0.673 for RM 005. The He ranged from 0.795 for RM 1370 to 0.254 for RM 480 with a mean of 0.531. The observed heterozygosity (Ho) varied from 0 to 0.071 for RM 515, with an average of 0.006 across all the 90 loci. The Ho was zero for most of the markers, indicating that genotypes used in the present study were genetically pure and completely homozygous and this may be attributed to the self-pollinated nature of the reproduction of rice. The observed heterozygosity (0.006) was far lower than the total expected heterozygosity (0.574). The mean value of the gene flow (Nm) was recorded at 12.412. Allelic counts presented in Fig. 4 a indicate that it was higher than expected in Population (Pop) 10, 8 and 4; almost equal to expected in Populations 2 and 3, and less than expected in all others, although the level varied. The test of significance (Fig. 4 b) showed that the difference in allelic counts (allelic diversity) was significant ($p = 0.001$) between Pop10 and Pop9, Pop9 and Pop8, Pop9 and Pop6, Pop9 and Pop 1-4. The differences between Pop 7 and Pop6 ($p = < 0.001$), Pop 7 and Pop3 ($p = 0.01$), Pop5 and Pop4 and Pop3 and Pop4 ($p = 0.05$) were also significant.

Population structure and genetic relationships

A Bayesian clustering model-based approach using admixture and K values ranging from 1 to 10 with 10 iterations was used to examine the best population stratification for 130 rice landraces. The number of clusters (K) was plotted against the "rate of change of likelihood" (K), which revealed a strong peak at K = 2 to determine the ideal K-value (Fig. 5). The population can be separated into two subpopulations, P1 and P2, with the optimal K-value indicating the highest likelihood of population clustering (Fig. 6). Based on the proportion of the genomic area shared by various subpopulations, the genotypes were divided into pure and admixture groups. The genotypes attributed to the relevant subpopulation as pure were those with a population membership fraction of less than 80%, while the others were classified as admixtures. About 52 genotypes made up the subpopulation P1, of which 48 (92.3%) were pure and 4 (7.69%) were admixtures. There were 78 genotypes in the subpopulation P2, of which 58 (74.35%) were pure and 20 (25.64%) were admixtured.

For grain yield, the landraces Ongshou, Meche, Vepvu Tsuk, and Tsuksa were the P1 subpopulation's best genotypes. The P2 subpopulation's short-duration (less

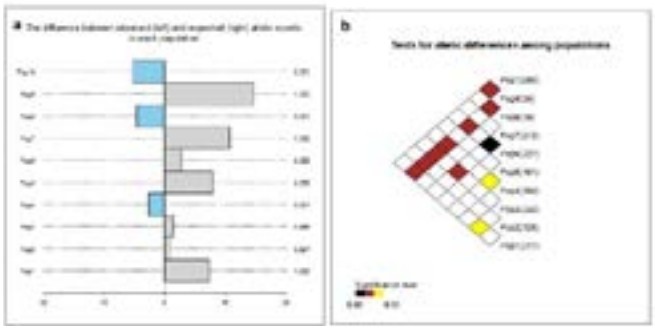

Fig. 4. (a) The difference between the observed (left) and expected (right) allelic counts in each population and (b) Test of allelic differences among populations

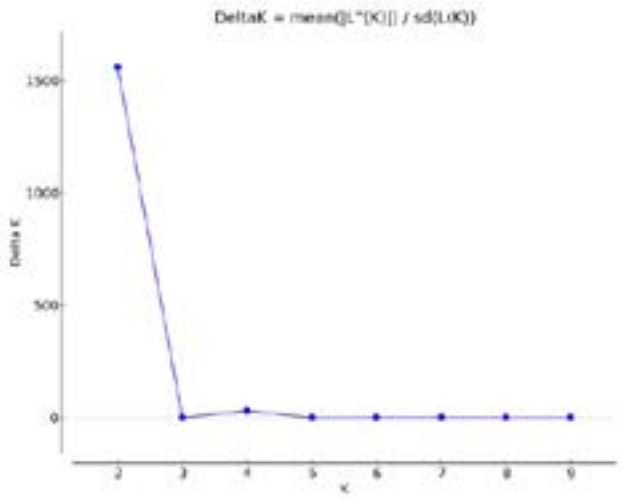

Fig. 5. A plot of delta K values from the Structure analyses of 114 rice accessions, obtained through Structure harvester ver. 0.6

than 100 days) landraces of rice are the Yunghah Hakla and Aongsho varieties. The softer gel consistency was observed in the P2 genotypes Maisa Tsuk, Phoural utlou 252, Nyari, KBA Stem, Chakhao Senapati, Ketki-u, Kemenya, and Yamuk. Long panicles are a hallmark of the few landraces of the P1 subpopulation Deihou, Mapok Tsuk, Chachak Hou, and Epyo Tsuk Longsa. Breeding programs for the genetic improvement of yield and grain quality features can use landraces with high yield potential and desired grain qualities. Based on the yield and grain quality variables examined in the current study, the population was not substantially categorized.

Using the STRUCTURE software, fixation indices FST, FIS, and FIT were computed to measure the pattern of population subdivisions. Significant divergence within both subpopulations was suggested by the FST values of 0.276 and 0.107 for subpopulations P1 and P2, respectively, with an average of 0.192. The mean alpha value in the model-based study was 0.1566. This genetic divergence could be

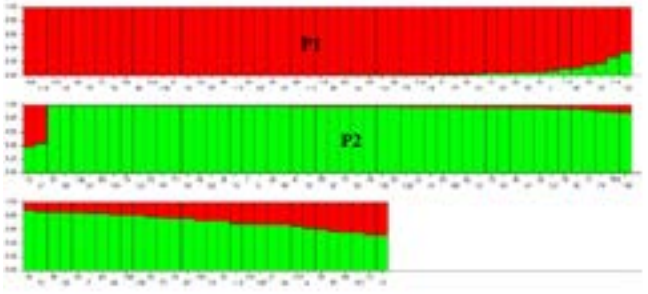

Fig. 6. Population structure of 114 rice accession based on 65 SSR markers. Note: Numbering of genotypes corresponds to the serial number in Supplementary Table 1

the result of human and natural selection favoring a diverse set of alleles in different ecologies as well as distinctions in geography and habitat, which result in low allele sharing between different subpopulations (Nakamichi et al. 2022).

The genotypes' genetic diversity was also examined at the subpopulation level in terms of the allele counts (Na), effective allele counts (Ne), observed heterozygosity (Ho), gene diversity (He), unbiased anticipated heterozygosity (uHe), and fixation index (F). In comparison to subpopulation P1, subpopulation P2 showed greater gene diversity. The mean anticipated heterozygosity in both subpopulations was greater than the mean observed heterozygosity. The Mean fixation index of both subpopulations, which ranged from 0.981 to 0.990, supported this.

Analysis of Molecular Variance (AMOVA)

The two subpopulations identified by model-based STRUCTURE analysis were subjected to the calculation of AMOVA and genetic diversity indices for detecting differentiation in the subpopulations. AMOVA revealed that 13% of the total variation was among populations, 86% among individuals within populations and 1% within individuals (Table 2). Wright's F statistic, genetic differentiation or fixation index ($F_{ST)}$ was 0.126, whereas inbreeding coefficient (F_{IS}) and F_{IT} were 0.986 and 0.988, respectively (Table 2). A high value of F_{IS} indicated a lack of heterozygosity and high distinctness of subpopulations due to the highly self-pollinated nature of the rice. In model-based analysis, F_{ST} for subpopulation P2 was 0.107 which indicates low differentiation that is further supported by a high gene flow value of 1.742 in AMOVA analysis. Nm value of less than one indicates limited gene flow among subpopulations and more than one indicates high gene flow (Wright 1965).

Neighbor-joining based clustering

An unweighted neighbor-joining tree based on the genotypic data generated by the 90 SSR markers explained the genetic relationship among the 130 rice genotypes (Fig. 7). The rice genotypes were separated into two groups along with the admixed genotypes spreading over the

Table 2. AMOVA of 130 rice germplasm

Source	df	SS	MS	Est. Var.	%
Among Pops	1	485.132	485.132	3.499	13%
Among Individuals	128	6199.603	48.434	24.052	86%
Within Individuals	130	43.000	0.331	0.331	1%
Total	259	6727.735		27.882	100%
F-Statistics	Value	P(rand >= data)			
F_{ST}	0.126	0.001			
F_{IS}	0.986	0.001			
F_{IT}	0.988	0.001			
Nm	1.742				

Table 3. Significant marker-trait association for various grain quality and yield related traits

Trait	Marker Name	Chromosome	F value	P value	R^2
Days to flowering	RM101	12	17.4665	5.48E-05	0.0472
Decorticated grain width	RM515	8	5.2488	6.22E-04	0.0355
Days to maturity	RM210	8	7.5080	8.39E-04	0.0456
Days to maturity	RM105	9	13.0216	4.45E-04	0.0398
Gel consistency	RM240	2	7.6777	1.51E-05	0.1192
Gelatinization temperature	RM112	2	13.0405	4.41E-04	0.0372
Plant height	RM1256	3	4.7851	2.09E-04	0.0958

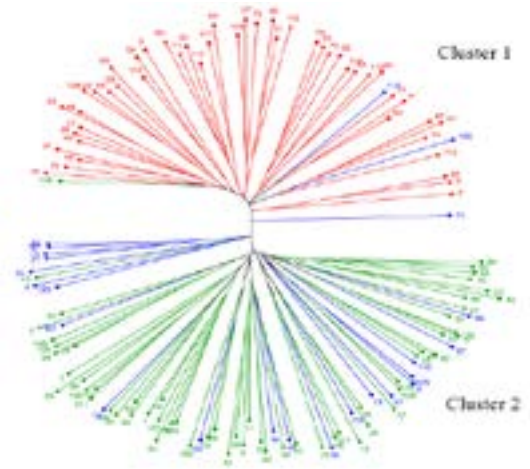

Fig. 7. Unrooted neighbour-joining tree of 130 rice genotypes using SSR markers

two groups by cluster analysis based on the unweighted neighbor-joining clustering method. Venn diagram was used to compare the genetic relationship or clustering results of the model-based analysis with the unweighted neighbour-joining clustering method.

Principle coordinate analysis

PCoA, using SSR allelic data, assessed the genetic relationship among the rice genotypes. The first two axes of differentiation explained 26.95% of the total variation. The first coordinate explained 20.02% of the variation and the second coordinate explained 6.93% of the variation. The results of the PCoA revealed two groups that correspond to model-based subdivisions of the 130 rice genotypes. A high correspondence was observed for the grouping of germplasm using model-based STRUCTURE analysis with PCoA and unweighted neighbor-joining clustering, which further confirms the population STRUCTURE analysis results. A close correspondence was recorded in the results of model-based STRUCTURE analysis related to genetic diversity and fixation indices with AMOVA and FST analysis. These results revealed that the population under study has high genetic diversity and moderate population structure.

