International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 15s,2025
https://theaspd.com/index.php

Morphological Characterisation, Genetic Diversity And Ethnobotanical Study Of Rice Landraces Of Lohit District, Arunachal Pradesh, India

Silikta Manchey[1], Dunyak Ado Boje[2], Sumpam Tangjang[3] , Tonlong Wangpan*[4]
[1,2,3]Department of Botany, Rajiv Gandhi University, Rono Hills, Doimukh-791112, Arunachal Pradesh,
[4*]DBT-APSCS&T Centre of Excellence for Bioresources and Sustainable Development, Department of Science and Technology, Kimin-791121, Arunachal Pradesh, India,
Silikta.manchey@rgu.ac.in[1], (Research Scholar)
Sumpam.tangjang@rgu.ac.in[2], (Professor)
dunyak.a@gmail.com[3], (scholar)
tonlong.wangpan@rgu.ac.in[4] (Assistant Professor)
*Corresponding Authors
Tonlong Wangpan
Email ID: tonlong.wangpan@rgu.ac.in (Assistant Professor)

Abstract

*This study investigates the agromorphological traits, genetic diversity, and ethnobotanical significance of indigenous rice landraces from the Lohit district of Arunachal Pradesh. Altogether fifteen landraces were collected from ten villages and subjected to phenotypic characterisation using standard descriptors, revealing substantial diversity in grain length, breadth, weight, presence of awn, lemma colour, and kernel characteristics. Genetic diversity was assessed using SSR markers and matK-based DNA barcoding. STRUCTURE, PCA, and PCoA analyses revealed two genetic subpopulations, with high intra-population variability and limited inter-population differentiation. AMOVA results showed 92.34% of the variance occurred within individuals. The mean pairwise genetic distance was 0.0331%, confirming low inter-landrace divergence. Ethnobotanical documentation revealed six major cultural uses of rice, including its role in traditional fermentation (P-aa and Yu-shie), ritual ceremonies (Tam-La, Tam-La-Du), and medicinal preparation (Ke-Chaie). Among the varieties, Ke-Tuilu showed the highest cultural importance index. The integration of molecular, morphological, and ethnobotanical analyses underscores the significance of these landraces as reservoirs of genetic and cultural heritage, warranting their conservation and utilisation in future rice breeding and food security programs.***

Key words: *Rice landraces, Ethnobotany, Northeast India, Mishmi tribe, Indigenous rice varieties*

INTRODUCTION

Rice (Oryza sativa L.), often referred to as the “grain of life”, sustains nearly half of the global population and has significant economic, cultural, and ecological value (Shivani et al., 2021). As a member of the Poaceae family, rice has developed into a crop of global importance, with India emerging as a prominent cultivator and exporter. India produces both basmati and non-basmati varieties and is anticipated to produce 150 million metric tons of milled rice in the 2024–2025 marketing year, representing a 9% increase from the previous year and marking the second highest production on record (Madhu et al., 2023; USDA FAS, 2025). As global demand escalates, rice breeding programs face the dual challenge of improving yield and grain quality. Grain quality is multifaceted and is influenced by intrinsic traits such as texture, taste, and colour, as well as extrinsic factors such as branding and appearance. Regional preferences vary; for instance, East Asian consumers often prefer short, sticky rice with a low amylose content, whereas consumers in tropical and subtropical regions favour long, slender, translucent grains (Lee et al., 2025). This complexity is further compounded by the fact that rice quality is governed by polygenic traits that are affected by genotype, environment, and management practices (Sultana et al., 2022).Traditional rice landraces, preserved and cultivated by indigenous farmers over generations, represent a significant repository of genetic diversity. These landraces occupy a distinct evolutionary position, serving as a bridge between wild ancestors and contemporary cultivars (Cui et al., 2019). Their inherent resilience to biotic and abiotic stresses renders them essential for future breeding initiatives focused on climate adaptation, disease resistance, and nutritional enhancement. However, a substantial portion of their potential remains unexplored, highlighting the critical need for their documentation, characterisation, and conservation (Sinha, 2015; Caldo et al., 1996).

Recent advancements in molecular biology have significantly enhanced our ability to investigate plant genetic resources. Techniques such as SSR markers, QTL mapping, GWAS, and genome sequencing have enabled the precise analysis of both monogenic and polygenic traits (Sharma et al., 2021). Molecular markers that remain unaffected by environmental factors provide dependable insights into genetic variability, structure, and gene flow. Notably, DNA barcoding using conserved chloroplast gene regions, such as matK and rbcL, has demonstrated high efficacy for species-level identification, particularly in cases where morphological distinctions are ambiguous (Aitken et al., 2024; Shadrin, 2021).Beyond their genetic significance, traditional rice varieties are essential cultural assets. In Northeast India, rice is intricately interwoven into the daily life of the populace, serving as a staple food and playing a crucial role in rituals, medicinal practices, and the production of fermented beverages. The jhum cultivation system, which remains prevalent in Arunachal Pradesh, has facilitated the preservation and diversification of these landraces (Wangpan et al., 2018). Each variety embodies narratives of adaptation, resilience, and community identity.The present study investigates the agromorphological characteristics, genetic diversity, and ethnobotanical significance of indigenous rice landraces from the Lohit district of Arunachal Pradesh. By integrating field-based morphological assessments with SSR marker analysis and matK-based DNA barcoding, alongside ethnobotanical documentation, this study aims to underscore the dual scientific and cultural value of these traditional varieties. The findings are expected to contribute to conservation strategies and support the sustainable utilisation of rice genetic resources in addressing global food security.

METHODOLOGY

Study area

Located in the northeastern region of Arunachal Pradesh, India, the Lohit district is characterised by its natural beauty (Figure 1). Tezu serves as the district's administrative centre. Geographically, the district is situated between 28°22'25.532" N and 27°34'11.765" N latitude and 95°46'15.332" E and 96°41'59.555 longitude. It shares its borders with the Namsai district of Arunachal Pradesh, Tinsukia district of Assam to the west, Anjaw district to the east, Changlang district to the south, and Lower Dibang Valley district to the north. The district presents a captivating landscape, ranging from snow-capped mountains to evergreen forests covering the lower hills of the region. The Mishmi people, who speak the Midzu branch of the Tibeto-Burman language family, constitute the district's tribal population. They are of Mongoloid descent and have a rich biocultural heritage.

Data collection

Household surveys were administered using structured questionnaires to gather data on indigenous rice-based ethnobotanical applications (Wangpan et al., 2019). The field study adhered to the methodology outlined by Jain & Mudgal, (1999). Surveys were conducted in ten villages (Tafragam, Changliang, Duraliang, Loiliang, Upper Loiliang, Paya, Hatidoba, Chailiang, Kabruliang and Yealiang) to collect the samples and document the unique traditional knowledge of indigenous rice utilisation and ethnobotanical applications. Personal interviews were conducted with village Gaon Bura and older individuals actively involved in agriculture. In each village, ten households were randomly surveyed, resulting in a total of 100 households surveyed across the district. The questionnaire covered a wide range of topics related to the ethnobotanical utility of indigenous rice, including cultural, traditional, rituals, and agricultural practices and landrace knowledge. Prior Informed Consent (PIC) of informants was obtained prior to interview, after consulting the village headmen to conduct interviews.

Processing of sample

The collected rice grains were immersed in a carbendazim fungicide solution (Pedireddi et al., 2018), thoroughly washed with warm water (30-35°C), and allowed to germinate at 30-32°C (Li and Yang, 2020). Soil was collected from same agricultural field and utilized to transplant the seeds into pots measuring 30 cm in diameter and 20 cm in height. Each rice sample was cultivated and maintained using standardized agronomic practices (Sato and Uphoff, 2007). After 18 days, leaves from each sample were harvested for DNA extraction.

DNA extraction and PCR amplification

DNA was extracted from 18-day-old leaves using the method outlined by Doyle & Doyle (1987) with minor modifications. The extracted DNA was quantified using a UV-VIS spectrophotometer. Subsequently, the DNA was amplified using SSR primers (Table 1). The PCR reaction was conducted in a total volume of 25 µl, comprising 2 µl of 50 ng/µl DNA, 1 µl of each forward and reverse primer, 8.5

µl of nuclease-free water, and 12.5 µl of 2X Green Master Mix. The mixture was subjected to PCR amplification, beginning with an initial denaturation at 94 °C for 5 min, followed by 40 cycles of amplification, including denaturation for 24 s, annealing (between 53 °C to 71 °C) for 42 s, extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. For the matK primer, the reaction volume remained consistent with that of the SSR primer, but the conditions were as follows: initial denaturation at 94 °C for 3 min, followed by 35 cycles of amplification, with denaturation for 30 s, annealing at 39 °C for 42 s, extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. The amplified product was subsequently subjected to gel visualization using 2% agarose in TAE solution stained with ethidium bromide.

Data Analysis

The phenotypic characteristics were categorized into distinct groups following the guidelines of the PPV & FRA (2007). Genetic clustering analysis was conducted using STRUCTURE software (version 2.3.4), employing the Bayesian and admixture models to examine the population structure of all landraces utilizing SSR markers. Ten iterations were performed for each K value to determine the precise number of subpopulations. The run parameters included a 10,000 burn-in period with 100,000 Markov chain Monte Carlo iterations. To ascertain the K value, the mean estimate of the log posterior probability of the data [L (K)] was plotted against the designated K value. Delta K was estimated using the Structure Selector (https://lmme.ac.cn/StructureSelector/) according to the method of Earl and VonHoldt (2012) and Evanno et al. (2005). FST, which measures genetic differentiation among populations, was based on Wright's F-statistics. GeneAlEx 6.51 b2 (Peakall and Smouse 2012) was used to estimate molecular variance components within the population, among individuals, and between populations using an analysis of molecular variance (AMOVA) with 999 permutations. DARwin 6.0.021 was used to generate Principal Coordinate Analysis (PCoA), a similarity matrix (Jacquemoud and Perrier, 2006) based on a covariance matrix. Principal Component Analysis (PCA) was performed using GraphPad Prism (version 10) to visualize the underlying patterns among the landraces and reduce the dimensionality of the multivariate dataset. The binary data matrix produced by polymorphic SSR markers was further examined using NTSYS-pc (version 2.11). DNA sequences derived from the matK were submitted to the National Center for Biotechnology Information (NCBI) to obtain accession numbers from GenBank. Additionally, these sequences were submitted to the Barcode of Life Data System (BOLD) (https://boldsystems.org/) to acquire sample and process IDs for the rice samples. The BOLD system provides representative barcodes. Multiple sequence alignments were conducted using Molecular Evolutionary Genetics Analysis (MEGA) version 11 software. A phylogenetic tree for the rice landraces was constructed, employing maximum likelihood methods to infer their evolutionary history. Genetic distance matrices for the DNA sequences were generated using the Kimura-2-parameter model implemented via ClustalW. Phylogenetic reliability was evaluated using the bootstrap method with 1,000 replicates under default settings that assumed uniform substitution rates and homogeneous patterns across them. The genetic pairwise distance was calculated using the Kimura-2-Parameter model. Parameters such as total pairwise comparison, mean pairwise comparison, minimum pairwise distance, and maximum pairwise distance were also calculated following Tamura et. al. (2021).

The ethnobotanical data were first arranged as per the requisition of the software; thereafter, the data were qualitatively analysed on ethnobotanyR software version 0.1.9. Additionally, the ethnoChord plots are made using functions from the circlise package (Gu et al. 2014).

RESULTS

Agromorphological characterisation of rice landraces

The grain length varied from 5.81 ± 0.16 mm to 10.51 ± 0.30 mm (Table 2). The longest grain, PV442135, measured 10.51 ± 0.30 mm, while the shortest, PV442145, measured 5.81 ± 0.16 mm among the rice landraces. The rice grains were categorized according to PPV & FRA (2007) into three categories: very short, short, and medium length. The medium category was the most prevalent, accounting for 66.67%, followed by short at 20% and very short at 13.33%. The grain breadth ranged from 2.60 ± 0.16 mm to 4.00 ± 0.24 mm. The broadest grain width was observed in PV442134, measuring 4.00 ± 0.24 mm, whereas the thinnest, PV442132, measured 2.60 ± 0.16 mm among all the landraces. The grain width was categorized into medium, broad, and very broad categories following PPV & FRA (2007), with the broad category being the most common at 60%, followed by medium at 26.67% and very broad at 13.33%. Subsequently, the ratio of grain length (GL) to grain breadth (GB) was calculated, and the rice landraces were categorized accordingly (PPV & FRA, 2007). The GL:GB ratio exhibited significant variation, with the highest ratio found in PV442135 (3.27 ± 0.19 mm) and the lowest in PV442145 (1.62 ± 0.21 mm).The rice landraces were categorized into three types: medium, slender, and bold types. The medium category constituted the highest percentage at 66.67%, followed by slender at 26.67% and bold at 6.67%. The heaviest grain was observed in PV442136, weighing 44.81 ± 0.70 g, whereas the lightest grain was recorded in PV442145, weighing 17.91 ± 0.68 g. The 1000-grain weights were classified into four categories: low, medium, high, and very high. The analysis revealed that low-weight grains were the most prevalent at 40%, followed by medium-weight at 33.33%, very high-weight at 20%, and high-weight at 6.67%. Kernel length varied from 3.87 ± 0.39 mm to 7.53 ± 0.75 mm, with the longest kernel length reported in PV442135 (7.53 ± 0.75 mm) and the shortest in PV442140 (3.87 ± 0.39 mm) (Table 3). According to PPV & FRA (2007), kernel length was categorized as long (33.33%), medium (33.33%), short (20%), and extra-long (13.33%). Kernel breadth ranged from 2.36 ± 0.24 mm to 3.11 ± 0.31 mm, with the broadest kernel breadth found in PV442145 (3.11 ± 0.31 mm) and the thinnest in PV442143 and PV442142, both measuring 2.36 ± 0.24 mm. The decorticated grain width was classified into medium and broad categories, with broad accounting for 86.67% and medium for 13.33% of the samples. The kernel length-to-breadth ratio exhibited significant variation, with the highest ratio being 2.92 ± 0.15 mm, and PV442140 having the smallest ratio of 1.37 ± 0.07 mm (Table 4). The rice grains were classified according to PPV & FRA (2007) into long bold and short bold categories, with long bold grains predominating at 86.67%, and short bold grains being relatively rare at 13.33%. The most common lemma color among the local rice varieties was straw (40%), followed by purple-black (20%) and gold (26.67%), with brown furrows on straw accounting for 13.33%. The remaining varieties exhibited red and brown lemmas, each accounting for 6.67%. The most common kernel colour among the local rice varieties was white, which accounts for 53.33% of the landraces. This was followed by light brown, seen in 13.33% of the varieties, and light red, which makes up 13.33%. Dark purple kernels were found in 13.33% of the varieties, while brown and red were less common, appearing in 6.67% each. Among the landraces of the Lohit district, only 13.33% possessed awns, whereas 86.67% did not. Landrace PV442135 had the longest awn at 7.81 ± 0.74 mm, whereas PV442134 had the shortest awn at 6.98 ± 0.20 mm. All other landraces were awnless.Following the characterization of morphological traits, a correlation matrix heatmap was constructed incorporating 18 agronomic traits (Figure 2). This matrix provides a comprehensive overview of the interrelationships among various grain characteristics, highlighting both strong and moderate correlations between them. GL and GB exhibited a strong positive correlation of 0.79, indicating that an increase in grain length was associated with an increase in grain breadth. This relationship suggests that grain shape is proportionally related to its overall size, implying that selection for one dimension may simultaneously affect the other. Similarly, 1000-grain weight (1000 GW) was positively correlated with GL (0.56) and kernel breadth (KB) (0.77), indicating that both the length and breadth of the grain influence the total grain weight. Grain size plays a significant role in determining the overall grain weight, reinforcing the importance of grain length and breadth. Kernel length-to-breadth ratio (KLC) has a strong positive correlation with GL (0.76) and KB (0.72), suggesting that the shape of the grain, as determined by this ratio, is closely linked to its length and breadth. Grain bulk density (GBC) also exhibited a strong positive correlation with GL (0.68) and KB (0.70), indicating that denser grains are larger and possess more robust physical characteristics. Awn length (AL) was moderately correlated with GL (0.60) and KB (0.50), whereas awn presence (AP) showed a weaker correlation with most traits but had a moderate positive correlation with GL (0.56).

The heatmap further revealed significant correlations among GL/GB, grain length-to-circumference ratio (GLC), and GLC: GBC, which were significantly correlated with KB, 1000 GW, and KLC: kernel bulk density (KBC) had a moderate positive correlation with KB (0.72) and 1000 GW. The panicle length-to-circumference ratio (PLC) and days to grain color (DGC) exhibited weaker correlations with other traits. KLC and KB were negatively correlated (-0.47), whereas 1000 GW was weakly negatively correlated with

AL (-0.23) and AP (-0.19). Overall, the heatmap revealed a complex set of positive and negative correlations, with central traits, such as 1000 GW, KLC, and GBC, influencing multiple other traits, particularly grain size and weight. In summary, the correlation matrix elucidates a complex network of relationships among traits, particularly those related to grain size, shape, and density.

The analysis identified 1000 GW, KLC, and GBC as the central traits influencing other important characteristics of rice grains. The strong correlations among these traits suggest that selecting for larger and better-shaped grains can enhance other important quality attributes such as grain weight and density. Negative correlations, such as those between GL and GL/GB and KLC and KB, indicate trade-offs that breeders must consider when selecting specific traits. This information is crucial for designing breeding strategies, as it provides a clearer understanding of how selection of one trait affects others, thereby optimizing grain quality, yield, and resilience.For all the landraces, a Bayesian clustering model-based approach was employed, with the K value ranging from 1 to 10, and ten iterations were conducted to identify the optimal stratifications of the population (Table 5). The cluster number (K) was plotted against the "rate of change of likelihood" (K), revealing a pronounced peak at K=2, which determined the ideal K-value (Figure 3 & Figure 4). Consequently, the landraces (populations) of the Lohit district can be categorized into two subpopulations: P1 and P2. These were further divided into pure and admixture groups based on the percentage of the genomic region shared by each subpopulation (Table 6). The proportion of the pure genotype was significantly higher than that of the admixture groups. Approximately 16.67% of the genotype was attributed to admixture, while 83.33% was pure in population I. Similarly, approximately 11.11% of the genotype was admixture, and 88.98% was pure in P-II. Principal Coordinate Analysis (PCoA) was used to assess the genetic relationships among the rice genotypes using SSR allelic data (Figure 5). The genetic relationship among the 15 rice landraces from the Lohit district represented two populations. The first three primary coordinates accounted for 50.93% of the overall genetic variation, with Axis 1, Axis 2, and Axis 3 explaining 24.45%, 14.73%, and 11.75% of the variation, respectively (Table 7). Members of population I predominantly clustered in the upper right quadrant, indicating genetic homogeneity. Conversely, members of population II exhibited greater dispersion across the left and lower right quadrants. Beyond the overall clustering, some landraces appeared to be deviant. PV442130, a member of population I, was located near the cluster of population II. Similarly, PV442136, a member of population II, was found among the cluster of population I, and PV442139 from population II was positioned near the population I cluster. Members of population II, PV442137 and PV442145, were positioned near population I. Additionally, PV442131 and PV442132, members of population I, were located on the lower side of the plot, close to members of population II, reflecting a higher intra-population diversity.

Cluster analysis using the Simple Matching coefficient categorized the landraces into two distinct clusters, as depicted in Figure 6, along with their respective subclusters, members, and number of members, as detailed in Table 8. Among all the landraces, the closest relationship was observed between PV442142 and PV442145, which clustered together at a similarity coefficient of ~ 0.97. Cluster A is further subdivided into five subclusters: A1 comprises two members (PV442130 and PV442132), A2 includes two members (PV442142 and PV442145), A3 also contains two members (PV442143 and PV442144), A4 consists of two members (PV442131 and PV442141), and A5 encompasses two members (PV442140 and PV442139). Cluster B is similarly divided into three subclusters: B1 has a single member (PV442133), B2 includes two members (PV442134 and PV442135), and B3 comprises two members (PV442136 and PV442137). The PCA score plot illustrates the distribution of landraces based on Principal Component 1 (PC1) and Principal Component 2 (PC2), with PC1 accounting for the highest variance, followed by PC2 (Figure 7). Landraces PV442140 and PV442145 were positioned on the right side of PC1 with the highest positive scores and also exhibited positive scores for PC2. In contrast, PV442135 and PV442130 were located on the left side of PC1 with high negative scores. PV442134 is situated on the upper side of PC2, displaying a high positive score for PC2 and a negative score for PC1. PV442142 and PV442144 also appeared on the right side of PC1 with positive scores but had negative loadings on PC2, placing them in the lower quadrant. PV442130 and PV442139 were situated on the left of PC1 and the lower side of PC2, indicating negative scores on both axes. Landraces PV442136, PV442141, and PV442133 were located near the origin with negative scores on PC1 and positive scores on PC2, whereas PV442137 also lay near the origin with positive scores on both PC1 and PC2. PV442143 was positioned in the lower-left quadrant, reflecting negative scores for both components. Finally, PV442131 and PV442132 were found near the origin with positive scores on PC1 and negative scores on PC2. The PCA loading plot shows how each character relates to PC1 and PC2 (Figure 8). In PC1, traits like GL, KL, GLC, KLC, LPC, KL/KB, KLC: KBC, 1000GWC and ratios GL/GB load negatively, meaning higher values for these traits are associated with negative PC1 scores. Traits like KB, KBC, DGC, AL, GB, AP, GLB and GLC: GLB load positively in PC1. In PC2, traits like AP, AL, GB, KBC, KB, DGC, GLC: GLB, GL, KL, GLC, KLC, LPC, GLB and ratios GL/GB have high positive loadings, while the traits like KL/KB, 1000GW, and KLC: KBC have negative loading. The traits grouped into meaningful clusters GL, KL, 1000GW on the left, negative PC1, awn and grain width traits AP, AL, GB at the top, positive PC2 and shape ratios KL/KB, GL/GB also on negative PC1. These clusters mean long, slender grains are heavier and awned grains are broader. Trait correlations are inferred from vector angles. Traits pointing in the same direction (e.g., GL and KL) are positively correlated. Opposite vectors (e.g., GL vs GB) are negatively correlated, and perpendicular ones (e.g., AP vs GL/GB) are independent. Overall, PC1 is grain slenderness-to-breadth, and PC2 is awn presence and grain robustness. The PCA helps to differentiate genotypes by key morphological and quality traits, and shows the underlying trade-offs and associations among agronomic characteristics.The AMOVA (Analysis of Molecular Variance) results indicate that most genetic variation, 92.34%, is found within populations, while only 7.66% is among populations (Figure 9). This suggests that the rice landraces from different villages in the Lohit district are genetically similar, with most diversity existing within each village's landraces. This pattern implies there might be significant gene flow or shared ancestry among these groups, meaning seeds or pollen could be moving freely between villages, keeping them genetically alike.