Marker trait association

Marker trait association analysis using the MLM model (mixed linear model) based on the Q matrix generated in STRUCTURE and kinship matrix of TASSEL revealed a total of seven associations at $p < 0.0001$ for grain quality and yield attributing traits with R^2 ranging from 3.55 to 11.91% under upland ecosystem without type 1 error (Table 3). The marker RM240, located on chromosome 2 was associated with gel consistency, explaining 11.91 % of the variation, whereas marker RM112, located on the same chromosome,

is associated with GT and explained 3.17% of the variation. The marker RM1256, located on chromosome 3 is associated with plant height, explaining 9.58% of the variation. Two associations were detected for maturity duration on chromosomes 8 and 9 by RM210 and RM105, respectively and, explaining 4.55 and 3.98% of the variation, respectively. The marker RM101 located on chromosome 12 is associated with days to flowering, explaining 4.72% of the variance. The marker RM515, located on chromosome 8 is associated with decorticated grain width, explaining 3.55% of the variation.

Discussion

This study focused on tall, medium-mature NE Indian hill accessions. Farmers here chose taller plants to limit weed competition and early to medium-duration landraces to avoid chaffy grains. A recent study found low amylose levels and bold long and short-grain morphologies in numerous rice landraces. The majority of Nagaland and Manipur rice landraces have low-amylose content with long and short bold grain shape morphologies (Roy et al. 2015). The study employed a K-means clustering approach to identify structures within the population that exhibited the greatest discrimination. Additionally, the resulting differentiation pattern was compared with the genotypic origins. In this study's population, the Elbow method identified three clusters with minimal overlap in the K-means clustering, indicating variability within the analyzed population and complete dissimilarity between the three formed groups. Bollinedi et al. (2020), employing K-means Cluster analysis and, classified 190 Indian rice landraces into two groups based on grain physico-chemical traits and three groups based on nutritional traits. The deliberate cultivation of these landraces in conditions of isolation or through rigorous natural and human selection has resulted in their unique characteristics and wide-ranging diversity (Tiwari et al. 2020). Cluster I characterized by medium duration, low yielding with preferred grain quality traits suggesting that genotypes belonging to this cluster could serve as valuable resources for improvement of grain quality. Cluster-II genotypes featured early to medium maturity and, lengthy panicles and high grains per panicle. These characteristics suggest that the genotypes belonging to cluster II may serve as valuable resources for the development of medium-duration HYVs.Cluster III landraces are characterized by tall, early to medium-duration, high-yield landraces. The Tsuksa landrace of this cluster matures in 112 days and has the best production potential of all genotypes (4.77 tons ha^{-1}) followed by Ongshou (4.54 tons ha^{-1}) with medium maturity. This finding suggests that these genotypes hold potential for utilization in breeding programs aimed at enhancing rice yield.

The principal component analysis conducted on 130 rice germplasm populations indicated that the highest level of diversity was attributed to various factors, including maturity duration, date of 50% flowering, grain width and length, plant height, panicle length, spikelet fertility, and yield. Tiwari et al. (2020) reported comparable outcomes in rice restorer lines. This analysis showed that the rice landraces used in the current study had a sufficient amount of diversity, and the best genotypes from distinctive divergent clusters could be used in the hybridization program to produce desirable segregants with desirable yield potential and grain quality traits.

The polymorphism information content (PIC) metric measures a marker's information and polymorphism identification ability. Thus, genetic research relies on the PIC value of genetic markers to identify and choose them (Serrote et al. 2020). The present investigation revealed that the SSR markers exhibited substantial polymorphic information content (PIC) values, which ranged from 0.19 (RM 480) to 0.80 (RM 5891), with an average value of 0.52. This shows the marker loci had high allelic diversity and were evenly dispersed among the population tested. Several SSR markers in this analysis had PIC values above 0.60, indicating that they were informative. SSR markers' multi-allelic properties explain the result. SSR polymorphism analysis showed that the markers were polymorphic and could distinguish the accessions used in this study. Das et al. (2013) found 0.575 PIC in northeast Indian rice accessions. The current investigation found similar PIC values. Umakanth et al. (2017) found 0.44 PIC in 232 northeast Indian rice landraces. Anupam et al. (2017) reported a PIC value of 0.47 for rice landraces in Tripura. Anandan et al. (2016) documented a value of 0.240 for the PIC in a collection of 85 Assam rice specimens, encompassing both indica and japonica rice germplasm. The present study's markers were deemed informative and useful, as evidenced by the findings. The quantification of genetic variation has been approached through the utilization of the total number of alleles at a given locus, which is considered a crucial metric for assessing the evolutionary potential of populations over extended periods. Nevertheless, it is important to note that the number of alleles is subject to a significant limitation, as it is heavily influenced by the size of the sample, unlike heterozygosity, which is not similarly affected. The current investigation identified 237 alleles through the use of 90 polymorphic markers among a sample size of 130 individuals. The allelic count ranged from 2 to 4 per locus, with an average of 2.6 alleles per locus. Umakanth et al. (2017) observed an average of 3 alleles per locus in North East Indian rice landraces, with a range of 2 to 8 alleles. In rice samples from India, Southeast Asia, and America, Nachimuthu et al. (2015) found 2–7 alleles, averaging 3 per locus. Das et al. (2013) found 4.9 alleles per locus in Northeast Indian rice landraces, while Zhang et al. (2011) found 3.88 among 150 South Asian and Brazilian rice types. According to Suvi et al. (2020), Asian and African rice genotypes have 2

to 20 alleles, averaging 7.43 per locus. Aljumaili et al. (2018) found 4.09 alleles per locus in Malaysian rice, ranging from 2 to 7. Roy et al. (2016) found a higher frequency of alleles (2–21, average 8.49) in rice accessions from Arunachal Pradesh landraces, basmati rice, Meghalayan aromatic rice, aus, *japonica*, and *indica* rice. Bhuvaneswari et al. (2020) found 2 to 7 alleles in fragrant rice from Manipur, with an average of 3.5 alleles per marker. The observed disparity in the number of alleles identified could potentially be attributed to the utilization of genetically heterogeneous material and variable DNA markers in their investigation.

The gene diversity observed in this study is consistent with the results reported by Umakanth et al. (2017), wherein a mean gene diversity of 0.51 was observed in a sample of 232 rice genotypes from various regions, including northeastern landraces, indica varieties, basmati, temperate japonica, and tropical japonica rice genotypes. Nachimuthu et al. (2015) observed 0.52 gene diversity in 192 rice accessions from India, Asia, and America. Unlike Singh et al. (2016) and Anandan et al. (2016), the current research identified significant gene diversity. Singh et al. (2016) discovered 0.33 gene diversity in 729 rice varieties, while Anandan et al. (2016) found 0.30 in 426 ARC accessions, 25 tropical japonica, 57 indica landraces, and 127 breeding lines. We genotyped 409 rice accessions from 79 countries using SSR markers. The genetic diversity was 0.68, higher than Liakat et al. (2011). According to Garris et al. (2005) and Ni et al. (2002), most rice global diversity panels have gene diversity between 0.45 and 0.7. These findings on international accessions of indica, tropical japonica, temperate japonica, and wild relatives help to conclude that a panel of 130 rice accessions collected from North East India represents a sizeable portion of the genetic diversity that has been cultivated in the main rice-growing Asian continent.

The assessment of gene pool diversity is more informative when the focus is on overall allelic diversity rather than allele frequency, using markers like SSRs, which demonstrate co-dominance. As suggested by various studies (e.g., Fu et al. (2003) in oats and Russel et al. (2000) in barley), average genetic diversity (allelic counts in our study) is less sensitive to breeding or directional selection than allelic diversity at specific loci. This study found that average genetic diversity is constant among populations based on allelic counts. This is predicted for a traditionally cultivated species with low selection pressure. In five groups in our investigation, allelic numbers remained predicted or higher than expected. Overall, allelic counts indicate that all these populations can adjust to a wide range of environmental variables.

For additional marker-trait association analyses or genome-wide association research, an understanding of the population structure is essential. The current investigation utilized a model-based Bayesian clustering algorithm to determine the population structure. The results indicated that the peak value of ΔK was observed at K = 2, which is a crucial finding for understanding genetic diversity. Out of a total of 130 genotypes, 106 were identified as pure, while the remaining 24 rice accessions were classified as admixtures. The presence of admixture individuals may be attributed to gene flow between distinct rice genotypes facilitated by natural cross-pollination, which is likely to occur given the proximity of farmers' rice fields. On occasion, agriculturalists cultivate rice landraces that frequently comprise a blend of genotypes. According to Alemu et al. (2020) admixtures may be caused by informal seed systems that promote regional and national seed exchange among farming communities. The study found a low alpha value (α =0.1566), indicating a restricted number of admixed people. Small alpha values suggest that the individuals being studied belong to different populations, while high alpha values indicate that a considerable fraction of population accessions are admixed. The current investigation involved an analysis of population structure, which resulted in the classification of rice landraces into two distinct subpopulations, namely P1 and P2. This classification was based on a threshold value of greater than 80%. The categorization of 232 Northeast Indian rice varieties (Umakanth et al. 2017), 64 Northeast Himalayan rice cultivars (Roy et al. 2016), and 192 rice landraces from India, Southeast Asia, and America resulted in similar findings. Using SSR markers, Choudhury et al. (2023) divided Indian rice landraces into four subpopulations. Bhuvaneswari et al. (2020) detected three subpopulations in 93 Manipur aromatic rice accessions. The outcomes of various classification methodologies based on genotypic data, including model-based analysis, neighbor-joining clustering, and PCoA, exhibited a substantial level of consistency. This consistency implies the existence of genuine genetic variations among the genotypes examined at the DNA level. Furthermore, the findings suggest that the rice accessions scrutinized in this study can be segregated into two primary subpopulations.

The AMOVA analysis indicated a greater proportion of variation within individuals and a lesser proportion of variation among populations. A greater degree of variability among individuals within the rice population has been documented earlier by many workers (Malik et al. 2022; Choudhury et al. 2021; Singh et al. 2016). The possible explanation for this phenomenon could be attributed to the acquisition of landraces from diverse ecological areas in Nagaland and Manipur, encompassing the cultivation of rice on wetland terraces and various upland jhum scenarios. An intra-individual variance of 1% was detected, suggesting a high level of germplasm purity. The existence of genetic variability among individuals in a population contributes to genetic differentiation, thereby reinforcing population divergence.

Nachimuthu et al. (2015) proposed that a high FIT score in

rice indicates a heterozygosity shortage due to inbreeding. Wright (1978) introduced F-statistics to characterize genetic diversity within and between populations. Wright's (1978) concept classifies subpopulations by FST. Values above 0.25 indicate significant subpopulation differentiation, while 0.15 to 0.25 suggests great differentiation. FST values between 0.05 and 0.15 indicate moderate genetic differentiation, and values below 0.05 imply negligible differentiation. The present study identified a noteworthy distinction between two subpopulations based on the FST values obtained from the model-based analysis. This differentiation can be attributed to the diverse ecological and topographical conditions under which the genotypes were collected. The AMOVA results revealed a moderate genetic differentiation between the two subpopulations, as evidenced by the FST value of 0.126. A value of Nm less than one is indicative of restricted gene exchange among subpopulations. However, in the current investigation, the Nm value was recorded as 1.742, which is relatively high. This suggests that there may be a significant genetic exchange or gene flow occurring between subpopulations, as noted by Eltaher et al. (2018). As a result, there is moderate genetic differentiation between the subpopulations. The high Nm value observed may be attributed to various factors such as limited cross-pollination over an extended period, the introduction of landraces from one region to another by farmers, the cultivation of rice crops near each other, and the practice of seed mixing by some farmers before sowing. The findings of these studies indicate that the rice landraces analyzed in this study demonstrate significant genetic diversity, which may prove advantageous for initiatives aimed at enhancing crop quality.