DNA BARCODING

DNA barcoding is a process through which plants and animals can be identified without relying on morphological characteristics. Plant DNA barcoding focuses on specific gene segments, such as rbcL or matK, which are generally between 400 and 800 nucleotides in length; for this particular research, the matK gene segment was utilized. These sequences consist of a unique arrangement of the four nitrogenous bases (A, T, G, C), akin to a coded text string. The process involves sequencing a defined portion of a plant's DNA and translating it into this base sequence. The provided illustrations demonstrate the typical nucleotide lengths and variations for certain plant DNA barcodes. It is important to clarify that the barcode is not a visual graphic but rather a genetic fingerprint used to distinguish plant species by matching it against known sequences in existing databases. Rice landraces were sequenced with the matK primer, and their NCBI accession numbers were obtained. The BOLD platform provides a visual representation of the DNA barcode sequence for each landrace. The graphical barcode encodes the genetic sequence of the DNA barcode. The BOLD system also provides a sample ID and a process ID associated with the sequence, which is highly beneficial for tracking and identification of the landraces. The accession number (from NCBI), sample ID, process ID, sequence length, and number of different nitrogenous bases are listed in Table 9. Additionally, all the barcode figures are presented in Figure 10.The Maximum Likelihood dendrogram identified five distinct clusters, each representing a group of related taxa (Table 10). The variability observed in the dendrogram can be attributed to genetic divergence among the landraces. These landraces were categorized into two primary clades, A and B, with an outlier. PV442132 served as the sole outgroup, distinguishing itself from the other clades. Clade A was further subdivided into subclades A1, comprising PV442130, PV442131, and PV442133; A2, consisting of PV442134, PV442135, and PV442136; and A3, which included PV442137 and PV442139. Similarly, Clade B was divided into B1, containing PV442140; B2, which includes the closely related pair PV442141 and PV442142; and B3, which forms a compact cluster with PV442143, PV442144, and PV442145.
The mean pairwise genetic distance among the landraces from Lohit ranged from 0 to 0.48% (Table 11). The absolute minimum genetic distance was observed between PV442134-PV442130, PV442134-PV442135, and PV442139-PV442140. The maximum genetic distance of 0.48% was noted between PV442130-PV442132, PV442132-PV442135, PV442132-PV442136, and PV442132-PV442141. Additionally, the mean genetic distance was calculated to be 0.0331%.

Ethnobotany of rice in the Lohit district

Rice in yeast cake preparation

P-aa, a yeast cake, embodies centuries-old knowledge transmitted through generations and constitutes a significant aspect of the Mishmi tribe's indigenous knowledge system. This unique fermentation starter is composed of locally sourced rice landraces. Initially, the raw rice is mashed into a paste using a stone pestle. Subsequently, the leaves of Scoparia dulcis are incorporated into the mixture and pounded into a paste. The resultant paste is formed into small cakes, which are then gently dried over a fireplace. This drying process not only prolongs the freshness of the cakes but also enhances their fermentative capacity. The P-aa exemplifies the Mishmi community's deep connection to local biodiversity and traditional practices, playing a crucial role in various rituals and in the production of Yu-shie, a traditional rice beer. Historically, rice-based yeast cakes have served as fermentation starters in numerous Asian cultures, particularly in China, Japan, and Southeast Asia. The traditional preparation of these cakes involves the use of specific microbial cultures, such as yeasts and molds like Saccharomyces cerevisiae and Amylomyces rouxii, to ferment cooked rice (Dung et al., 2005).

Rice in Rituals

Yu-shie crafted from the Ke-Tuilu, Ke-mebo, Bajdhan, and Karthik rice landraces. This rice beer holds significant cultural and ritualistic importance for the indigenous people of the district, particularly in ceremonies such as Tam-La and Tam-La-Du. The production of Yu-shie involves a meticulous process: initially, the rice is cooked and subsequently combined with a starter culture known as P-aa. This mixture then undergoes anaerobic incubation to facilitate fermentation, culminating in the extraction of the rice beer. Within the community, the consumption of alcohol serves as a symbol of respect and a gesture of hospitality. The detailed production process is outlined in the Table 12. The diversity of Indian fermented foods is intricately linked to the ethnic and culinary distinctiveness of each community. The microbial diversity encompasses mycelial fungi, alcohol- or enzyme-producing yeasts, and a limited number of Gram-positive and Gram-negative bacteria, each possessing a range of functional traits. These functional microbes play a pivotal role in traditional fermentation processes due to their functional characteristics, offering consumers various health-promoting benefits (Tamang et al., 2016). Scientific investigations have demonstrated that the microflora present in natural or starter cultures is instrumental in bio-enriching rice with a variety of health-promoting macronutrients and phytochemicals during fermentation (Ray et al., 2016).The Tam-La ritual holds significant cultural importance within the Mishmi community, symbolizing their deep-rooted connection to nature and spirituality. This ceremonial practice, which has been perpetuated across generations, is a complex form of worship directed towards the spirits, with the primary aim of securing blessings for land fertility, prosperity, and the abundant yield of essential crops such as O. sativa, Brassica rapa (mustard), Sesamum indicum (sesame), Setaria sp. (millet), and Solanum tuberosum (potato), among others. These crops are indispensable for the sustenance of the Mishmi people, and the ritual serves as an invocation for a plentiful harvest. In addition to agricultural prosperity, Tam-La is performed to safeguard cattle from diseases and to ensure their health, acknowledging the vital role livestock plays in the community's economic well-being. The scope of the ritual extends beyond agricultural and livestock concerns, as it is also performed for the health and well-being of individuals, particularly in instances of illness. In such situations, the clergyman, who is integral to the ritual, appeals to the spirits for healing and protection. Essential offerings for the ritual include Yu-shie and Chambai. Consequently, Tam-La is a multifaceted ritual, performed on various occasions, tailored to the needs of the tribe and individual circumstances, encapsulating the Mishmi people's spiritual practices and their harmonious relationship with the environment.

Tam-La-Du represents a more communal iteration of the traditional Tam-La ritual, wherein the entire Mishmi tribe congregates to honor and venerate the spirits, seeking blessings for collective well-being, prosperity, and an abundant harvest. This significant event is observed annually on the 14th of February, a tradition that has been consistently maintained since 1970. In contrast to the smaller, more localized Tam-La rituals, Tam-La-Du unites every clergyman within the tribe, who collaborate in leading the worship ceremonies. These gatherings emphasize the importance of unity and shared purpose within the community, as they collectively endeavor to ensure health, prosperity, and abundance not only for individuals but for the entire tribe. Tam-La-Du functions as both a spiritual and social cornerstone for the Mishmi people, reinforcing their cultural heritage and the bonds that unite the community through shared rituals and collective aspirations.

Rice as medicine

Ke-Chaie is a traditional remedy rooted in the cultural customs of the local community, representing knowledge passed down through generations. This method involves cooking rice and allowing it to soak in water overnight. It is believed that this soaking process facilitates the release of beneficial chemical compounds from the rice, which can be utilized as a potent therapeutic beverage. The local population holds that this rice water, having been soaked overnight, is particularly effective in restoring health and energy, especially for those who are ill or frail. Within the community, where Ke-Chaie is employed as a natural remedy to enhance vitality and recovery, there exists a strong belief in its restorative properties. Ke-Chaie is prepared using any available rice landraces, selected based on their availability. This practice reflects the community's strong ties to regional agriculture and its reliance on traditional knowledge to maintain health.

Rice as Food

Chambai is intricately woven into the bio-cultural traditions of the Mishmi community, serving as a pivotal element in various rituals. This dish is indispensable at ceremonies such as Tam-La, Tam-La-Du, weddings, and funerals, symbolizing the transmission of tradition across generations. In contemporary contexts, Chambai has transcended its traditional role in celebrations, now featuring in birthdays and family gatherings. This cherished traditional delicacy of the Mishmi people embodies their rich cultural and ceremonial heritage. The dish is characterized by the unique flavor and texture of roasted sesame seeds (Sesamum indicum) combined with locally sourced cooked rice. A notable feature of Chambai is its spherical form, often accompanied by boiled chicken meat. Its capacity to adapt while preserving its cultural essence has rendered Chambai an integral component of Mishmi cuisine, acting as a conduit between the past and present.

Quantitative ethnobotany of rice in Lohit district

The Frequency of Citation (FC) per species among all reported species was highest for Ke-Tuilu (50), followed by Bajdhan (43), as illustrated in Figure 12. The Use Report (URs) for all species was also highest for Ke-Tuilu (239), followed by Bajdhan (181), as depicted in Figure 13. Ke-Tuilu, with a Cultural Importance Index (CI) score of 2.39, is identified as the most significant variety among all species. Bajdhan follows with a CI score of 1.81, as shown in Figure 14. In addition to rice being a staple food, six traditional uses of rice were documented: P-aa, Yu-shie, Ke-Chaie, Tam-La, Tam-La-Du, and Chambai, as presented in Figure 15.

DISCUSSION

In agricultural research, the agromorphological variation observed among rice landraces is vital for breeding, adaptation, and conservation (Ganavi et al., 2025; Kaur & Kaur, 2024). The current study revealed considerable diversity in grain length, breadth, weight, and kernel dimensions, aligning with findings in other indigenous rice collections (Touthang et al., 2021; Marone et al., 2021). Such morphological variability represents not only aesthetic differences but also adaptations to local environmental stresses, including diseases, pests, and climatic fluctuations (Dwivedi et al., 2016).

The predominance of medium grain size and shape among Lohit landraces mirrors trends reported in northeastern and eastern Himalayan regions (Touthang et al., 2021; Sundarrao & Roy, 2022). Strong correlations among grain length, grain breadth, and thousand-grain weight underline these traits as key determinants for selecting higher-yielding and market-preferred varieties (Rani et al., 2022; Jahan et al., 2021). High heritability estimates and genetic advance for these traits suggest that improvement through selective breeding is feasible and can be efficient (Singh et al., 2018; Kioko et al., 2015). Moreover, kernel-related traits contribute significantly to cooking quality and consumer preferences, further supporting their prioritization in breeding programs (Kumari et al., 2021). Multivariate analyses, including PCA and cluster analyses, revealed clear structuring among the landraces. Traits like grain length, kernel length, and kernel breadth accounted for major axes of variation, supporting their use as diagnostic traits in germplasm characterization (Christina et al., 2021; Roy et al., 2023). The clear separation of clusters, along with overlaps, indicates both distinctiveness and gene flow, reflecting complex historical seed exchange and adaptation processes. The genetic analysis, showing a low mean pairwise genetic distance (0.0331%), suggests extensive seed sharing among communities and a high level of genetic homogeneity (Chakraborty & Ray, 2019; Zahra et al., 2020). Despite this, the presence of two genetic subpopulations points to localized selection pressures or cultural preferences influencing genetic makeup (Parida et al., 2024). Such patterns highlight the need for conservation strategies that maintain intra-population diversity and support local seed systems (Aremu, 2011; Wangpan et al., 2018). DNA barcoding using the matK gene provided deeper insights into genetic relationships and confirmed identities of landraces (Singh & Banerjee, 2018; Mursyidin et al., 2021). The presence of distinct clades and unique outliers emphasizes hidden genetic diversity that may not be apparent from morphology alone. This suggests the importance of integrating molecular tools with field-based assessments to avoid underestimating diversity (Ratnasingham et al., 2024). Ethnobotanical data further enrich our understanding by linking genetic resources with cultural practices. Landraces such as Ke-Tuilu, which exhibited the highest cultural importance index, serve pivotal roles in rituals (Tam-La and Tam-La-Du), food preparations (Yu-shie and Chambai), and medicinal uses (Ke-Chaie) (Wangpan et al., 2019; Tamang et al., 2016; Sinha, 2015). The high frequency of citation and use reports for Ke-Tuilu and Bajdhan reflect deep community attachment and highlight priorities for in situ conservation. Such varieties represent biocultural heritage that supports food security, cultural identity, and resilience against social and environmental changes. The multifunctionality of these rice landraces stresses the urgent need to protect them from genetic erosion and cultural loss. Community-based conservation initiatives, participatory breeding, and local seed banks can be effective in safeguarding both genetic and cultural integrity (Dwivedi et al., 2016; Roy et al., 2023). Promoting traditional knowledge alongside scientific interventions strengthens community empowerment and ensures that conservation efforts are sustainable and culturally sensitive. Overall, integrating morphological, molecular, and ethnobotanical analyses offers a comprehensive framework to guide breeding programs, policy-making, and conservation efforts. This holistic approach not only supports crop improvement but also aligns with global calls for safeguarding agrobiodiversity and cultural diversity (Wangpan et al., 2018; Chakraborty & Ray, 2019). Future research should further explore genotype-environment interactions and socio-economic factors influencing landrace maintenance, ensuring these invaluable resources remain resilient and relevant for generations to come.

CONCLUSION

The ethnobotanical applications of rice among the Mishmi people elucidate its cultural, ritualistic, and medicinal significance in the community. The utilization of rice, ranging from the production of Yu-shie to its role in rituals Tam-La and Tam-La-Du, underscores the profound connection between the community and their environment. The preparation of P-aa and the creation of Yu-shie are culturally important, particularly in rites associated with fertility, prosperity, and spiritual well-being. These processes also demonstrate the role of microbial cultures in rice fermentation. From a quantitative standpoint, Bajdhan and Ke-Tuilu exhibit high cultural importance, as evidenced by the frequency of citation and use reports, with Ke-Tuilu being the most culturally significant variety. DNA barcodes using the matK gene were obtained from the COBL Maximum likelihood analysis revealed two major clades (A and B) with an outgroup, highlighting genetic divergence among the landraces. The genetic distance ranged from 0.00% to 0.48%, with a mean distance of 0.0331%, indicating low genetic variation among the landraces studied. The matK barcode, along with phylogenetic analysis, is a useful method for the accurate identification, characterization, and preservation of rice landraces. The study of agromorphological characteristics of rice landraces from the Lohit district revealed diverse grain morphology. A wide range was observed in grain length, breadth, and weight, with both slender and bold grain types present and different variations in lemma color. Variation in kernel color and awn presence further contributes to phenotypic variation. Characteristics such as grain length and 1000-grain weight exhibited a strong positive correlation with each other. The multivariate analysis, including STRUCTURE, PCA, and cluster analysis, identified two genetic subpopulations. Molecular studies have revealed a clear population structure, specifically two populations. PCA and PCoA were congruent with the population structures. Only a small fraction of the total genetic variance was found among the populations, with the majority of the diversity found within the population.

DECLARATIONS

The authors declare that they have no known competing financial interests or personal relationships that may have influenced the work reported in this study.

AUTHOR CONTRIBUTIONS

SM contributed to the design and implementation of the research; DAB did the analysis; ST conceived the study; and TW supervised the research. All the authors have approved the final version of the manuscript.

ACKNOWLEDGEMENTS

The authors extend their heartfelt gratitude to the villagers of Lohit District, Arunachal Pradesh, for generously sharing their invaluable knowledge. Their cooperation and insights were crucial to the success of this study.

REFERENCES

1. Aitken, S., Gupta, R., & Nair, R. (2024). Assessment of genetic variation and evolutionary dynamics in saffron (Crocus sativus L.) accessions using chloroplast DNA barcodes matK and rbcL. Scientia Horticulturae, 300, 112388.
2. Aremu, C. O. (2011). Genetic diversity: A review for need and measurements for intraspecies crop improvement. Journal of Microbiology and Biotechnology Research, 1(2), 80–85.
3. Caldo, R. A., Sebastian, L. S., & Hernandez, J. E. (1996). Morphology-based genetic diversity analysis of ancestral lines of rice in Philippine rice cultivars. Philippines Journal of Crop Science, 21, 86–92.
4. Chakraborty, D., & Ray, A. (2019). Population genetics analyses of North-East Indian indigenous rice landraces revealed divergent history and alternate origin of aroma in aus group. Plant Genetic Resources: Characterization and Utilization, 17(5), 437–447.
5. Christina, G. R., Thirumurugan, T., Jeyaprakash, P., & Rajanbabu, V. (2021). Principal component analysis of yield and yield related traits in rice (Oryza sativa L.) landraces. Electronic Journal of Plant Breeding, 12(3), 907–911.
6. Cui, D., Lu, H., Tang, C., Li, J., Xinxiang, A., Yu, T., Ma, X., Zhang, E., Wang, Y., Cao, G., Xu, F., Qiao, Y., Dai, L., Li, R., Tian, S., Koh, H.-J., & Han, L. (2019). Genomic analyses reveal selection footprints in rice landraces grown under on-farm conservation conditions during a short-term period of domestication. Evolutionary Applications, 13(8), 1706–1721.
7. Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19(1), 11–15.
8. Dwivedi, S. L., Ceccarelli, S., Blair, M. W., Upadhyaya, H. D., Are, A. K., & Ortiz, R. (2016). Landrace germplasm for improving yield and abiotic stress adaptation. Trends in Plant Science, 21(1), 31–42.
9. Ganavi, B. N., Yashwanth, C. R., Mahesh, H. B., & Harinikumar, K. M. (2025). Analysis of morphological diversity in traditional and improved rice varieties. Mysore Journal of Agricultural Sciences, 59(2), 104–114.
10. Jahan, M. S., Ahmed, M. K., & Talukder, S. (2021). Grain properties of indigenous rice cultivars from Khagrachari district of Bangladesh. South Asian Journal of Agriculture, 8(1–2), 51–56.
11. Jain, S. K., & Mudgal, V. (1999b). A handbook of ethnobotany. Bishen Singh Mahendra Pal Singh.
12. Kaur, M., & Kaur, S. (2024). Screening of wheat germplasm for studying grain length variation during maturity and grain development stages. Plant Archives, 24, 336–343.
13. Kioko, W. F., Mawia, A. M., Piero, N. M., Muriira, K. G., Wavinya, D. W. N., Chemutai, R. L., Matheri, F., Makori, A. W., & Mwenda, N. S. (2015). Phenotypic characterization on selected Kenyan and Tanzanian rice (Oryza sativa L.) populations based on grain morphological traits. Rice Research: Open Access, 3(4), 155.
14. Kumari, S. D., Nirmaladevi, G., Lakshmi, K., & Chamundeswari, N. (2021). Evaluation of grain quality traits in popular rice varieties of Andhra Pradesh. Journal of Pharmacognosy and Phytochemistry, 10(1S), 217–224.
15. Lee, Y. K., Lee, G., & Koh, H.-J. (2025). Molecular signatures of selection for amylose and protein content in Northeast Asian rice (Oryza sativa L.). Plant Biotechnology Reports, 19, 147–155.
16. Li, X., & Yang, Y. (2020). Comparative studies on seed germination of two rice genotypes with different tolerances to low temperature. Plants, 9(10), 1351.
17. Madhu, D., Kamani, H., & Mesara, A. (2023). A study on production and export of rice from India. International Research Journal of Modernization in Engineering Technology and Science, 5(3), 2782–2786.
18. Marone, D., Russo, M. A., Mores, A., Ficco, D. B. M., Laidò, G., Mastrangelo, A. M., & Borrelli, G. M. (2021). Importance of landraces in cereal breeding for stress tolerance. Plants, 10(7), 1267.
19. Parida, M., Gouda, G., Chidambaranathan, P., Umakanta, N., Katara, J. L., Sai, C. B., Samantaray, S., Patra, B. C., & Mohapatra, T. (2024). Genetic relatedness of rice landraces in North East India with wild relatives using chloroplast markers. Plant Genetic Resources, 21(6), 571–578.

20. Pedireddi, U. R., Subba Rao, L. V., Choudhary, R., Patroti, P. D., Pallay, S., Kranthi, K. V. V. S., Kumar, A., & Deepak, N. G. (2018). Effect of seed infection on seed quality and longevity under storage of three rice varieties produced at different environments. Journal of Pharmacognosy and Phytochemistry, 7(1S), 3289–3298.

21. Protection of Plant Varieties and Farmers' Rights Authority. (2007). Guidelines for the conduct of tests for distinctiveness, uniformity, and stability on Oryza sativa L. (rice) (SG/01/2007). Government of India. Published by the Registrar, on behalf of the Chairperson, PPV & FR Authority.

22. Rani, M. H., Faruquee, M., Khanom, M. S. R., & Begum, S. N. (2022). Genetic variability and multivariate studies on the grain physical properties of rice (Oryza sativa L.) landraces. SABRAO Journal of Breeding and Genetics, 54(1), 1–10.

23. Roy, S., Patra, B. C., Kumar, J., Sar, P., Jogi, U. S., Konyak, Z., Banerjee, A., Basak, N., Mandal, N. P., & Bansal, K. C. (2023). Ethnolinguistic associations and genetic diversity of rice landraces in Nagaland, India. Plants, People, Planet, 5(6), 685–698.

24. Sato, S., & Uphoff, N. (2007). A review of on-farm evaluations of System of Rice Intensification (SRI) methods in Eastern Indonesia. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources, 2(54), 1–11.

25. Shadrin, D. M. (2021). DNA barcoding: An effective molecular tool for species identification, molecular authentication, and phylogeny studies in plant science research. Plant Science Journal, 8(2), 45–60.

26. Sharma, S., Pokhrel, A., Dhakal, A., & Poudel, A. (2020). Agro-morphological characterization of rice (Oryza sativa L.) landraces of Lamjung and Tanahun District, Nepal. Annals of Plant Sciences, 9(2), 105–151.

27. Shivani, D., Jabeen, F., Chaithanya, K., Koushik, M. B. V. N., Dileep, G. D., Punnia Koti, E., Supriya, K., Sundaram, R. M., Aravind Kumar, J., & Abdul Fiyaz, R. (2021). Assessment of genetic diversity of rice germplasm using microsatellite markers. The Pharma Innovation Journal, 10(5), 1393–1397.

28. Singh, J., & Banerjee, S. (2018). Utility of DNA barcoding tool for conservation and molecular identification of intraspecies of rice genotypes belonging to Chhattisgarh using rbcL and matK gene sequences. Plant Archives, 18(Special Issue ICAAAS-2018), 69–75. ISSN 0972-5210.

29. Singh, S. K., Vennela, P. R., Singh, R., Gayatonde, V., & Singh, D. K. (2018). Studies on character association, path analysis and genetic variability in rice (Oryza sativa) genotypes. International Journal of Current Microbiology and Applied Sciences, 7(5), 2715–2723.

30. Sinha, A. K. (2015). Traditional rice of West Bengal: Diversity, conservation and agronomic characterization. In P. K. Tyagi & P. K. 'Bharti' (Eds.), Seed technology, plant growth and cropping system (pp. 105–115). Discovery Publishing House Pvt. Ltd.

31. Sultana, S., Faruque, M., & Islam, M. R. (2022). Rice grain quality parameters and determination tools: A review on the current developments and future prospects. International Journal of Food Properties, 25(1), 1063–1078.

32. Thapa, P., Mainali, R. P., Karkee, A., Ghimire, K. H., & Joshi, B. K. (2022). Agro-morphological characterization and diversity assessment of rice landraces in Nepal. Agriculture Development Journal, 16(1), 72–82.

33. Touthang, L., Kalita, H., Makdoh, B., Angami, T., & Singh, R. (2021). Agro-morphological characterization, genetic variation and heritability analysis of rice landraces (Oryza sativa L.) of Arunachal Pradesh, Northeast India. Indian Journal of Plant Genetic Resources, 34(2), 185–195.

34. USDA Foreign Agricultural Service. (2025, April 2). Grain and feed annual: India (Report No. IN2025-0023). U.S. Department of Agriculture.

35. Wangpan, T., Gogoi, H., Tapi, T., & Tangjang, S. (2019). Bio-cultural diversity and ethnobotanic utility of indigenous rice (Oryza sativa L.) in Lakhimpur District, Assam, India. Notulae Scientia Biologicae, 11(1), 138–144.

36. Wangpan, T., Taka, T., & Tangjang, S. (2018). On-farm diversity of indigenous rice (Oryza sativa L.) landraces in border of Eastern Himalaya. Pertanika Journal of Tropical Agricultural Science, 41(1), 239–254.

37. Wangpan, T., Taka, T., & Tangjang, S. (2018). On-farm diversity of indigenous rice (Oryza sativa L.) landraces in the border of Eastern Himalaya. Pertanika Journal of Tropical Agricultural Science, 41(1), 393–410.

Figure captions:

Figure 1. Map of the Lohit district (showing study sites) in Arunachal Pradesh, India, showing village locations where indigenous rice landraces were collected for morphological, genetic, and ethnobotanical analyses.

Figure 2. Heatmap of correlations among 18 agromorphological traits of Lohit rice landraces, indicating strength and direction of trait associations.

Figure 3. Delta K plot showing the optimal number of genetic clusters (K = 2) in STRUCTURE analysis of Lohit rice landraces.