The analysis of allelic patterns and genetic diversity indices yielded valuable insights into the genetic diversity present within subpopulations. Both subpopulations displayed discernible genetic diversity; however, subpopulation 2 exhibited a relatively higher degree of genetic diversity. The comprehension of the genetic variability present in the rice genotypes of Nagaland and Manipur could potentially aid in the development of effective strategies for preserving genetic diversity within breeding programs. The results of the PCoA analysis demonstrated the diversity and distinctiveness of the populations under investigation. Specifically, the first two principal coordinates accounted for 20.02 and 6.93% of the total variation, respectively. Nachimuthu et al. (2015) have all reported a comparable molecular variance pattern in rice. Choudhury et al. (2014) reported increased genetic variation as per PCoA and identified three population subgroups in the northeast rice germplasm.

The association mapping of yield and yield-related traits revealed seven significant marker-trait associations (MTA) with R2 ranging from 0.035 to 0.119 using MLM analysis in the present study. The R2 value can indeed be important as strong evidence of high marker-trait association, especially in the case of genetic studies, such as GWAS. The *p-value* determines the association of QTLs with markers and the R2 value predicts the magnitude of the QTL effects (Ashfaq et al. 2023). Similarly, Kaldate et al. (2023) reported 23 QTLs for yield and related traits with LOD (Logarithm of odds) values ranging between 2.50 and 7.83 and R2 values of 2.95–12.42% in rice. Zhang et al. (2014) reported 65 marker-trait associations having R2 value less than 0.10 for grain quality and yield contributing traits using 150 Ting's rice core collection and 274 SSR markers. An SSR marker, RM240, was associated with gel consistency on chromosome 2, explaining 11.9% of the overall variance. In the DH population of the *indica* cultivar TN1 and the typical *japonica* cultivar CJ06, Su et al. (2011) showed a correlation between RM240 and gel consistency. This finding supports the notion that RM240 is indeed associated with GC. In their study, Swamy et al. (2012) identified a QTL denoted as GC2.1 located on chromosome 2, which was associated with the trait of GC. Wang et al. (2017) have identified the presence of QGC2 on chromosome 2, which has an impact on the trait of GC. In contrast to other studies, Ramchander et al. (2021) found a correlation between RM240 and days to flowering, while Sharma et al. (2021) reported an association with days to maturity. Zhang et al. (2020b) observed a relationship between RM240 and grain yield, and Leng et al. (2014) reported an association between protein content and grain hardness. Thus, it is imperative to obtain additional verification using a high-density marker in the vicinity of this locus to establish the pleiotropic impact or co-localization of markers for GC on chromosome 2. The co-occurrence or pleiotropy of markers associated with diverse attributes related to grain quality and yield presents an avenue for investigating the enhancement of said traits. The genetic locus situated on chromosome 3, namely RM1256, exhibits a correlation with the trait of plant height and accounts for 9.58% of the observed variability in the phenotype of plant height. Previous studies have also confirmed the existence of genes responsible for plant height, which are primarily located on chromosomes 12, 1, 3, and 4 (Yang et al. 2022; Sitoe et al. 2022; Shearman et al. 2022). Han et al. (2017) have documented a noteworthy correlation between this characteristic and chromosome 3. Thus, the current investigation validates the existence of the gene responsible for plant height located on chromosome 3. Donde et al. (2020) have documented a correlation between RM1256 and the ratio of seed length to breadth in rice.

The genetic marker RM101, situated on chromosome 12, exhibited a significant correlation with days to 50% flowering trait. Prince et al. (2015) discovered that this is associated with the occurrence of leaf rolling under moisture stress. Lei et al. (2013) documented a correlation between tolerance to rice blast disease. The present study has identified a novel

QTL for days to 50% flowering on chromosome 4, which is significant due to the lack of available information on any QTL for this trait in this particular region. Additional examination is necessary to authenticate the current results.

The markers RM210 and RM105 were associated with maturity duration on chromosomes 8 and 9, respectively. Previous studies have identified a correlation between RM210 and various traits, including grain weight (Xie et al. 2006), susceptibility to rice bacterial leaf blast (Hasan et al. 2015), spikelet fertility, and heat susceptibility index (Prasanth et al. 2016). Talukdar et al. (2017) have documented a correlation between RM105 and the number of grains per panicle in aromatic rice from Assam. The current study has identified a novel QTL for days to maturity, as information is scarce regarding the presence of QTL for this trait on chromosomes 8 and 9. Additional examination is necessary to authenticate the current discoveries.

The genetic marker RM515, located on chromosome 8, has been found to exhibit a significant correlation with decorticated grain width. This marker accounts for 3.55% of the overall variance observed in the trait under consideration. The existence of QTL related to grain width on chromosome 8 has also been documented by others (Aslam et al. 2022; Yang et al. 2021). The RM515 marker is located at 20.3MB on chromosome 8 as per primer blast, NCBI, Aslam et al. (2022) reported a qGW8 QTL on chromosome 8 for the same trait in between 24.21-26.84MB. Fahliani et al. (2011) documented a correlation between RM515 and the characteristics of grain length and grain shape in rice. Hashemi et al. (2015) have discovered a correlation between the RM515 marker and the 2-acetyl-1-pyrroline (2AP) content in rice. Thus, there is a need to obtain additional verification using a high-density marker in the vicinity of this locus to establish the pleiotropic impact or co-localization of markers for DGW on chromosome 2. The co-occurrence or pleiotropy of markers associated with diverse attributes related to grain quality and yield provides the opportunity for improvement of these traits.

Supplementary material

Supplementary Tables S1 to S3 can be accessed at www.isgpb.org

Authors' contribution

Conceptualization of research (HV, SPD, AK), Designing of the experiments (SPD, AK, HV), Contribution of experimental materials (HV, K S, LD), Execution of field/lab experiments and data collection (HV, ST, CA, DR, LKB, HK), Analysis of data and interpretation (RNS, HV, PPB), Preparation of manuscript (HV, ST, RNS).

Acknowledgments

We thank the Indian Council of Agricultural Research, Govt. of India, for providing funds to carry out this work under the Institute project.

References

Alemu A., Feyissa T., Letta T. and Abeyo B. 2020. Genetic diversity and population structure analysis based on the high-density SNP markers in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). BMC Genetics, **21**: 1-12.

Aljumaili JS., Rafii MY., Latif MA., Sakimin SZ., Arolu IW. and Miah G. 2018. Genetic diversity of aromatic rice germplasm revealed by SSR markers. BioMed Research International.

Anandan A., Anumalla M., Pradhan SK. and Ali J. 2016. Population structure, diversity and trait association analysis in rice (*Oryza sativa* L.) germplasm for early seedling vigor (ESV) using trait-linked SSR markers. PloS One, **11**(3): e0152406.

Anupam A., Imam J., Quatadah SM., Siddaiah A., Das SP., Variar M. and Mandal NP. 2017. Genetic diversity analysis of rice germplasm in Tripura State of Northeast India using drought and blast-linked markers. Rice Sci.., **24**(1): 10-20.

Aslam K., Naveed SA., Sabar M., Shabir G., Shah SM., Khan AR. and Arif M. 2022. Identification of QTLs for rice grain size and weight by high-throughput SNP markers in the IR64 x Sadri population. Front. Genet.., **13**: 955347.

Aznan A., Viejo CG., Pang A. and Fuentes S. 2003. Review of technology advances to assess rice quality traits and consumer perception. Food Res. Inter., **172**: 113105.

Berdugo C.J., Valbuena R.I., Sanchez B.E., Barrero L.S. and Yockteng R. 2017. Genetic diversity and association mapping in the Colombian Central Collection of *Solanum tuberosum* L. Andigenum group using SNPs markers. PloS One, **12**(3): e0173039.

Bhuvaneswari S., Gopala K.S., Bollinedi H., Saha S., Ellur R.K., Vinod K.K. and Singh A.K. 2020. Genetic architecture and anthocyanin profiling of aromatic rice from Manipur reveals divergence of Chakhao landraces. Front. Genet., **11**: 570731.

Bollinedi H., Vinod K.K., Bisht K., Chauhan A., Krishnan S.G., Bhowmick P.K. and Singh A.K. 2020. Characterising the diversity of grain nutritional and physico-chemical quality in Indian rice landraces by multivariate genetic analyses. Indian J. Genet.Plant Breed., **80**(01): 26-38.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y. and Buckler E.S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, **23**: 2633-2635.

Choudhury D.R., Kumar R., Maurya A., Semwal D.P., Rathi R.S., Gautam R.K. and Singh R. 2023. SSR and SNP marker-based investigation of Indian rice landraces in relation to their genetic diversity, population structure, and geographical isolation. Agriculture, **13**(4): 823.

Choudhury D.R., Kumar R., Singh K., Singh N.K. and Singh R. 2021. Identification of a diverse core set panel of rice from the East Coast region of India using SNP markers. Front. Genet., **12**: 726152.

Choudhury D.R., Singh N., Singh A.K., Kumar S., Srinivasan K., Tyagi R.K. and Singh R. 2014. Analysis of genetic diversity and population structure of rice germplasm from the northeastern region of India and development of a core germplasm set. PLoS One, **9**(11): e113094.

Custodio M.C., Cuevas R.P., Ynion J., Laborte A.G., Velasco M.L. and Demont M. 2019. Rice quality: How is it defined by consumers, industry, food scientists, and geneticists? Trends Food Sci. Technol., **92**: 122-137.

Das B., Sengupta S., Parida S.K., Roy B., Ghosh M., Prasad M. and Ghose T.K. 2013. Genetic diversity and population structure of rice landraces from Eastern and North Eastern States of India. BMC Genetics, **14**: 1-14.

Dellaporta S.L., Wood J. and Hicks J.B. 1983. A plant DNA minipreparation: version II. Plant Mol. Biol. Reporter, **1**: 19-21.

Donde R., Mohapatra S., Baksh S.Y., Padhy B., Mukherjee M., Roy S. and Dash S.K. 2020. Identification of QTLs for high grain yield and component traits in new plant types of rice. PLoS One, **15**(7): e0227785.

Earl D.A. 2012. VonHoldt BM STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources., **4**: 359-361.

Eltaher S., Sallam A., Belamkar V., Emara H.A., Nower A.A., Salem K.F. and Baenziger P.S. 2018. Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. Front. Genet., **9**: 76.

Evanno G., Regnaut S. and Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol., **14**(8): 2611-2620.

Fahliani R.A., Khodambashi M., Houshmand S., Arzani A. and Sorkheh K. 2011. Heritability for some agronomic characters of rice (*Oryza sativa* L.) and their linked microsatellites identification. Turkish J. Agric. Forestry, **35**(5): 481-490.

Falush D., Stephens M. and Pritchard J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol.Notes, **7**(4): 574-578.

FAOSTAT. 2018. FAOSTAT database collections.