Figure 4. Bar plot of assignment probabilities of individual rice landraces to two genetic clusters (K = 2) as determined by STRUCTURE analysis, highlighting levels of admixture and genetic homogeneity among accessions.

Figure 5. PCoA plot of genetic relationships among 15 rice landraces based on SSR marker data, separating accessions into two genetic clusters

Figure 6. UPGMA dendrogram of genetic similarity among Lohit rice landraces based on SSR markers (Simple Matching coefficients).

Figure 7. PCA score plot illustrating the distribution and clustering of rice landraces based on major morphological traits, highlighting variation along the first two principal components.

Figure 8. PCA loading biplot displaying contributions and correlations of morphological traits to Principal Components 1 and 2, providing insights into key agronomic characteristics driving variation among rice landraces.

Figure 9. AMOVA results indicating high genetic variability within populations and low differentiation among populations.

Figure 10. DNA barcode representations of 15 Lohit rice landraces (matK gene), including accession numbers and local names.

Figure 11. Maximum likelihood phylogenetic tree of Lohit rice landraces based on matK sequences, showing clade relationships.

Figure 12. Frequency of citation (FC) values for different rice landraces in ethnobotanical surveys conducted, indicating relative importance and popularity among the local communities.

Figure 13. Use report (UR) values for each rice landrace recorded reflecting the diversity and extent of traditional uses in local cultural and medicinal practices.

Figure 14. Cultural importance (CI) scores of rice landraces summarizing their overall significance in the community’s ethnobotanical knowledge and cultural practices.

Figure 15. EthnoChord plot illustrating the connections between the rice landraces of Lohit and their documented traditional uses highlighting the multifunctionality and cultural embedding of each landrace.

Table 1. List of SSR and matK primers used for genetic variability analysis and DNA barcoding of rice landraces, including primer sequences and their respective melting temperatures (Tm).

SSR Marker	Forward Primer (5′→3′)	Reverse Primer (5′→3′)	Tm (Forward)	Tm (Reverse)
RM10	TTGTCAAGAGGAGGCATCG	CAGAATGGGAAATGGGTCC	58 °C	58 °C
RM11	TCTCCTCTTCCCCCGATC	ATAGCGGGCGAGGCTTAG	58 °C	58 °C
RM100	CATGGAGAGGAACTGGTGTT	CTCTGATTTCTACCTCTCTC	60 °C	58 °C
RM118	CCAATCGGAGCCACCGGAGAGC	CACATCCTCCAGCGACGCCGAG	74 °C	74 °C
RM131	TCCTCCCTCCCTTCGCCCACTG	CGATGTTCGCCATGGCTGCTCC	74 °C	72 °C
RM135	CTCTGTCTCCTCCCCCGCGTCG	TCAGCTTCTGGCCGGCCTCCTC	76 °C	74 °C
RM153	GCCTCGAGCATCATCATCAG	ATCAACCTGCACTTGCCTGG	62 °C	62 °C
RM154	ACCCTCTCCGCCTCGCCTCCTC	CTCCTCCTCCTGCGACCGCTCC	76 °C	76 °C
RM225	TGCCCATATGGTCTGGATG	GAAAGTGGATCAGGAAGGC	58 °C	58 °C
RM320	CAACGTGATCGAGGATAGATC	GGATTTGCTTACCACAGCTC	62 °C	60 °C
RM333	GTACGACTACGAGTGTCACCAA	GTCTTCGCGATCACTCGC	66 °C	58 °C
RM341	CAAGAAACCTCAATCCGAGC	CTCCTCCCGATCCCAATC	60 °C	58 °C
RM520	AGGAGCAAGAAAAGTTCCCC	GCCAATGTGTGACGCAATAG	60 °C	60 °C
RM1031	GTGAAGGCACACCAACCG	GACGAGGATCGAATTCGAAG	58 °C	60 °C
RM1048	CAAGCCTATAATGTGAATTG	AATTTTTAGTTTGGGGTAGA	54 °C	52 °C
RM1063	GTGATTGGCTGCTGTCATTG	TGGCAAGTGCAACTGCAAG	60 °C	58 °C
RM1075	CCAGTTCAGTAGTTCACACACC	GTTGGGTTGCTGTGTTGTTC	66 °C	60 °C
RM1134	ACACCCAACTTTTCTCACGC	AGCTAGGGTTTCGATCTCCC	60 °C	62 °C
MatK 22	CGATCTATTCATTCAATATTC	TCTAGCACACGAAAGTCGAAG	54°C	64°C

Table 2. Morphological characteristics of rice landraces from Lohit district, including grain length, grain breadth, and their ratio (GL:GB), categorized based on PPV & FRA (2007) standards.

Variety	**Grain Length (GL) (mm) ± SD**	**Grain Length Category**	**Grain Breadth (GB) (mm) ± SD**	**Grain Breadth Category**	**GL: GB Ratio ± SD**	**GL: GB Category**
PV442130	9.34 ± 0.28	Medium	2.83 ± 0.17	Medium	3.30 ± 0.17	Slender
PV442135	9.86 ± 0.31	Medium	3.06 ± 0.18	Broad	3.22 ± 0.18	Slender
PV442132	8.72 ± 0.22	Medium	2.67 ± 0.16	Medium	3.26 ± 0.16	Slender
PV442136	9.99 ± 0.35	Medium	4.00 ± 0.24	Very Broad	2.50 ± 0.24	Medium
PV442133	9.70 ± 0.29	Medium	3.27 ± 0.20	Broad	2.97 ± 0.20	Medium

PV442131	8.33 ± 0.18	Short	3.19 ± 0.19	Broad	2.61 ± 0.19	Medium
PV442137	9.18 ± 0.27	Medium	3.24 ± 0.19	Broad	2.83 ± 0.19	Medium
PV442139	7.52 ± 0.20	Short	3.14 ± 0.19	Broad	2.40 ± 0.19	Medium
PV442134	9.81 ± 0.33	Medium	3.35 ± 0.20	Broad	2.93 ± 0.20	Medium
PV442140	5.92 ± 0.15	Very Short	2.70 ± 0.16	Medium	2.19 ± 0.16	Medium
PV442141	7.24 ± 0.19	Short	2.60 ± 0.16	Medium	2.79 ± 0.16	Medium
PV442145	10.51 ± 0.30	Medium	3.22 ± 0.19	Broad	3.27 ± 0.19	Slender
PV442142	8.90 ± 0.25	Medium	3.37 ± 0.20	Broad	2.64 ± 0.20	Medium
PV442143	9.91 ± 0.32	Medium	3.27 ± 0.20	Broad	3.03 ± 0.20	Medium
PV442144	5.81 ± 0.16	Very Short	3.58 ± 0.21	Very Broad	1.62 ± 0.21	Bold

Table 3. Kernel traits of rice landraces from Lohit district, showing 1000-grain weight, kernel length, and kernel breadth, with categories assigned according to PPV & FRA (2007) guidelines.

Accession No.	Grain Weight (g) ± SD	Grain Weight Category	Kernel Length (KL) (mm) ± SD	Kernel Length Category	Kernel Breadth (KB) (mm)± SD	Kernel Breadth Category
PV442130	25.73 ± 0.53	Medium	7.00 ± 0.70	Long	2.54 ± 0.25	Broad
PV442135	40.98 ± 2.38	Very high	6.59 ± 0.66	Medium	2.59 ± 0.26	Broad
PV442132	38.78 ± 2.43	Very high	6.23 ± 0.62	Medium	2.67 ± 0.27	Broad
PV442136	44.81 ± 0.70	Very high	8.33 ± 0.83	Extra long	2.85 ± 0.29	Broad
PV442133	19.72 ± 1.25	Low	6.59 ± 0.66	Long	2.54 ± 0.25	Broad
PV442131	19.78 ± 0.68	Low	6.95 ± 0.70	Long	2.72 ± 0.27	Broad
PV442137	23.00 ± 1.15	Medium	7.00 ± 0.70	Long	2.54 ± 0.25	Broad
PV442139	19.59 ± 0.98	Low	4.31 ± 0.43	Short	3.04 ± 0.30	Broad
PV442134	22.36 ± 0.65	Medium	3.87 ± 0.39	Short	2.83 ± 0.28	Broad
PV442140	22.30 ± 0.72	Medium	6.46 ± 0.65	Medium	2.36 ± 0.24	Medium
PV442141	18.83 ± 1.47	Low	6.54 ± 0.65	Medium	2.67 ± 0.27	Broad
PV442145	17.91 ± 0.68	Low	7.53 ± 0.75	Extra long	2.59 ± 0.26	Broad
PV442142	21.81 ± 0.55	Medium	6.460 ± 0.65	Medium	2.36 ± 0.24	Medium
PV442143	27.04 ± 1.59	High	6.61 ± 0.66	Long	2.54 ± 0.25	Broad
PV442144	18.50 ± 0.75	Low	5.16 ± 0.52	Short	3.11 ± 0.31	Broad

Table 4. Morphological variation of rice landraces based on kernel length-to-breadth (KL:KB) ratio, lemma colour, kernel colour, and awn presence, including awn length measurements.

Accession No.	KL: KB Ratio ± SD	KL: KB Ratio Category	Lemma Colour	Kernel Colour	Awn (Y/N)	Awn Length (mm) ± SD
PV442130	2.75	Long Bold	Straw	Brown	No	-
PV442135	2.54	Long Bold	Brown furrows on straw	Red	No	-
PV442132	2.33	Long Bold	Gold	White	No	-
PV442136	2.92	Long Bold	Purple black	White	Yes	7.81 ± 0.74
PV442133	2.59	Long Bold	Purple black	White	No	-
PV442131	2.55	Long Bold	Brown furrow on straw	White	No	-
PV442137	2.76	Long Bold	Purple black	White	No	-
PV442139	1.42	Short Bold	Straw	Dark purple	No	-
PV442134	1.37	Short Bold	Straw	Dark purple	Yes	6.98 ± 0.20
PV442140	2.74	Long Bold	Gold	Light brown	No	-
PV442141	2.45	Long Bold	Red	Light red	No	-
PV442145	2.9	Long Bold	Gold	Light brown	No	-
PV442142	2.74	Long Bold	Straw	White	No	-
PV442143	2.6	Long Bold	Brown	Light red	No	-
PV442144	1.66	Short Bold	Straw	White	No	-

Table 5. Mean log-likelihood values (LnP(K)), standard deviations, and Delta K (ΔK) values for different numbers of clusters (K) in STRUCTURE analysis, used to determine optimal population structure of rice landraces of Lohit, Arunachal Pradesh.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln"(K)|	Delta K
1	10	-716.2	0.56569	NA	NA	NA
2	10	-602.38	0.32592	113.82	47.54	145.8651
3	10	-536.1	0.96609	66.28	13.26	13.7254
4	10	-483.08	6.33821	53.02	4.61	0.72733
5	10	-434.67	21.15105	48.41	12.46	0.5891
6	10	-398.72	26.44268	35.95	20.08	0.75938
7	10	-382.85	114.8267	15.87	47.77	0.41602
8	10	-319.21	40.93435	63.64	74.34	1.81608
9	10	-329.91	54.92614	-10.7	8.85	0.16113
10	10	-331.76	58.26993	-1.85	NA	NA

Table 6. Assignment probabilities of rice landraces to two genetic clusters (K = 2) based on STRUCTURE analysis, indicating degree of admixture and final population assignment.

Variety	Cluster 1	Cluster 2	Assignment	Population
PV442130	0.723	0.277	Admixed	I
PV442131	0.006	0.994	Pure	II
PV442132	0.007	0.993	Pure	II
PV442133	0.018	0.982	Pure	II
PV442134	0.012	0.988	Pure	II
PV442135	0.008	0.992	Pure	II
PV442136	0.007	0.993	Pure	II
PV442137	0.020	0.980	Pure	II
PV442139	0.100	0.900	Pure	II
PV442140	0.909	0.091	Pure	I
PV442141	0.864	0.136	Pure	I
PV442142	0.995	0.005	Pure	I
PV442143	0.988	0.012	Pure	I
PV442144	0.993	0.007	Pure	I
PV442145	0.286	0.714	Admixed	II

Table 7. Percentage of genetic variation explained by the first three principal coordinate axes in PCoA analysis of rice landraces from Lohit district.

Axis	Percentage of Variation	Cumulative Percentage
1	24.45	24.45
2	14.73	39.18

3	11.75	50.93

Table 8. Cluster grouping of rice landraces based on genetic similarity analysis, including subcluster membership and number of members per subcluster.

Cluster	**Sub-clusters**	**Members**	**No. of Members**
A	A1	PV442130, PV442132	2
	A2	PV442142, PV442145	2
	A3	PV442143, PV442144	2
	A4	PV442131, PV442141	2
	A5	PV442139, PV442140	2
B	B1	PV442133	1
	B2	PV442134, PV442135	2
	B3	PV442136, PV442137	2

Table 9. Details of rice landrace DNA barcodes, including local names, accession numbers, sample and process IDs, sequence length, and nucleotide base composition (A, G, C, T).

Sl. No.	**Local Name**	**Accession number**	**Sample ID**	**Process ID**	**Length**	**A**	**G**	**C**	**T**
1	Bajdhan	PV442130	RGUL1	RGUL001-25	835	251	157	137	290
2	Ke-botha	PV442137	RGUL2	RGUL002-25	834	251	156	137	290
3	Ke-chama	PV442139	RGUL3	RGUL003-25	740	219	140	126	255
4	Ke-jawya	PV442134	RGUL4	RGUL004-25	739	219	140	126	254
5	Kala lahi	PV442133	RGUL5	RGUL005-25	742	220	141	126	255
6	Karthik	PV442131	RGUL6	RGUL006-25	748	222	142	127	257
7	Ke-ne	PV442141	RGUL7	RGUL007-25	835	251	157	137	290
8	Ke-tuilu	PV442144	RGUL8	RGUL008-25	746	221	142	127	256
9	Ke-mebo	PV442140	RGUL9	RGUL009-25	739	219	140	126	254
10	Ke-piyong	PV442142	RGUL10	RGUL0010-25	832	250	157	136	289
11	Chinni lahi	PV442132	RGUL11	RGUL0011-25	829	249	156	136	288
12	Bora	PV442135	RGUL12	RGUL0012-25	846	254	160	139	293
13	Ke-shi	PV442143	RGUL13	RGUL0013-25	831	250	156	136	289
14	Joha bor	PV442136	RGUL14	RGUL0014-25	835	251	157	137	290
15	Ke-phobo	PV442145	RGUL15	RGUL0015-25	742	220	141	126	255

Table 10. Clade and subclade membership of rice landraces based on maximum likelihood phylogenetic analysis (MEGA v11), including cluster composition and identification of outgroup.

Clade	Sub-clade	Members
–	Outgroup	PV442132
A	A1	PV442130, PV442131, PV442133
	A2	PV442134, PV442135, PV442136
	A3	PV442137, PV442139
B	B1	PV442140
	B2	PV442141, PV442142
	B3	PV442143, PV442144, PV442145

Table 11. Summary of genetic distances among rice landraces calculated using the Kimura-2-Parameter model, showing sample numbers, total pairwise comparisons, and distance statistics.

Taxon	Value
Number of Samples	15
Total Comparisons	105
Mean Distance (%)	0.0331
Minimum Distance (%)	0.00
Maximum Distance (%)	0.48

Table 12. Step-wise description of Yu-shie preparation by the Mishmi Tribe, including cooking, inoculation, incubation, and extraction stages.

Steps	Description	Remarks
	Glutinous rice is usually preferred over the other rice varieties, which are milled and then cooked normally and let cool down.	Cooking
	The rice is then mixed with the **P-aa** till every grain of rice is covered with the **P-aa**.	Inoculation
	The rice is kept packed in anaerobic conditions for 7-14 days.	Incubation
	Once the incubation period has been completed, which is about 5-7 days, it is ready for extraction. It is usually done with the hands in a clean muslin cloth through which it is squeezed.	Extraction

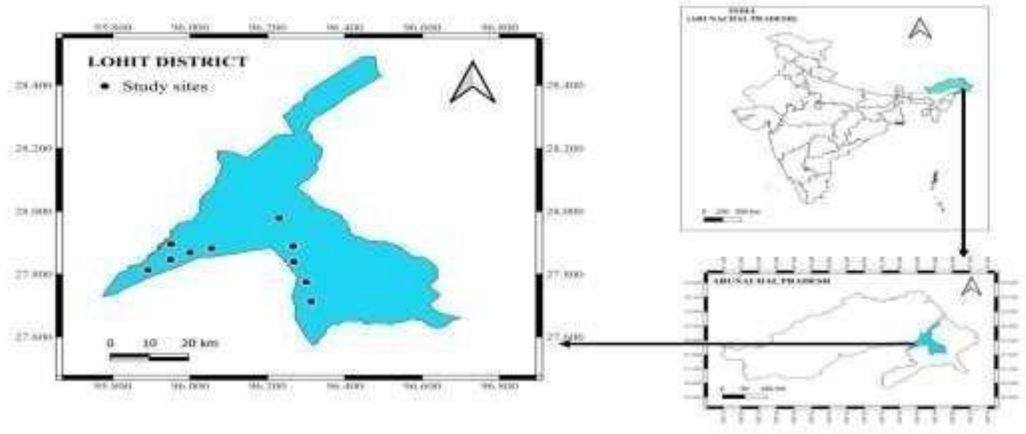


Figure 16. Map of the Lohit district (showing study sites) in Arunachal Pradesh, India, showing village locations where indigenous rice landraces were collected for morphological, genetic, and ethnobotanical analyses.

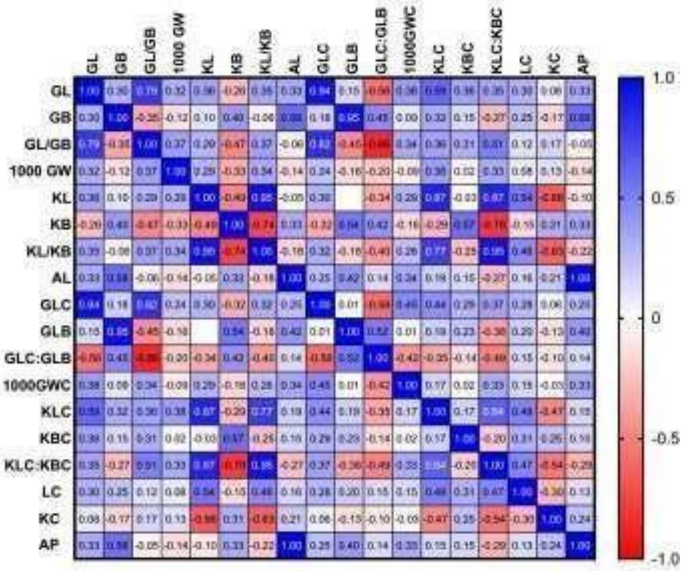


Figure 17. Heatmap of correlations among 18 agromorphological traits of Lohit rice landraces, indicating strength and direction of trait associations.

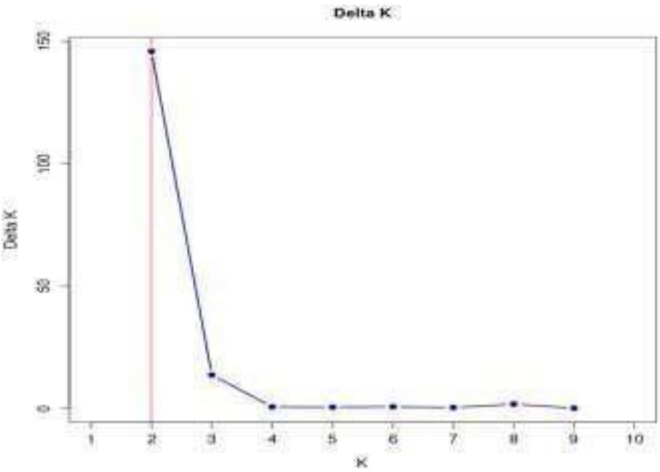


Figure 18. Delta K plot showing the optimal number of genetic clusters (K = 2) in STRUCTURE analysis of Lohit rice landraces.

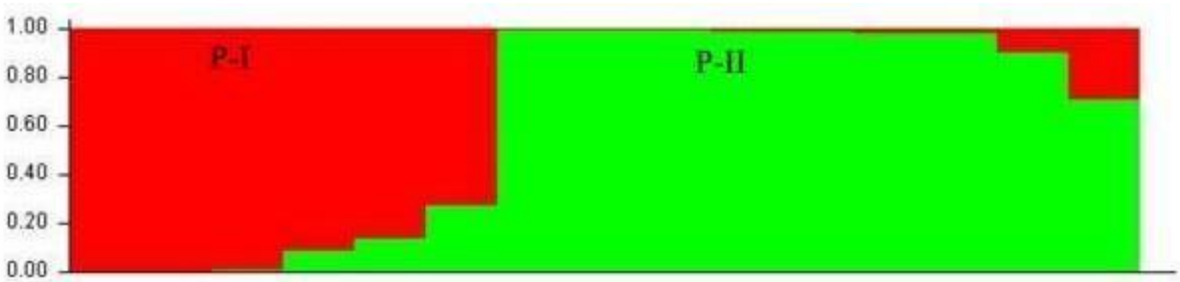


Figure 19. Bar plot of assignment probabilities of individual rice landraces to two genetic clusters (K = 2) as determined by STRUCTURE analysis, highlighting levels of admixture and genetic homogeneity among accessions.

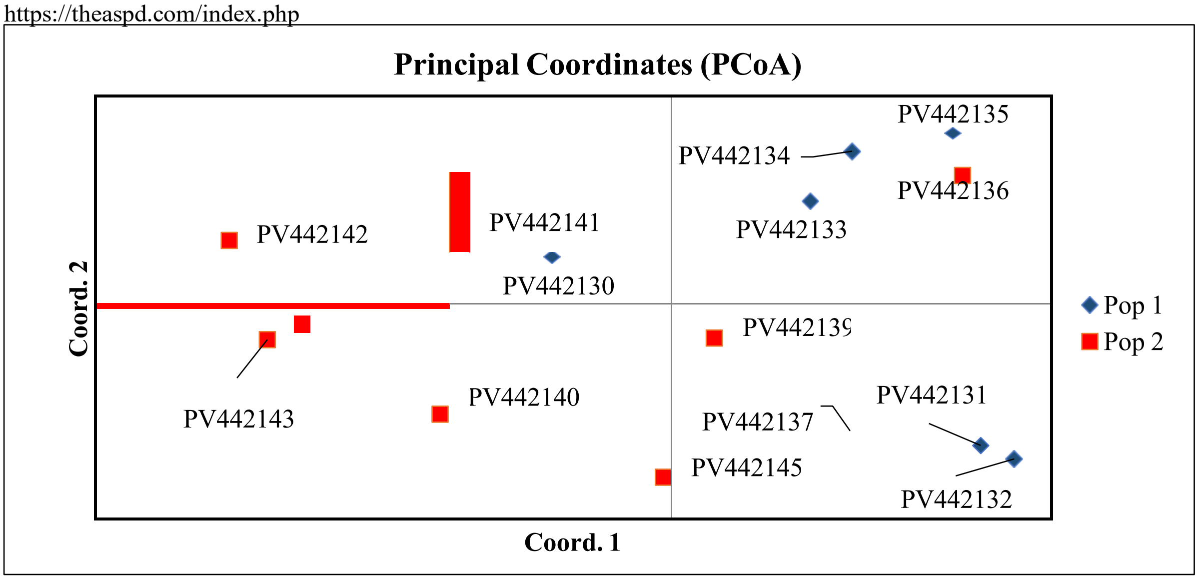


Figure 20. PCoA plot of genetic relationships among 15 rice landraces based on SSR marker data, separating accessions into two genetic clusters