FAOSTAT. 2022. Food and Agriculture Organization of the United Nations. Rome: FAO. http://www.fao.org/faostat/en/#data

Fu Y.B., Peterson G.W., Scoles G., Rossnagel B., Schoen D.J. and Richards K.W. 2003. Allelic diversity changes in 96 Canadian oat cultivars released from 1886–2001. Crop Sci., **43**: 1989–1995.

Galili T. 2015. Dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics, **31**(22): 3718-3720.

Garris A.J., Tai T.H., Coburn J., Kresovich S. and McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics, **169**(3): 1631-1638.

Han Z., Hu W., Tan C. and Xing Y. 2017. QTLs for heading date and plant height under multiple environments in rice. Genetica, **145**.

Hasan M.M., Rafii M.Y., Ismail M.R., Mahmood M., Rahim H.A., Alam M.A. and Latif M.A. 2015. Marker-assisted backcrossing: a useful method for rice improvement. Biotechnol. Biotechnol. Equip., **29**(2): 237-254.

Hashemi F.S.G., Rafii M.Y., Ismail M.R., Mohamed M.T.M., Rahim H.A., Latif M.A. and Aslani F. 2015 Opportunities of marker-assisted selection for rice fragrance through marker–trait association analysis of microsatellites and gene-based markers. Plant Biol.., **17**(5): 953-961.

Hori K. and Sun J. 2022. Rice grain size and quality. Rice., **15**(1): 33.

Juliano BO. and Villareal CP. 1993. Grain quality evaluation of world rices. International Rice Research Inst.

Kaiser H.F. 1974. An index of factorial simplicity. Psychometrika., **39**(1): 31–36.

Kaldate R., Verma R.K., Chetia S.K., Dey P.C., Mahalle M.D., Singh S.K., Baruah A.R. and Modi M.K. 2023. Mapping of QTLs associated with yield and related traits under reproductive stage drought stress in rice using SNP linkage map. Mol. Biol. Rep., **50**(8): 6349-6359.

Kassambara A. and Mundt F. 2020. Factoextra: Extract and visualize the results of multivariate data analyses. R Package, Version 1.0.7.

Le S., Josse J. and Husson F. 2008. FactoMineR: an R package for multivariate analysis. Journal of Statistical Software., **25**: 1-18.

Lei C., Hao K., Yang Y., Ma J., Wang S., Wang J. and Wan J. 2013. Identification and fine mapping of two blast resistance genes in rice cultivar 93-11. The Crop J.., **1**(1): 2-14.

Leng Y., Xue D., Yang Y., Hu S., Su Y., Huang L. and Zeng D. 2014. Mapping of QTLs for eating and cooking quality-related traits in rice (*Oryza sativa* L). Euphytica, **197**: 99-108.

Little R.R., Hilder G.B. and Dawson E.H 1958. Differential effect of dilute alkali on 25 varieties of milled white rice. Cereal Chem., **35**: 111-126.

Malik P., Huang M., Neelam K., Bhatia D, Kaur R., Yadav B. and Singh K. 2022. Genotyping-by-sequencing based investigation of population structure and genome wide association studies for seven agronomically important traits in a set of 346 *Oryza rufipogon* accessions. Rice, **15**(1): 1-23.

Nachimuthu V.V., Muthurajan R., Duraialaguraja S., Sivakami R., Pandian B.A., Ponniah G. and 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**(1): 1-25.

Nakamichi R., Kitada S. and Kishino H. 2022. Exploratory analysis of multi-trait coadaptations in light of population history. Ecol. Evol., **12**(3): e8755.

Ni.J., Colowit P.M. and Mackill D.J. 2002. Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci., **42**(2): 601-607.

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

Pattanayak A., Ingrai B., Khongwir D.E.A., Gatpoh E.M., Das A. and Chrungoo N.K. 2018. Diversity analysis of rice bean (*Vigna umbellata* (Thunb.) Ohwi and Ohashi)) collections from North Eastern India using morpho-agronomic traits. Sci Hort., **242**: 170-180.

Peakall R. and Smouse P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics, **28**(19): 2537-9.

Perrier X. and Jacquemoud-Collet J.P. 2006. DARwin software: Dissimilarity analysis and representation for Windows. http://darwin.cirad.fr/darwin

PPV & FRA. 2007. Guidelines for the conduct of test for distinctiveness, uniformity and stability on (*Oryza sativa* L.). Government of India.

Prasanth V.V., Basava K.R., Babu M.S., V.G.N V.T., Devi S.R., Mangrauthia S.K. and Sarla N. 2016. Field level evaluation of rice introgression lines for heat tolerance and validation of markers linked to spikelet fertility. Physiol. Mol. Biol. Plants, **22**: 179-192.

Prince S.J., Beena R., Gomez S.M., Senthivel S. and Babu R.C. 2015. Mapping consistent rice (*Oryza sativa* L) yield QTLs under drought stress in target rainfed environments. Rice, **8**(1): 1-13.

Pritchard J.K., Stephens M. and Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**(2): 945-59.

Ramchander S., Leon M.A.P., Souframanien J. and Arumugam P.M. 2021. Genetic diversity, allelic variation and marker trait associations in gamma irradiated mutants of rice (*Oryza sativa* L.). Int. J. Radiation Biol., **98**(1): 90-99.

Rincent R., Moreau L., Monod H., Kuhn E., Melchinger A., Malvar R.A., MorenoGonzalez J., Nicolas S., Madur D., Combes V., Dumas F., Altmann T., Brunel D., Ouzunova M., Flament P., Dubreuil P., Charcosset A. and MaryHuard T. 2014. Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genetics, **197**: 375-387.

Roy S., Banerjee A., Mawkhlieng B., Misra AK., Pattanayak A., Harish G.D. and Bansal K.C. 2015. Genetic diversity and population structure in aromatic and quality rice (*Oryza sativa* L.) landraces from North-Eastern India. PloS One, **10**(6): e0129607.

Roy S., Marndi B.C., Mawkhlieng B., Banerjee A., Yadav R.M., Misra A.K. and Bansal K.C. 2016. Genetic diversity and structure in hill rice (*Oryza sativa* L.) landraces from the North-Eastern Himalayas of India. BMC Genet., **17**(1): 1-15.

Roy S.C. and Shil P. 2020. Assessment of genetic heritability in rice breeding lines based on morphological traits and caryopsis ultrastructure. Scientific Reports, **10**(1): 1-17.

Russell J.R., Ellis R.P., Thomas W.T.B., Waugh R., Provan J., Booth A., Fuller J., Lawrence P., Young G. and Powell W. 2000. A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Molecular Breeding., **6**: 553–568.

Serrote C.M.L., Reiniger L.R.S., Silva K.B., dos Santos Rabaiolli S.M. and Stefanel C.M. 2020. Determining the polymorphism information content of a molecular marker. Gene, **726**: 144175.

Sharma M., Gangurde S.S., Salgotra R.K., Kumar B., Singh A.K. and Pandey M.K. 2021. Genetic mapping for grain quality and yield-attributed traits in Basmati rice using SSR-based genetic map. J. Biosci., **46**(3): 50.

Shearman J.R., Vejchasarn P., Naktang C., Phansenee Y., Jomchai N., Lanceras-Siangliw J. and Toojinda T. 2022. Rice height QTLs in KDML105 chromosome segment substitution lines. Genomics, **114**(1): 482-487.

Singh N., Choudhury D.R., Singh A.K., Kumar S., Srinivasan K., Tyagi R.K. and Singh R. 2013. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PloS One, **8**(12): e84136.

Singh N., Choudhury D.R., Tiwari G., Singh A.K., Kumar S., Srinivasan K. and Singh R. 2016. Genetic diversity trend in Indian rice varieties: an analysis using SSR markers. BMC Genet., **17**: 1-13.

Sitoe H.M., Zhang Y., Chen S., Li Y., Ali M., Sowadan O. and Hong D. 2022. Detection of QTLs for plant height architecture traits in rice (*Oryza sativa* L) by association mapping and the RSTEP-LRT method. Plants, **11**(7): 999.

Su Y., Rao Y., Hu S., Yang Y., Gao Z., Zhang G. and Zeng D. 2011. Map-based cloning proves qGC-6, a major QTL for gel consistency of japonica/indica cross, responds by Waxy in rice (*Oryza sativa* L.). Theor. Appl. Genet., **123**: 859-867.

Sultana S., Faruque M. and Islam M.R. 2022. Rice grain quality parameters and determination tools: A review on the current developments and future prospects. Int. J. Food Properties, **25**(1): 1063-1078.

Sun J., Yang L., Wang J., Liu H., Zheng H., Xie D. and Zou D. 2018. Identification of a cold-tolerant locus in rice (*Oryza sativa* L.) using bulked segregant analysis with a next-generation sequencing strategy. Rice, **11**: 1-12.

Suvi W.T., Shimelis H., Laing M., Mathew I. and Shayanowako A.I.T. 2020. Assessment of the genetic diversity and population structure of rice genotypes using SSR markers. Acta Agriculturae Scandinavica. Soil Plant Sci., **70**(1): 76-86.

Swamy B.M., Kaladhar K., Shobha Rani N., Prasad G.S.V., Viraktamath B.C., Reddy G.A. and Sarla N. 2012. QTL analysis for grain quality traits in 2 BC2F2 populations derived from crosses between *Oryza sativa* cv Swarna and 2 accessions of *O. nivara*. J. Hered., **103**(3): 442-452.

Talukdar P.R., Rathi S., Pathak K., Chetia S.K. and Sarma R.N. 2017. Population structure and marker-trait association in indigenous aromatic rice. Rice Sci., **24**(3): 145-154.

Tarang A., Kordrostami M., Shahdi K.A., Hosseini C.M., Forghani S.A., Ghanbarzadeh M. and Sattari M. 2020. Study of genetic diversity in rice (*Oryza sativa* L) cultivars of Central and Western Asia using microsatellite markers tightly linked to important quality and yield-related traits. Genetic Resour. Crop Evol., **67**: 1537-1550.

Tiwari S., Singh Y., Upadhyay P.K. and Koutu G.K. 2020. Principal component analysis and genetic divergence studies for yield and quality-related attributes of rice restorer lines. Indian J. Genet. Plant Breed., **82**(01): 94-98.

Umakanth B., Vishalakshi B., Sathish Kumar P., Rama Devi S.J.S., Bhadana V.P., Senguttuvel P. and Madhav M.S. 2017. Diverse rice landraces of North-East India enable the identification of novel genetic resources for Magnaporthe resistance. Front. Plant Sci., **8**: 1500.

Verma H., Borah J.L. and Sarma R.N. 2019. Variability assessment for root and drought tolerance traits and genetic diversity analysis of rice germplasm using SSR markers. Scientific Reports, **9**(1): 1-19.

Wang H., Zhu S., Dang X., Liu E., Hu H., Eltahawy M.S., Zaid I.U. and Hong D. 2019. Favorable alleles mining for gelatinization temperature, gel consistency and amylose content in *Oryza sativa* by association mapping. BMC Genet., **20**: 34.

Wang L., Xu J., Wang H., Chen T., You E., Bian H. and Shen Y. 2023. Population structure analysis and genome-wide association study of a hexaploid oat landrace and cultivar collection. Front.Plant Sci., **14**.