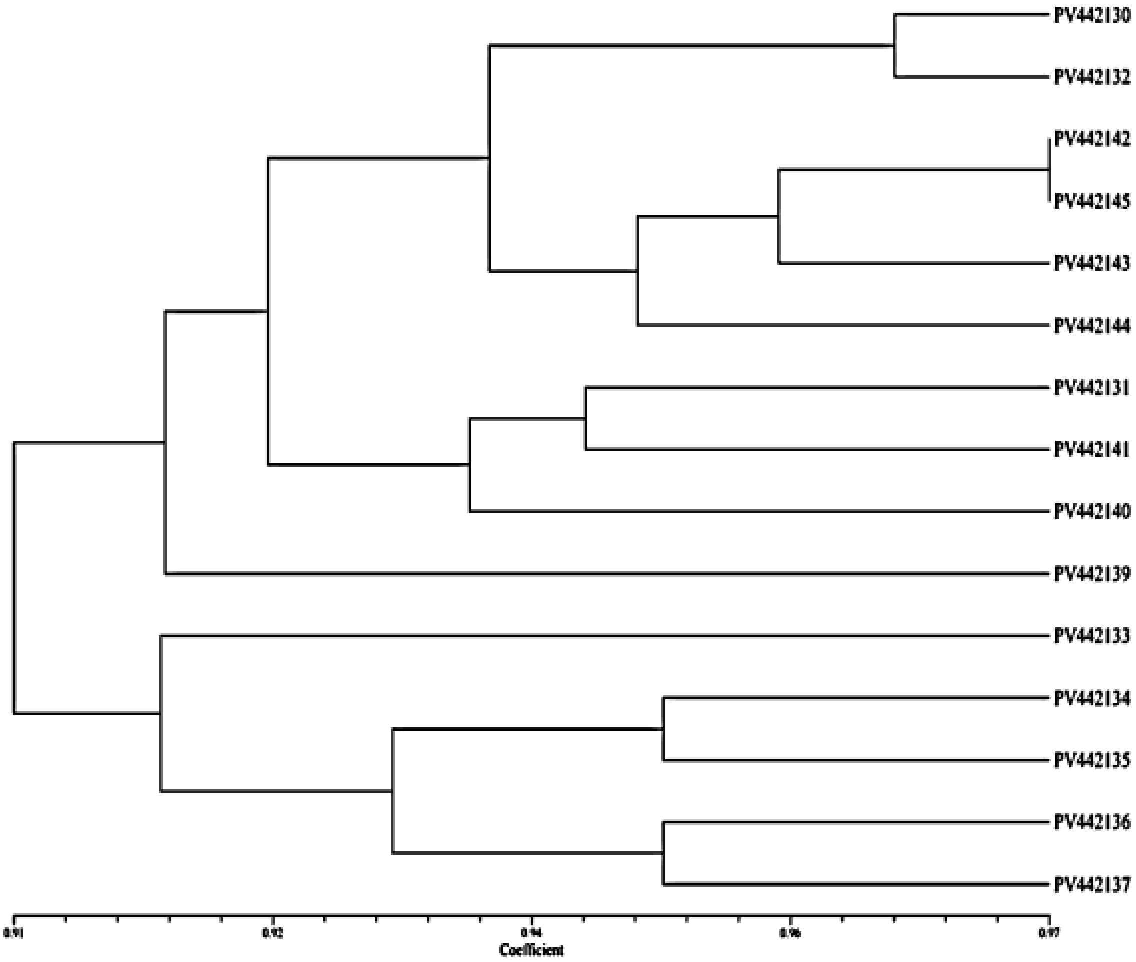


Figure 21. UPGMA dendrogram of genetic similarity among Lohit rice landraces based on SSR markers (Simple Matching coefficients).

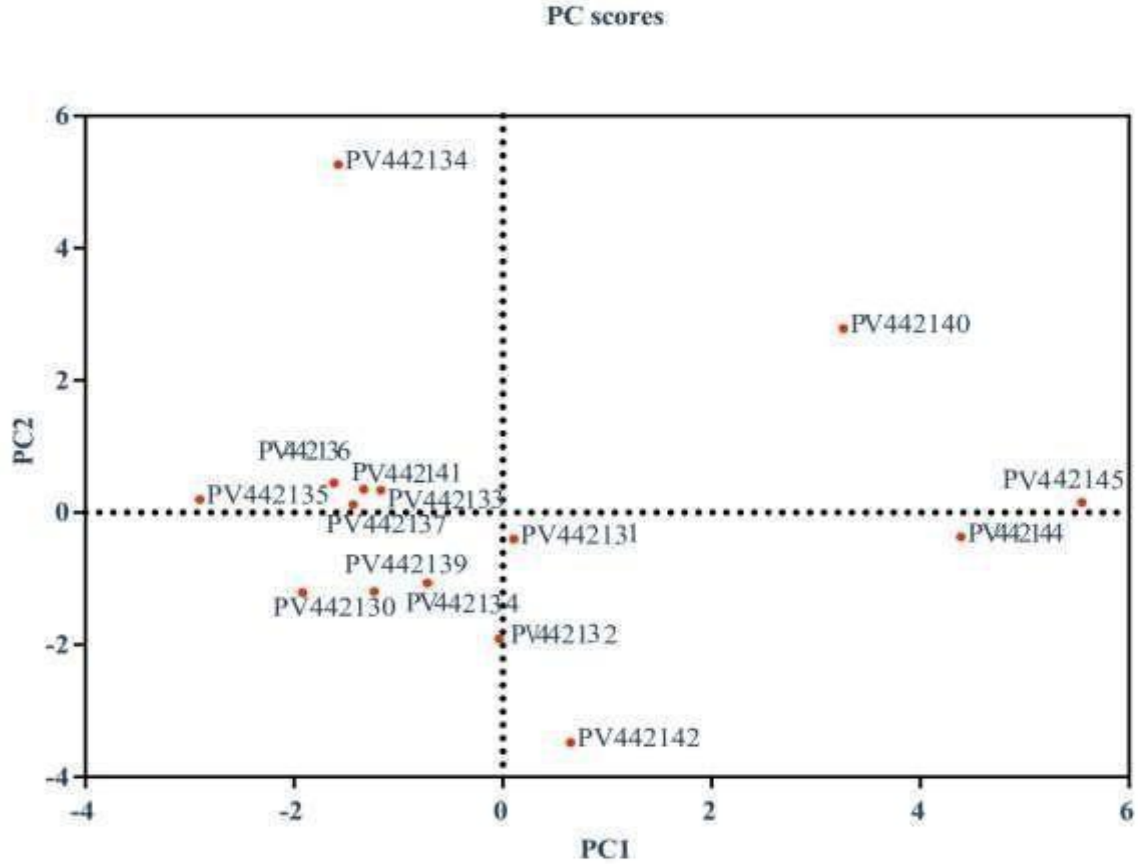


Figure 22. PCA score plot illustrating the distribution and clustering of rice landraces based on major morphological traits, highlighting variation along the first two principal components.

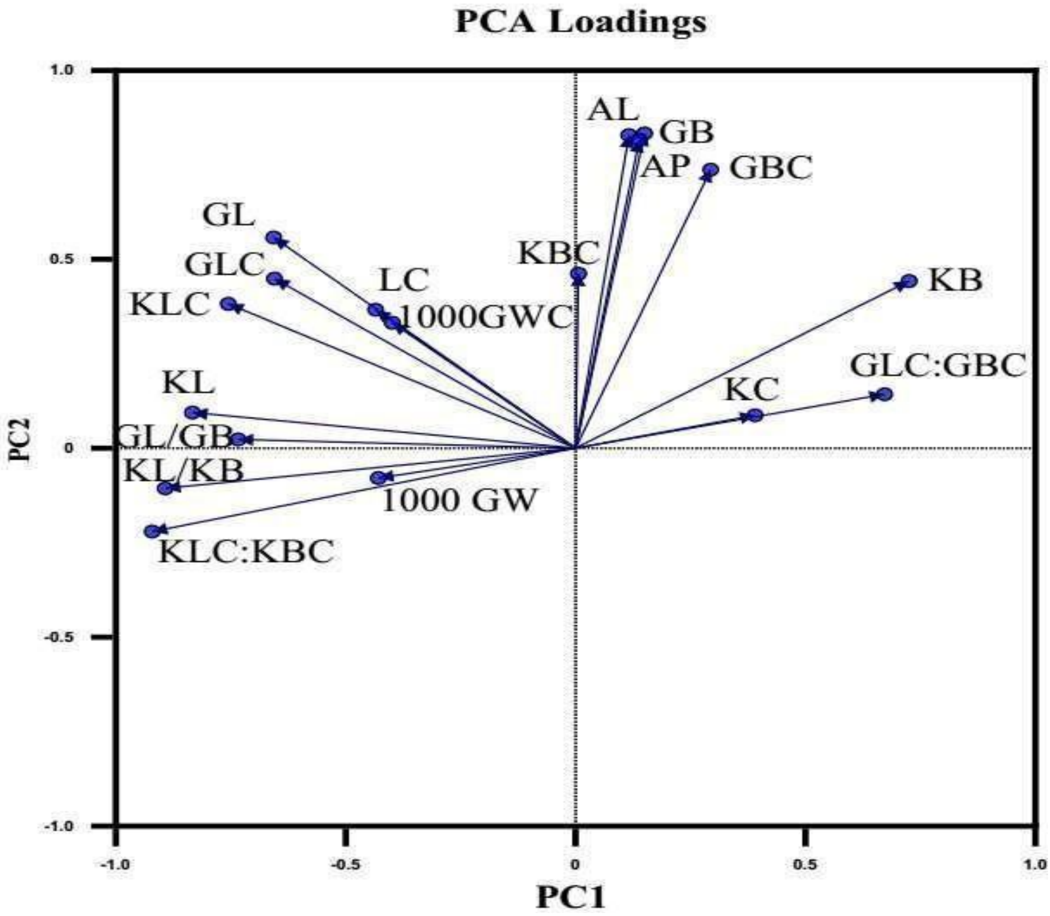


Figure 23. PCA loading biplot displaying contributions and correlations of morphological traits to Principal Components 1 and 2, providing insights into key agronomic characteristics driving variation among rice landraces.