Wang X., Pang Y., Zhang J., Wu Z., Chen K., Ali J. and Li Z. 2017. Genome-wide and gene-based association mapping for rice eating and cooking characteristics and protein content. Scientific Reports, **7**(1): 17203.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution, **395**: 420.

Wright S. 1978. The relation of livestock breeding to theories of evolution. J. Animal Sci., **46**(5): 1192-1200.

Xie X., Song M.H., Jin F., Ahn S.N., Suh J.P., Hwang H.G. and McCouch S.R. 2006. Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*. Theor. Appl. Genet., **113**: 885-894.

Yang H., Yang Q., Kang Y., Zhang M., Zhan X., Cao L. and Zhang Y. 2022. Finding Stable QTL for plant height in super hybrid

rice. Agriculture, **12**(2): 165.

Yang W., Xiong L., Liang J., Hao Q., Luan X., Tan Q. and Zhang G. 2021. Substitution mapping of two closely linked QTLs on chromosome 8 controlling grain chalkiness in rice. Rice, **14**(1): 1-11.

Zhang A., Gao Y., Li Y., Ruan B., Yang S., Liu C. and Qian Q. 2020a. Genetic analysis for cooking and eating quality of super rice and fine mapping of a novel locus qGC10 for gel consistency. Front. Plant Sci., **342**.

Zhang H., Zhu YJ., Zhu A.D., Fan Y.Y., Huang T.X., Zhang J.F. and Zhuang J.Y. 2020b. Identification and verification of quantitative trait loci affecting milling yield of rice. Agronomy, **10**(1): 75.

Zhang P., Li J., Li X., Liu X., Zhao X. and Lu Y. 2011. Population structure and genetic diversity in a rice core collection (*Oryza sativa L*.) investigated with SSR markers. PloS One, **6**(12): e27565.

Zhang P., Liu X., Tong H., Lu Y. and Li J. 2014. Association mapping for important agronomic traits in core collection of rice (*Oryza sativa L*.) with SSR markers. PLoS One, **9**(10): e111508.

Zhang Z., Ersoz E., Lai C.Q., Todhunter R.J., Tiwari H.K., Gore M.A., Bradbury J., Yu J., Arnett D.K., Ordovas J.M. and Buckler E. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genet., **42**: 355–360.

Supplementary Table S1. A list of experimental material and their location of collection

S. No.	Genotype Name	Location	Altitude (mamsl)	Altitud-inal group
1.	Akhan	Longleng, Nagaland	1100	2
2.	Akuk Moro	Wokha Village, Wokha	1300	2
3.	Angja	Tangha Village, Longleng	1100	2
4.	Aongsho	Tangha Village, Longleng	1100	2
5.	Asai	Imphal	786	1
6.	Bali Red	Dimapur	145	1
7.	Bhalum – I	Umiam	1000	1
8.	Boyoh	Dungkhao Village, Longleng	1100	2
9.	Chachak Hou	Longleng, Nagaland	1100	2
10.	Chakha	Longleng, Nagaland	1100	2
11.	Chakhao Poreiton	THOUBAL, MANIPUR	765	1
12.	Chakhao Senepati	SENAPATI, MANIPUR	1061	1
13.	Chali Yoh	Longleng, Nagaland	1100	2
14.	Chamnya Yoh	Dungkhao Village, Longleng	1100	2
15.	Ching Phouren Amubi	Imphal	786	1
16.	Chingphou Angouba	Imphal	786	1
17.	Chingtui Mah	Longleng, Nagaland	1100	2
18.	Chulietyio	Sanis Village, Wokha	1300	2
19.	Chupa Wing Rice	WAKHA, NAGALAND	1300	2
20.	Daramphou (258)	Imphal	786	1
21.	Deihou	Longleng, Nagaland	1100	2
22.	Dhaha	Longleng, Nagaland	1100	2
23.	Doiha	Longleng, Nagaland	1100	2
24.	Doulong	Hukpang village, Longleng	1100	2
25.	Engcha Yoh	Dungkhao Village, Longleng	1100	2
26.	Epyo Tsuk	Wokha Village, Wokha	1300	2
27.	Epyo tsuk longsa	Longsachung Village, Wokha	1300	2
28.	Ereima	Imphal	786	1
29.	Goyo Tsuk	Longkhum Village, Mokokchung	1325	2
30.	GP/K/10 Mema Katwa	UKHRUL,MANIPUR	1662	3
31.	Hah Shou	Nyang Village, Longleng	1100	2
32.	Hahnyak	Nyang Village, Longleng	1100	2
33.	Hukha Tssok	Nyang Village, Longleng	1100	2
34.	Hyung Yoh	Ukhrul	1662	3
35.	Jakho	Hukpang village, Longleng	1100	2
36.	Jaksa	Hukpang village, Longleng	1100	2
37.	Juchok	Phek, Nagaland	1525	3
38.	KBA Stem	Imphal	786	1
39.	KD - 5-2-8	MANIPUR	786	1
40.	KD - 5-3-14 (Manui Nira)	MANIPUR	786	1

41.	Kemenya	Kohima	1450	3
42.	Ketki-u	Kohima	1450	3
43.	Khemaru	Kohima	1450	3
44.	Konpemo Tsuk	Wokha Village, Wokha	1300	2
45.	Koyapvu Tsuk	Wokha, Nagaland	1300	2
46.	Kuza Shu	Porba, Nagaland	1525	3
47.	Lamjet	Hukpang village, Longleng	1100	2
48.	Laza	WAKHA, NAGALAND	1300	2
49.	Leimaphou	Imphal	786	1
50.	Lisem Tsuk	Longjang Village, Mokokchung,	1325	2
51.	Mahhak	Longleng, Nagaland	1100	2
52.	Maibo	Tangha Village, Longleng	1100	2
53.	Maisa Tsuk	Longjang Village, Mokokchung	1325	2
54.	Makhara	Imphal	786	1
55.	Malanken	Wokha Village, Wokha	1300	2
56.	Maloki	Longsachung Village, Wokha	1300	2
57.	Malpiri	Wokha Village, Wokha	1300	2
58.	Manabe	Phek, Nagaland	1525	3
59.	Manen Tsuk	Dibuia Village, Mokokchung	1325	2
60.	Manen Tsuk	Longjang Village, Mokokchung	1325	2
61.	Mapok	Longjang Village, Mokokchung	1325	2
62.	Mapok Tsuk	Mopungchuket Village, Mokokchung	1325	2
63.	Mapok Temesungla	Mopungchuket Village, Mokokchung	1325	2
64.	Maring (166)	Zunheboto	1850	4
65.	Maso Tsuk	Dibuia Village, Mokokchung	1325	2
66.	Mati pasi	Sanis Village, Wokha	1300	2
67.	Meba Peko	Hukpang village, Longleng	1100	2
68.	Meche	Sanis Zero Point, Sanis Village, Wokha	1300	2
69.	Meitak N- special	PEREN, NAGALAND	684	1
70.	Mekhrilha Kecha	Kohima, Nagaland	1450	3
71.	Mepongchuket Masu	Khensa Village, Mokokchung	1325	2
72.	Mesa	Longsachung Village, Wokha	1300	2
73.	Mesurong Tsuk	Mopungchuket Village, Mokokchung	1325	2
74.	Mipin	Zunheboto	1850	4
75.	Moro Etyo	Wokha Village, Wokha	1300	2
76.	Motiro	Sanis Village, Wokha	1300	2
77.	Motso Tsuk wokha	Wokha Village, Wokha	1300	2
78.	Moya Maso	Longkhum Village, Mokokchung	1325	2
79.	Moya Tsuk	Longkhum Village, Mokokchung	1325	2
80.	Nailong Mapok	Longjang Village, Mokokchung	1325	2
81.	Nam Yoh	Dungkhao Village, Longleng	1100	2

82.	Nari Chitpi	Imphal	786	1
83.	Neikedo-u lha Tsia	KOHIMA	1450	3
84.	Niphuthukpi (288)	Imphal	786	1
85.	Nukjan Nyakla	Tangha Village, Longleng	1100	2
86.	Nukjan Shola	Tangha Village, Longleng	1100	2
87.	Nuknyei	Nyang Village, Longleng	1100	2
88.	Nunkhumvu	Sanis Village, Wokha	1300	2
89.	Nyakmoh Yoh	Dungkhao Village, Longleng	1100	2
90.	Nyari	Kohima, Nagaland	1450	3
91.	Ongshou	Dungkhao Village, Longleng	1100	2
92.	Pfutsero Ru	Pfutsero	2133	4
93.	Phoremubi	MANIPUR	786	1
94.	Phoural Utlou 252	MANIPUR	786	1
95.	Pulu lha	Wokha Village, Wokha	1300	2
96.	Pumpha Mha	UKHRUL, MANIPUR	1662	3
97.	Ronayang 147	Imphal	786	1
98.	Rukhatang	Wokha Village, Wokha	1300	2
99.	Samaro	Sanis Village, Wokha	1300	2
100.	Samro Yoh	Sanis Zero Point, Sanis Village, Wokha	1300	2
101.	Semmeki	Khensa Village, Mokokchung	1325	2
102.	Senebumab	Imphal	786	1
103.	Shangha	Hukpang village, Longleng	1100	2
104.	Shangshak Local	UKHRUL, MANIPUR	1662	3
105.	Shangya	Nyang Village, Longleng	1100	2
106.	Shopvu	Sanis Village, Wokha	1300	2
107.	Shuphok	Nyang Village, Longleng	1100	2
108.	Tei Yoh	Hukpang village, Longleng	1100	2
109.	Teiri	Hukpang village, Longleng	1100	2
110.	Teke	Mokokchung	1325	2
111.	Thevuru	Kikruma, Phek	1525	3
112.	Toiya	Hukpang village, Longleng	1100	2
113.	Tsonyko	Wokha Village, Wokha	1300	2
114.	Tsuk Meren	Longsachung Village, Wokha	1300	2
115.	Tsuk Nakla	Longjang Village, Mokokchung	1525	3
116.	Tsuksa	Dibuia Village, Mokokchung	1525	3
117.	Tsuksemla	Dibuia Village, Mokokchung	1525	3
118.	Vam	Tangha Village, Longleng	1100	2
119.	Vepvu Tsok	Sanis Village, Wokha	1300	2
120.	Vepvu Tsuk	Wokha Village, Wokha	1300	2
121.	Wazruho Phek	NAGALAND	1525	3
122.	Yamchinga	Hukpang village, Longleng	1100	2

123.	Yamuk (M C -26-6-2-3)	Imphal	786	1
124.	Yimso Tsuk	Mopungchuket Village, Mokokchung	1325	2
125.	Yinching	Longleng, Nagaland	1100	2
126.	Yohjak	Hukpang village, Longleng	1100	2
127.	Yunghah	Tangha Village, Longleng	1100	2
128.	Yunghah Hakla	Nyang Village, Longleng	1100	2
129.	Yunghah Shula	Nyang Village, Longleng	1100	2
130.	Zu Tsuk	Wokha Village, Wokha	1300	2

Supplementary Table S2. Mean performance of 130 genotypes for grain quality, yield and yield attributing traits.

S. No.	Genotype Name	DF	DM	PH	ETN	PL	GPP	SF	YPP	GL	GW	DGL	DGW	AC	GT	GC
1.	Akhan	94.3	126.8	116.9	4.6	22.4	67.8	82.7	2.5	8.7	2.5	5.9	2.2	6.1	5.8	32.0
2.	Akuk Moro	96.8	110.8	104.3	3.6	23.6	172.4	81.6	3.1	8.3	2.5	5.6	2.3	15.1	1.9	56.7
3.	Angja	77.0	119.8	173.9	7.2	28.8	144.1	86.1	2.6	5.6	2.4	4.2	2.0	16.3	1.1	56.8
4.	Aongsho	74.5	98.2	167.3	7.0	27.6	150.6	96.3	4.8	5.8	3.1	4.2	2.8	13.5	1.1	48.8
5.	Asai	93.3	124.3	125.0	7.9	24.3	147.2	87.7	3.6	8.7	3.2	6.6	3.2	13.5	3.1	73.7
6.	Bali Red	89.0	129.5	156.5	8.0	28.5	142.7	84.8	3.4	7.5	2.5	6.6	2.7	18.4	6.5	29.2
7.	Bhalum – I	79.7	121.0	144.4	8.5	23.2	128.7	91.0	4.2	7.7	2.4	5.5	2.3	15.4	3.0	25.8
8.	Boyoh	85.0	112.3	150.3	7.3	26.4	132.4	86.0	3.8	6.2	3.2	4.3	2.8	7.4	2.0	30.3
9.	Chachak Hou	77.7	118.2	140.1	7.1	29.2	135.2	90.0	3.4	6.4	2.4	4.5	2.0	5.7	1.1	28.8
10.	Chakha	76.5	119.5	165.7	6.8	30.3	113.0	80.1	3.2	6.2	2.3	4.6	2.2	13.9	1.1	30.2
11.	Chakhao Poreiton	96.0	136.2	149.8	8.1	23.2	155.6	87.2	1.7	7.9	2.9	5.9	2.7	14.7	5.8	31.2
12.	Chakhao Senepati	97.7	136.3	93.8	4.5	20.0	133.0	87.1	3.8	7.2	3.1	4.5	2.4	5.8	3.8	111.5
13.	Chali Yoh	87.3	118.8	182.1	7.8	25.6	131.3	92.7	4.1	7.6	3.3	5.3	2.5	6.2	1.2	28.8
14.	Chamnya Yoh	83.2	107.5	159.6	8.5	27.9	153.9	68.2	1.3	6.8	2.4	4.8	2.0	15.4	1.1	40.0
15.	Ching Phouren Amubi	99.7	136.0	163.1	7.3	25.6	139.4	83.3	1.2	7.8	2.6	6.6	2.3	16.3	5.8	65.8
16.	Chingphou Angouba	95.8	129.3	128.6	4.7	22.9	149.0	90.2	2.5	8.3	3.5	5.8	3.0	16.0	5.9	30.7
17.	Chingtui Mah	96.5	131.5	123.8	6.6	25.1	78.5	74.6	4.0	8.3	2.9	6.0	2.1	14.6	6.1	46.5
18.	Chulietyio	88.8	114.8	155.9	5.0	26.0	138.5	87.7	2.8	9.5	3.1	6.5	2.5	6.8	2.1	53.7
19.	Chupa Wing Rice	96.5	133.5	107.4	5.3	21.0	102.3	90.0	3.7	8.4	2.9	6.7	2.6	17.2	6.1	92.0
20.	Daramphou (258)	96.5	139.7	142.8	4.6	21.6	105.1	94.0	2.3	8.2	2.5	6.0	2.1	7.5	5.0	31.5
21.	Deihou	88.8	119.7	171.0	8.4	31.3	138.6	85.0	3.0	8.3	2.9	6.8	2.6	14.5	2.2	29.8
22.	Dhaha	76.8	119.5	173.3	8.4	28.6	142.6	65.3	1.6	6.9	2.6	4.5	1.9	7.0	1.1	31.7
23.	Doiha	86.2	119.0	160.8	7.8	26.6	105.5	79.6	3.6	5.8	3.3	4.3	2.8	2.2	2.0	30.8
24.	Doulong	86.7	121.3	168.3	7.0	23.6	117.3	85.3	3.7	7.5	3.3	5.9	3.0	16.4	2.2	51.7
25.	Engcha Yoh	79.5	106.7	138.0	7.1	29.2	168.1	86.7	2.9	6.8	2.5	4.7	2.0	13.5	1.1	54.3
26.	Epyo Tsuk	85.5	108.7	148.5	8.8	26.2	151.5	83.2	3.5	8.5	2.8	6.9	2.6	5.0	1.2	72.8
27.	Epyo tsuk longsa	92.8	118.8	139.6	5.3	28.8	133.1	88.9	2.9	9.8	3.4	7.0	2.6	5.2	5.0	37.3
28.	Ereima	98.3	124.2	94.5	3.7	28.1	121.6	76.7	3.4	9.5	2.3	6.4	1.9	16.4	6.2	54.3
29.	Goyo Tsuk	83.2	105.0	147.5	6.9	27.7	158.9	78.2	4.2	9.4	2.9	6.8	2.7	13.4	2.0	52.7
30.	GP/K/10 Mema Katwa	98.5	131.7	157.2	7.9	28.3	156.3	78.1	2.2	6.9	2.7	6.0	2.2	7.2	1.2	31.8
31.	Hah Shou	89.0	120.0	116.3	5.1	22.6	141.7	74.4	3.4	5.3	3.1	4.5	2.6	16.4	1.9	113.8
32.	Hahnyak	78.2	118.5	150.9	6.8	28.1	137.8	82.4	3.7	6.4	2.5	4.6	2.1	7.9	1.1	70.0
33.	Hukha Tssok	87.7	121.3	144.0	8.1	22.3	140.8	82.0	2.3	8.9	3.0	6.4	2.6	17.0	4.5	31.7
34.	Hyung Yoh	87.3	119.7	185.6	6.8	28.8	131.8	83.8	3.4	5.5	1.9	5.6	3.0	16.1	1.9	80.0
35.	Jakho	81.7	122.5	164.8	6.7	26.1	130.1	67.4	1.4	6.2	2.8	4.6	2.0	14.6	1.1	86.3
36.	Jaksa	95.0	129.3	116.7	5.6	23.0	87.0	73.6	1.0	8.2	3.0	6.3	2.5	16.8	4.8	65.8
37.	Juchok	96.7	121.8	133.8	5.9	27.5	126.3	88.3	2.1	7.0	3.6	5.6	3.2	7.7	4.2	70.2
38.	KBA Stem	87.7	126.7	149.5	7.4	26.2	143.9	93.5	1.1	7.6	2.5	5.5	2.7	15.2	5.9	113.3
39.	KD - 5-2-8	96.7	131.5	107.2	5.0	22.8	69.9	88.3	4.0	6.7	2.5	5.9	2.4	28.0	6.8	108.8

40.	KD - 5-3-14 (Manui Nira)	97.5	130.2	167.4	7.0	24.6	133.2	83.9	2.6	6.7	2.8	5.5	2.6	28.8	3.0	90.3
41.	Kemenya	99.7	137.8	123.8	5.8	19.5	111.1	92.5	2.7	9.2	3.9	6.3	3.3	7.1	3.0	110.7
42.	Ketki-u	85.5	119.3	128.6	8.1	20.7	134.7	88.7	3.3	8.6	3.0	5.9	2.5	6.4	5.8	110.8
43.	Khemaru	98.3	121.5	128.7	5.8	28.3	118.1	75.6	3.6	9.5	3.2	6.6	2.6	15.4	2.0	31.7
44.	Konpemo Tsuk	87.0	113.5	143.7	4.9	25.8	124.3	88.7	3.9	9.0	3.1	6.8	2.8	15.6	2.0	46.3
45.	Koyapvu Tsuk	87.0	120.7	179.1	5.6	26.6	146.3	93.0	3.9	8.3	2.8	6.0	2.0	16.3	3.8	47.8
46.	Kuza Shu	96.8	135.2	130.2	6.7	25.5	116.0	79.1	3.5	7.8	2.8	5.8	2.6	16.4	2.0	30.5
47.	Lamjet		121.3	137.2	5.1	25.2	136.1	81.8	1.6	6.7	2.3	5.5	2.5	2.5	1.1	96.3
48.	Laza	93.7	121.3	124.5	7.0	27.1	108.2	85.8	3.8	8.2	2.4	5.2	2.2	22.9	2.0	92.7
49.	Leimaphou	96.3	134.0	151.8	8.8	27.5	121.1	93.6	2.2	9.6	2.7	6.4	2.4	2.0	6.0	29.8
50.	Lisem Tsuk	86.8	114.7	159.8	4.9	25.2	140.2	95.3	3.1	5.6	3.2	4.3	2.8	18.1	2.0	118.2
51.	Mahhak	79.8	119.7	160.5	6.0	28.6	117.8	79.1	3.7	6.9	2.6	5.2	2.3	27.3	2.0	79.2
52.	Maibo	74.5	116.0	172.2	7.7	27.6	137.6	87.9	3.7	6.3	2.4	4.9	2.0	18.0	1.1	51.7
53.	Maisa Tsuk	75.5	111.2	138.5	6.2	25.3	165.7	55.4	3.2	5.8	2.4	4.8	2.3	28.4	2.0	115.2
54.	Makhara	89.0	123.0	63.2	6.2	26.5	131.4	89.5	3.9	8.9	2.5	6.3	2.3	23.5	5.8	26.5
55.	Malanken	94.8	111.8	137.6	4.8	27.2	135.8	85.5	3.5	8.5	3.2	6.0	2.7	14.5	2.0	55.2
56.	Maloki	90.0	115.5	151.4	6.6	26.3	127.2	84.0	3.8	7.7	4.3	6.1	2.9	14.2	2.0	48.0
57.	Malpiri	80.3	109.7	127.9	3.6	19.6	88.7	53.1	1.1	9.9	3.2	6.9	2.2	2.2	1.1	68.3
58.	Manabe	86.2	119.0	108.8	4.1	23.2	188.1	82.6	2.4	9.3	3.3	6.7	2.3	7.7	4.0	30.3
59.	Manen Tsuk	76.3	107.2	161.8	7.7	28.3	135.9	82.5	2.9	6.3	2.3	4.9	2.1	16.2	1.1	50.2
60.	Manen Tsuk	80.2	105.5	161.6	6.3	27.1	136.6	90.3	3.3	6.0	2.4	4.5	2.2	14.3	1.1	45.7
61.	Mapok	81.7	108.0	139.8	6.7	26.3	128.8	84.3	2.7	7.4	3.0	6.1	3.0	5.1	2.0	53.8
62.	Mapok Tsuk	96.7	116.2	139.0	4.4	30.5	157.1	80.5	3.9	10.2	3.4	7.0	2.9	7.2	4.2	87.5
63.	Mapok Temesungla	86.5	113.3	143.0	8.6	24.1	140.7	82.7	4.0	7.1	2.9	5.2	2.6	2.1	5.0	32.5
64.	Maring (166)	87.0	127.7	152.9	5.9	23.6	83.4	87.9	2.1	10.1	2.5	6.3	2.2	15.6	2.0	32.0
65.	Maso Tsuk	84.0	106.7	136.5	8.3	27.6	119.6	91.2	4.0	8.4	3.1	5.6	2.6	7.2	4.2	96.8
66.	Mati pasi	87.2	113.7	160.4	5.0	26.7	109.3	92.0	4.0	8.5	3.0	6.5	3.0	6.2	1.1	48.0
67.	Meba Peko	95.0	121.5	164.9	8.0	26.1	155.7	88.1	2.2	7.4	3.3	5.8	3.0	7.8	2.0	48.2
68.	Meche	89.0	111.5	148.1	5.3	25.1	125.0	88.9	4.4	6.9	3.3	5.3	3.0	6.8	1.1	34.8
69.	Meitak N- special	97.3	135.0	95.0	5.4	20.1	73.2	76.7	1.9	8.4	3.1	6.0	2.7	7.5	2.1	37.8
70.	Mekhrilha Kecha	97.8	132.5	127.6	8.5	24.3	133.0	86.6	3.2	7.9	3.0	6.8	2.5	16.4	5.8	50.7
71.	Mepongchuket Masu	89.5	110.3	162.0	8.0	26.6	133.7	94.4	3.5	8.0	3.0	5.5	2.6	1.6	4.2	57.0
72.	Mesa	80.3	114.8	172.8	5.6	26.6	136.9	94.7	3.3	7.4	2.7	5.1	2.3	16.1	1.1	53.5
73.	Mesurong Tsuk	79.3	106.7	146.9	7.1	28.8	119.5	87.2	2.1	6.6	2.3	4.4	2.2	7.6	1.1	56.7
74.	Mipin	98.2	126.0	96.4	4.9	22.9	89.0	76.0	1.3	8.0	2.5	6.7	2.4	6.9	2.8	37.7
75.	Moro Etyo	78.7	111.8	140.3	4.3	28.5	143.2	86.3	2.7	8.7	2.4	5.4	3.0	7.6	5.2	82.5
76.	Motiro	95.5	112.5	141.4	5.8	22.6	136.9	88.3	2.2	6.8	3.2	5.9	2.7	7.2	4.7	83.3
77.	Motso Tsuk wokha	79.8	110.3	165.2	8.4	28.1	118.0	73.7	3.5	9.9	2.5	6.2	2.7	15.7	2.0	37.5
78.	Moya Maso	85.3	111.8	149.8	4.5	27.6	143.6	86.9	4.2	8.5	2.8	6.7	2.9	7.3	2.0	48.7
79.	Moya Tsuk	88.8	114.0	158.0	5.5	28.8	136.9	93.9	3.8	7.9	3.2	6.8	3.2	15.3	2.0	66.5

80.	Nailong Mapok	83.5	107.0	154.6	7.6	27.3	105.7	85.8	3.9	7.8	3.3	6.5	2.9	5.3	1.1	38.2
81.	Nam Yoh	77.0	120.3	170.9	6.5	28.9	157.2	94.3	1.3	6.7	2.6	5.0	2.2	7.6	1.1	38.0
82.	Nari Chitpi	97.0	132.8	88.2	7.9	27.1	134.6	85.9	3.6	8.1	2.9	5.7	2.8	17.7	5.0	52.5
83.	Neikedo-u lha Tsia	82.0	123.3	155.5	7.3	27.7	129.3	92.2	4.1	8.0	3.0	6.2	2.5	22.2	5.0	30.2
84.	Niphuthukpi (288)	90.5	121.5	125.0	9.0	18.6	67.7	80.7	1.2	7.8	2.4	6.1	2.4	15.9	6.0	53.7
85.	Nukjan Nyakla	84.3	119.0	182.4	8.8	28.7	153.1	78.9	3.6	7.8	2.6	5.7	2.5	5.6	3.0	81.2
86.	Nukjan Shola	80.5	118.2	172.4	8.0	26.6	136.2	90.3	3.0	7.5	2.5	5.9	2.2	7.9	4.0	73.5
87.	Nuknyei	96.3	120.7	149.7	7.5	24.5	109.9	77.3	3.7	7.8	2.9	5.6	2.3	6.8	5.0	104.7
88.	Nunkhumvu	89.5	109.2	134.1	7.6	25.7	142.1	94.5	2.4	7.2	4.1	5.4	3.4	7.6	3.0	76.7
89.	Nyakmoh Yoh	84.0	119.8	155.1	6.2	28.2	125.2	88.2	3.3	8.8	2.9	6.0	2.6	13.0	1.1	37.5
90.	Nyari	80.8	140.8	113.1	5.5	21.5	85.5	84.6	3.6	7.1	3.3	6.2	2.7	2.2	5.0	114.3
91.	Ongshou	87.2	122.3	150.6	5.1	26.2	109.1	86.7	4.5	6.7	2.8	5.1	2.8	13.7	2.1	38.7
92.	Pfutsero Ru	88.7	121.7	171.5	5.6	27.2	132.3	80.0	3.2	8.4	2.8	6.3	2.5	27.1	2.0	47.3
93.	Phoremubi	94.3	133.3	135.3	9.0	23.5	65.0	87.1	3.3	9.3	3.0	6.4	2.6	7.1	1.1	32.5
94.	Phoural Utlou 252	90.3	122.8	125.8	5.5	25.4	150.3	84.2	1.5	9.0	3.0	6.9	2.5	28.2	6.8	114.5
95.	Pulu lha	91.3	116.5	135.7	5.3	28.1	134.3	86.8	3.6	8.6	3.7	6.3	3.1	7.8	1.9	53.5
96.	Pumpha Mha	91.3	127.0	118.7	9.4	20.2	105.4	90.0	3.0	7.9	3.2	5.8	2.8	6.7	1.9	32.7
97.	Ronayang 147	93.3	120.3	112.4	4.5	27.5	119.8	71.4	2.8	8.6	2.9	6.7	2.6	4.7	1.1	34.8
98.	Rukhatang	91.0	117.5	151.0	6.9	26.9	99.0	80.8	2.1	8.9	3.2	5.2	3.0	15.1	2.0	71.0
99.	Samaro	95.5	110.3	128.7	6.7	26.6	160.6	93.3	3.7	8.2	3.6	6.1	3.0	8.8	2.0	31.8
100.	Samro Yoh	85.8	119.0	147.1	8.1	26.6	161.3	89.2	3.7	7.8	3.2	5.9	2.7	14.5	1.1	30.7
101.	Semmeki	94.0	113.7	143.5	4.8	25.3	162.5	90.9	4.2	8.5	3.5	6.8	2.9	8.3	3.0	49.5
102.	Senebumab	95.3	123.0	136.7	8.0	26.0	114.4	92.6	2.8	9.2	3.0	6.7	2.7	16.3	5.0	81.7
103.	Shangha	78.8	119.5	158.0	8.7	27.0	109.6	71.6	3.2	6.3	2.7	4.8	2.2	14.9	1.1	31.2
104.	Shangshak Local	97.0	134.3	121.0	5.4	24.1	102.8	86.6	3.7	6.9	2.6	5.8	2.5	17.2	6.0	45.2
105.	Shangya	86.2	118.5	142.5	5.1	26.4	148.2	79.0	1.6	6.4	2.7	4.6	2.1	13.9	5.0	51.2
106.	Shopvu	96.2	112.2	142.4	4.7	28.1	137.4	95.9	2.9	7.2	4.0	5.2	3.4	14.5	6.2	52.0
107.	Shuphok	84.3	120.3	148.6	6.9	25.3	131.2	93.1	2.8	5.9	3.4	4.1	2.8	14.1	1.1	54.5
108.	Tei Yoh	96.2	119.8	133.2	7.0	23.0	131.7	87.8	2.0	7.0	2.3	5.6	2.4	7.7	1.2	49.0
109.	Teiri	87.7	120.7	151.6	7.1	20.8	78.7	84.4	3.7	7.0	2.9	6.0	2.6	24.2	3.0	52.7
110.	Teke	82.3	114.0	136.6	6.0	25.2	113.6	81.1	4.1	7.5	2.7	5.7	2.4	6.9	2.0	29.5
111.	Thevuru	98.8	139.5	155.4	6.2	22.1	111.8	87.2	4.0	7.8	2.3	5.8	2.5	13.6	1.1	31.5
112.	Toiya	89.5	121.3	154.7	6.8	26.6	115.9	79.1	2.6	5.7	3.1	4.5	2.5	5.7	2.0	49.7
113.	Tsonyko	87.0	113.2	138.3	3.8	25.0	136.6	91.6	3.6	7.2	3.0	6.3	2.7	14.0	2.0	84.7
114.	Tsuk Meren	88.7	108.0	155.2	7.5	24.8	117.5	89.0	3.6	6.7	3.1	5.0	2.7	14.9	2.0	31.5
115.	Tsuk Nakla	85.5	107.8	149.9	6.8	23.2	120.1	93.8	4.1	5.4	2.5	4.5	2.4	29.7	2.0	80.3
116.	Tsuksa	84.5	112.2	145.6	7.7	25.7	158.0	95.3	4.8	6.2	3.3	4.2	2.8	6.6	1.8	43.3
117.	Tsuksemla	85.2	111.7	145.0	9.1	26.5	147.0	94.9	3.9	6.4	3.2	5.1	2.9	15.6	1.1	62.2
118.	Vam	89.8	118.3	166.6	7.5	20.8	139.3	86.5	4.0	7.0	3.6	5.0	3.1	1.7	1.8	109.2
119.	Vepvu Tsok	86.8	110.0	143.3	7.1	26.4	98.0	92.8	2.4	7.6	3.2	5.0	2.8	8.9	2.0	31.0
120.	Vepvu Tsuk	88.2	111.0	140.6	7.1	26.4	147.7	88.0	4.4	7.5	3.4	5.0	2.9	6.7	1.9	31.3
121.	Wazruho Phek	97.3	121.8	113.7	8.5	26.9	130.9	81.4	3.6	7.9	2.9	5.9	2.7	14.3	1.8	91.0

122.	Yamchinga	91.0	124.5	156.0	4.7	27.2	139.2	79.4	1.3	8.4	3.0	6.3	2.4	7.5	2.0	86.2
123.	Yamuk (M C -26-6-2-3)	96.3	134.7	129.3	4.9	21.9	118.5	91.1	4.0	8.7	3.2	6.1	2.5	22.6	6.8	110.3
124.	Yimso Tsuk	94.3	112.3	163.5	7.9	27.5	126.8	91.3	2.2	7.5	3.2	5.2	2.6	8.7	2.0	47.5
125.	Yinching	87.0	116.2	130.1	7.6	25.2	155.2	80.5	3.2	5.4	2.5	4.4	2.8	7.9	1.1	38.0
126.	Yohjak	82.0	121.0	146.9	5.7	28.2	153.9	88.1	2.3	6.1	2.3	4.5	2.1	5.8	1.1	57.5
127.	Yunghah	77.0	119.2	166.1	5.0	27.8	148.0	91.2	4.1	7.7	2.7	6.7	2.7	14.4	1.2	29.3
128.	Yunghah Hakla	74.3	97.5	134.1	6.9	24.7	110.4	90.7	1.4	8.7	2.6	6.5	2.2	12.6	1.2	65.7
129.	Yunghah Shula	85.0	118.8	147.8	5.7	25.9	116.8	80.7	3.2	7.8	2.9	5.8	2.7	7.9	1.2	54.3
130.	Zu Tsuk	95.3	111.7	133.8	4.5	25.5	105.6	76.9	2.5	7.6	3.6	6.0	3.0	13.4	1.1	30.8

DF = Days to flowering; DM = Days to maturity; PH = Plant height; ETN = Effective tiller number; PL = Panicle length; GPP = Grains per panicle; SF = Spikelet fertility; YPP = Yield per plant; GL Grain length; GW = Grain width; DGL = Decorticated grain length; DGW = Decorticated grain width; GT = Gelatinization temperature; GC = Gel consistency and AC = Amylose content

Supplementary Table S3. Genetic diversity of 90 SSR markers in the 130 rice genotypes

Marker	Na	Ne	Ht	He	Ho	FIS	FIT	FST	Nm	PIC
RM101	2	1.978	0.500	0.494	0.000	1.000	1.000	0.011	22.107	0.500
RM430	4	2.121	0.659	0.489	0.000	1.000	1.000	0.259	0.717	0.648
RM 112	2	1.753	0.446	0.411	0.000	1.000	1.000	0.078	2.937	0.443
RM 170	2	1.779	0.439	0.437	0.000	1.000	1.000	0.006	45.151	0.432
RM 174	2	1.992	0.499	0.498	0.000	1.000	1.000	0.002	114.334	0.499
RM 190	2	1.676	0.494	0.378	0.000	1.000	1.000	0.234	0.818	0.492
RM 210	3	2.315	0.624	0.559	0.000	1.000	1.000	0.105	2.141	0.632
RM 221	2	1.986	0.500	0.496	0.000	1.000	1.000	0.007	34.783	0.500
RM 495	2	1.651	0.469	0.349	0.000	1.000	1.000	0.257	0.722	0.473
RM 219	3	2.441	0.619	0.587	0.013	0.978	0.979	0.051	4.619	0.614
RM 231	2	1.594	0.373	0.372	0.000	1.000	1.000	0.000	813.512	0.371
RM 105	2	1.804	0.453	0.440	0.000	1.000	1.000	0.029	8.487	0.448
RM 1337	2	1.765	0.480	0.432	0.000	1.000	1.000	0.099	2.271	0.483
RM 1341	2	1.731	0.494	0.405	0.000	1.000	1.000	0.181	1.133	0.492
RM 236	3	2.075	0.538	0.512	0.000	1.000	1.000	0.048	4.940	0.519
RM 13	3	1.743	0.419	0.379	0.007	0.981	0.983	0.097	2.338	0.436
RM 144	3	2.149	0.545	0.533	0.006	0.989	0.990	0.022	11.299	0.538
RM 152	3	2.268	0.562	0.546	0.000	1.000	1.000	0.028	8.719	0.539
RM 169	2	1.423	0.297	0.297	0.000	1.000	1.000	0.001	323.476	0.301
RM 223	2	1.516	0.340	0.331	0.000	1.000	1.000	0.025	9.658	0.363
RM 253	3	2.269	0.559	0.559	0.000	1.000	1.000	0.000	1105.637	0.564
RM 267	3	2.535	0.657	0.605	0.000	1.000	1.000	0.079	2.895	0.665
RM 276	3	2.563	0.623	0.609	0.000	1.000	1.000	0.023	10.405	0.624
RM 314	3	2.109	0.550	0.519	0.000	1.000	1.000	0.055	4.267	0.558
RM 327	3	2.685	0.648	0.626	0.000	1.000	1.000	0.033	7.263	0.642
RM 3345	3	2.727	0.637	0.632	0.000	1.000	1.000	0.008	31.768	0.642
RM 340	4	2.163	0.630	0.486	0.000	1.000	1.000	0.229	0.842	0.641
RM 342	2	1.991	0.500	0.498	0.000	1.000	1.000	0.004	62.454	0.500
RM 36	3	1.987	0.509	0.495	0.000	1.000	1.000	0.027	8.991	0.515
RM 42	2	1.755	0.495	0.423	0.000	1.000	1.000	0.146	1.463	0.498
RM 421	2	1.939	0.487	0.484	0.000	1.000	1.000	0.006	39.957	0.486
RM 431	2	1.896	0.477	0.471	0.000	1.000	1.000	0.012	21.328	0.477
RM 480	2	1.371	0.268	0.254	0.000	1.000	1.000	0.053	4.497	0.189
RM 511	2	1.972	0.498	0.493	0.000	1.000	1.000	0.011	23.182	0.496
RM 1	3	2.177	0.630	0.512	0.028	0.945	0.955	0.186	1.092	0.616
RM 241	3	2.008	0.504	0.502	0.014	0.972	0.972	0.005	50.870	0.499
RM 22	2	1.835	0.484	0.452	0.000	1.000	1.000	0.064	3.636	0.490
RM 7563	2	1.634	0.480	0.343	0.000	1.000	1.000	0.286	0.625	0.483
RM 440	4	1.902	0.628	0.401	0.014	0.965	0.978	0.362	0.441	0.639
RM 219	3	2.409	0.620	0.578	0.017	0.971	0.973	0.068	3.430	0.604

Marker	Na	Ne	Ht	He	Ho	FIS	FIT	FST	Nm	PIC
RM 228	3	1.924	0.497	0.474	0.056	0.881	0.886	0.046	5.204	0.459
RM 515	4	2.256	0.658	0.556	0.071	0.873	0.893	0.155	1.367	0.613
RM 1235	2	1.918	0.494	0.477	0.000	1.000	1.000	0.034	7.157	0.486
RM 256	3	2.549	0.627	0.602	0.000	1.000	1.000	0.040	6.032	0.605
RM 530	3	2.833	0.653	0.646	0.000	1.000	1.000	0.010	23.589	0.646
RM 519	3	2.864	0.653	0.651	0.000	1.000	1.000	0.003	77.542	0.653
RM 520	2	1.978	0.500	0.494	0.000	1.000	1.000	0.011	21.630	0.500
RM 521	3	2.579	0.625	0.609	0.000	1.000	1.000	0.026	9.408	0.632
RM 523	3	2.625	0.627	0.616	0.000	1.000	1.000	0.017	14.680	0.630
RM 525	2	1.990	0.499	0.498	0.000	1.000	1.000	0.004	65.650	0.499
RM 53	2	1.995	0.499	0.499	0.007	0.986	0.986	0.001	192.857	0.492
RM 5371	2	1.976	0.494	0.494	0.000	1.000	1.000	0.001	263.795	0.494
RM 543	2	1.859	0.462	0.462	0.000	1.000	1.000	0.001	230.583	0.462
RM 5473	3	2.033	0.509	0.503	0.000	1.000	1.000	0.012	20.145	0.522
RM 6589	2	1.872	0.491	0.463	0.000	1.000	1.000	0.055	4.287	0.498
RM 7	3	2.280	0.569	0.561	0.000	1.000	1.000	0.014	17.772	0.573
RM 8231	3	2.254	0.558	0.552	0.000	1.000	1.000	0.011	23.331	0.553
RM 8264	2	1.929	0.487	0.481	0.000	1.000	1.000	0.013	19.565	0.486
RM 262	2	1.613	0.381	0.378	0.000	1.000	1.000	0.008	32.855	0.386
RM 5897	3	2.259	0.570	0.555	0.000	1.000	1.000	0.026	9.378	0.583
RM 434	2	2.760	0.662	0.635	0.000	1.000	1.000	0.040	5.968	0.496
RM 17305	3	2.417	0.598	0.573	0.000	1.000	1.000	0.042	5.703	0.370
RM 81	2	2.692	0.649	0.625	0.000	1.000	1.000	0.037	6.511	0.434
RM 248	3	2.985	0.704	0.627	0.000	1.000	1.000	0.109	2.045	0.534
RM 6887	2	2.364	0.631	0.521	0.000	1.000	1.000	0.175	1.177	0.440
RM 5748	2	2.096	0.653	0.465	0.000	1.000	1.000	0.289	0.615	0.475
RM 3286	2	2.489	0.655	0.586	0.000	1.000	1.000	0.104	2.150	0.494
RM 1370	4	3.714	0.795	0.730	0.036	0.950	0.954	0.081	2.822	0.678
RM 6838	3	2.994	0.717	0.665	0.050	0.925	0.931	0.073	3.176	0.469
RM 5891	3	2.904	0.704	0.645	0.028	0.956	0.960	0.084	2.732	0.798
RM 3773	3	2.749	0.691	0.636	0.007	0.989	0.990	0.080	2.856	0.536
RM 1341	3	2.659	0.648	0.616	0.007	0.988	0.989	0.049	4.878	0.466
RM 31	3	2.599	0.703	0.605	0.000	1.000	1.000	0.139	1.542	0.604
RM 1281	4	2.605	0.621	0.542	0.032	0.940	0.948	0.127	1.720	0.364
RM 243	3	2.587	0.641	0.604	0.000	1.000	1.000	0.057	4.132	0.463
RM 5638	3	2.305	0.578	0.549	0.011	0.979	0.980	0.049	4.835	0.323
RM 302	3	2.762	0.657	0.631	0.000	1.000	1.000	0.039	6.099	0.514
RM 005	3	3.072	0.729	0.673	0.000	1.000	1.000	0.076	3.033	0.639
RM 174	2	2.788	0.661	0.639	0.000	1.000	1.000	0.033	7.301	0.500
RM 279	3	2.745	0.681	0.629	0.006	0.991	0.992	0.075	3.064	0.529
RM 53	3	2.891	0.693	0.649	0.035	0.946	0.949	0.063	3.727	0.532

Marker	Na	Ne	Ht	He	Ho	FIS	FIT	FST	Nm	PIC
RM 240	3	2.661	0.711	0.540	0.028	0.949	0.961	0.240	0.790	0.588
RM 540	3	2.795	0.730	0.611	0.006	0.991	0.992	0.163	1.283	0.589
RM 482	2	2.277	0.603	0.538	0.000	1.000	1.000	0.107	2.089	0.427
RM 231	3	2.642	0.636	0.615	0.017	0.972	0.973	0.034	7.093	0.426
RM 1256	3	2.706	0.683	0.616	0.023	0.963	0.967	0.098	2.313	0.529
RM 1352	3	2.851	0.665	0.639	0.000	1.000	1.000	0.038	6.288	0.501
RM 15078	3	3.173	0.735	0.679	0.000	1.000	1.000	0.076	3.035	0.628
RM 15429	2	2.686	0.655	0.626	0.000	1.000	1.000	0.044	5.386	0.494
RM 15448	2	2.768	0.661	0.634	0.000	1.000	1.000	0.040	6.003	0.493
Mean	2.633	2.278	0.574	0.531	0.006	0.990	0.991	0.071	43.960	0.518

Na=Number of alleles, Ne=Number of effective alleles, Ht=Total expected heterozygosity, He=Gene Diversity, Ho=Observed Heterozygosity, FIS=inbreeding coefficient, FIT=inbreeding coefficient to total, FST=Fixation index, and Nm=Gene flow.