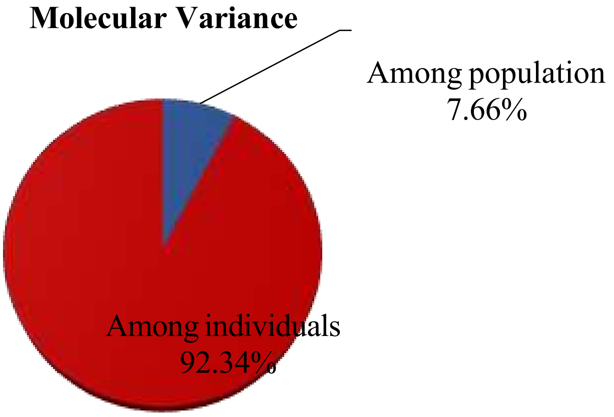
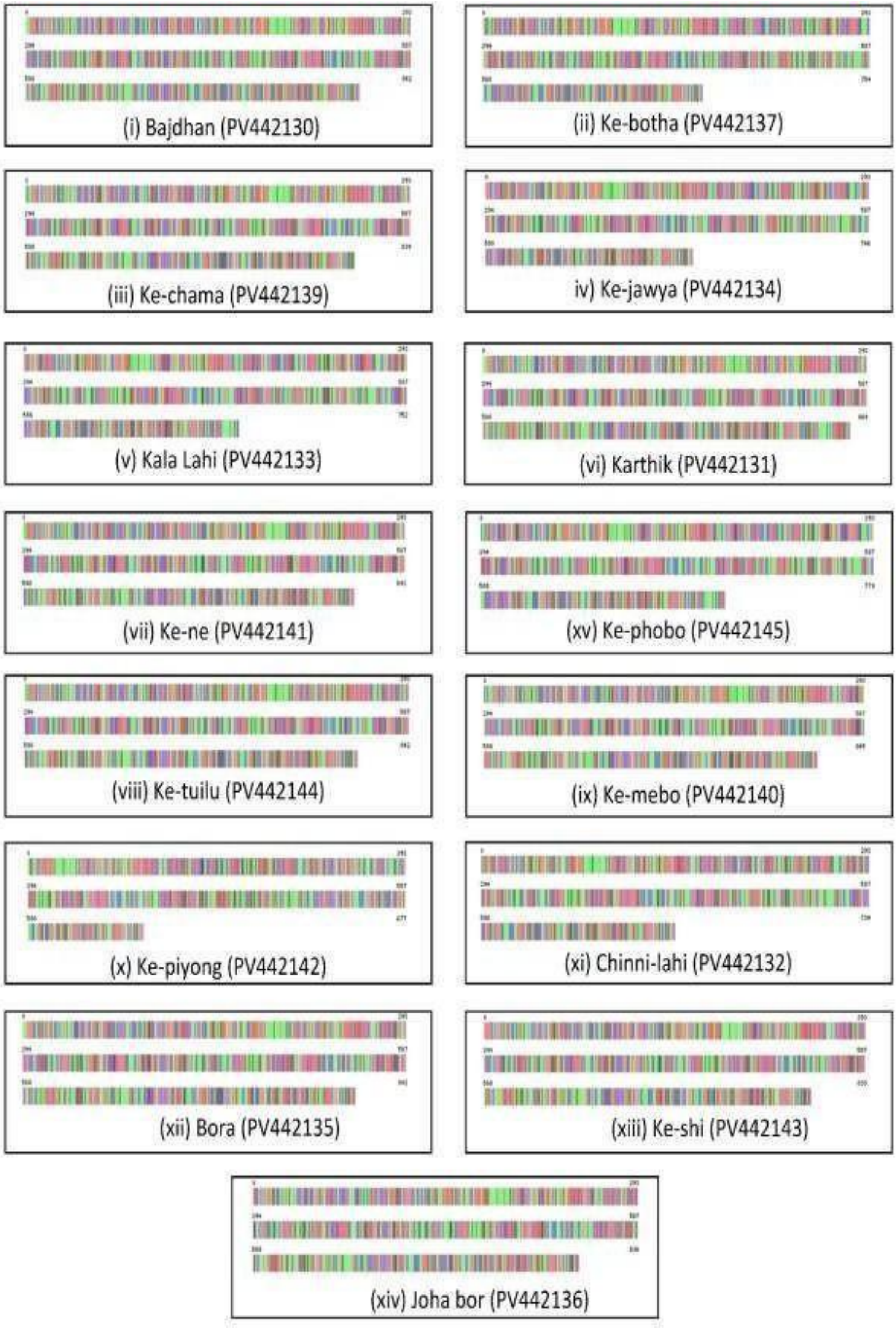


Figure 24. AMOVA results indicating high genetic variability within populations and low differentiation among populations.



(i) Bajdhan (PV442130)

(ii) Ke-botha (PV442137)

(iii) Ke-chama (PV442139)

iv) Ke-jawya (PV442134)

(v) Kala Lahi (PV442133)

(vi) Karthik (PV442131)

(vii) Ke-ne (PV442141)

(xv) Ke-phobo (PV442145)

(viii) Ke-tuilu (PV442144)

(ix) Ke-mebo (PV442140)

(x) Ke-piyong (PV442142)

(xi) Chinni-lahi (PV442132)

(xii) Bora (PV442135)

(xiii) Ke-shi (PV442143)

(xiv) Joha bor (PV442136)

Figure 25. DNA barcode representations of 15 Lohit rice landraces (matK gene), including accession numbers and local names.

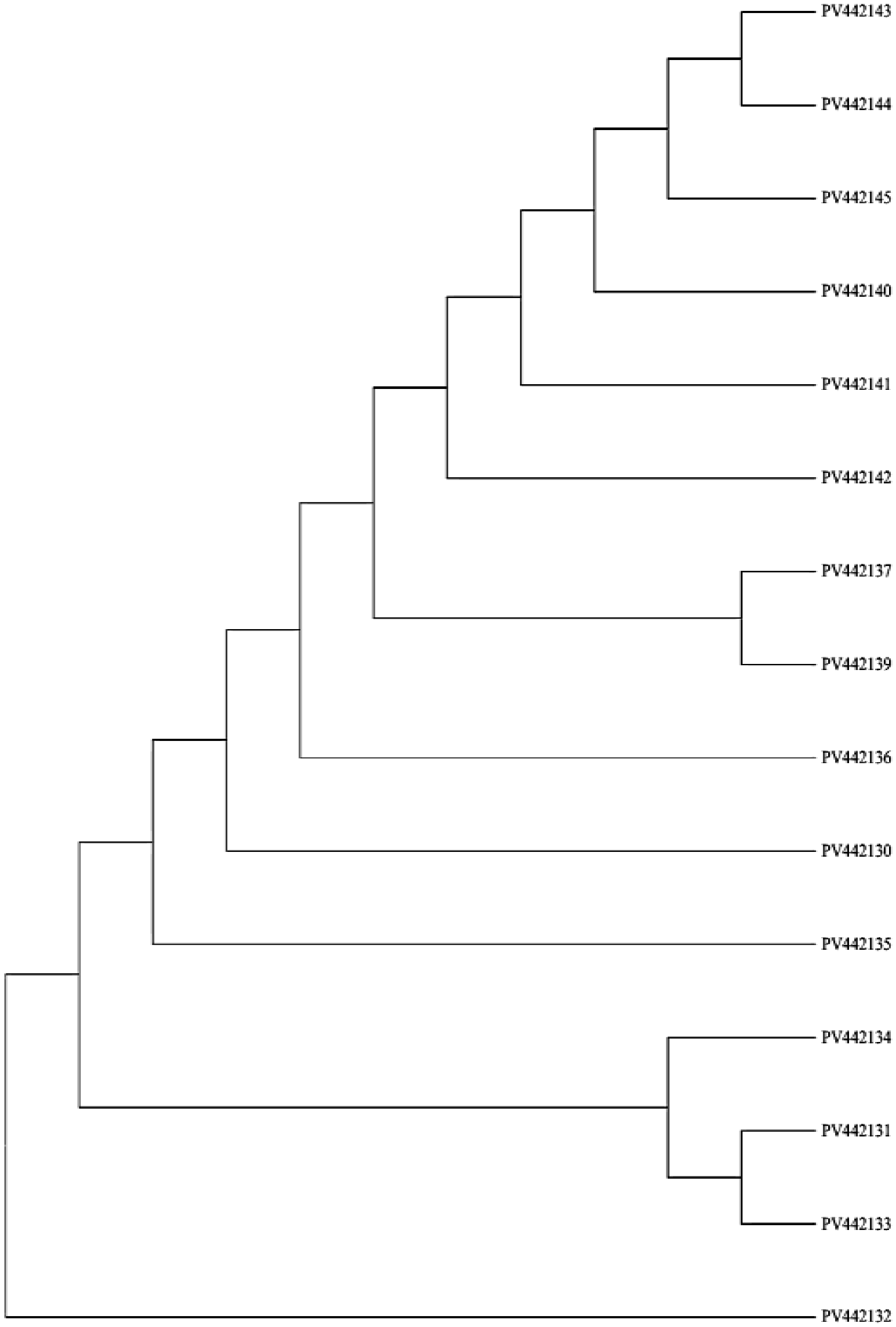


Figure 26. Maximum likelihood phylogenetic tree of Lohit rice landraces based on matK sequences, showing clade relationships.

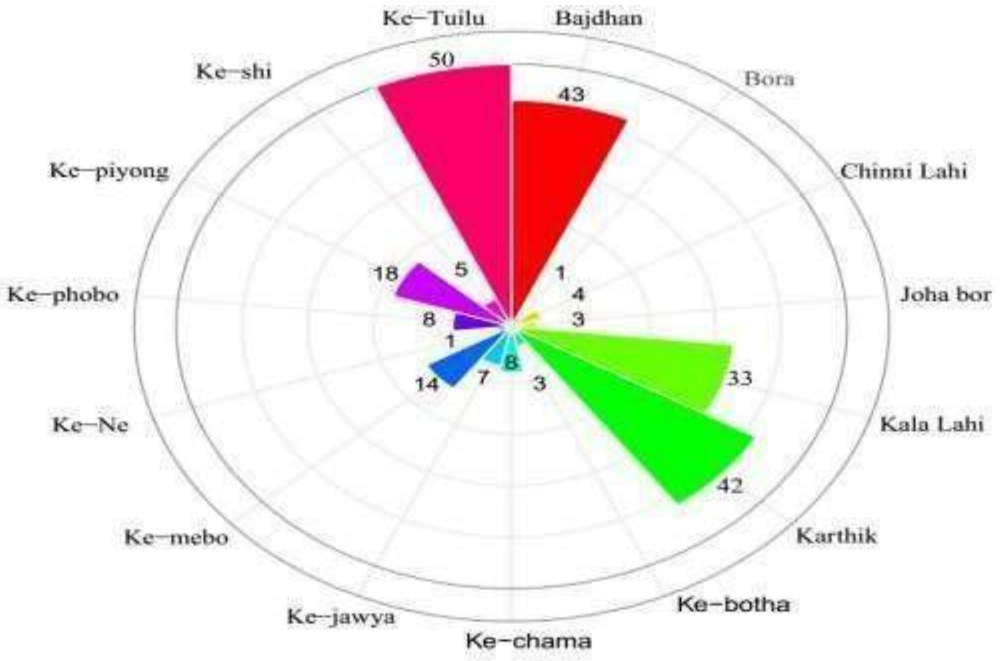


Figure 27. Frequency of citation (FC) values for different rice landraces in ethnobotanical surveys conducted, indicating relative importance and popularity among the local communities.

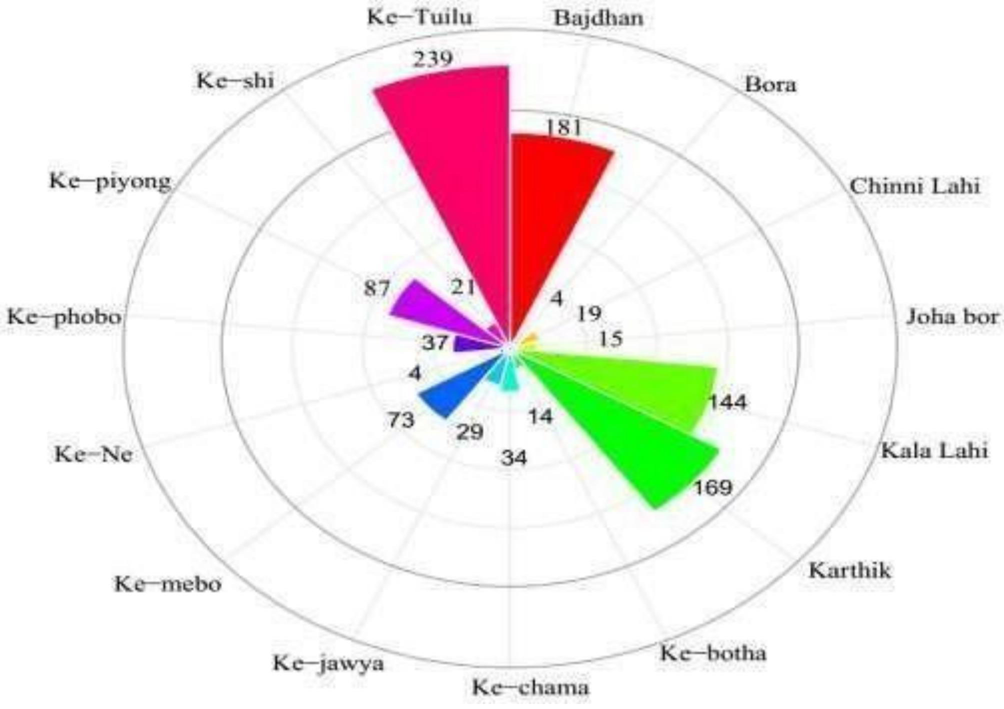


Figure 28. Use report (UR) values for each rice landrace recorded reflecting the diversity and extent of traditional uses in local cultural and medicinal practices.

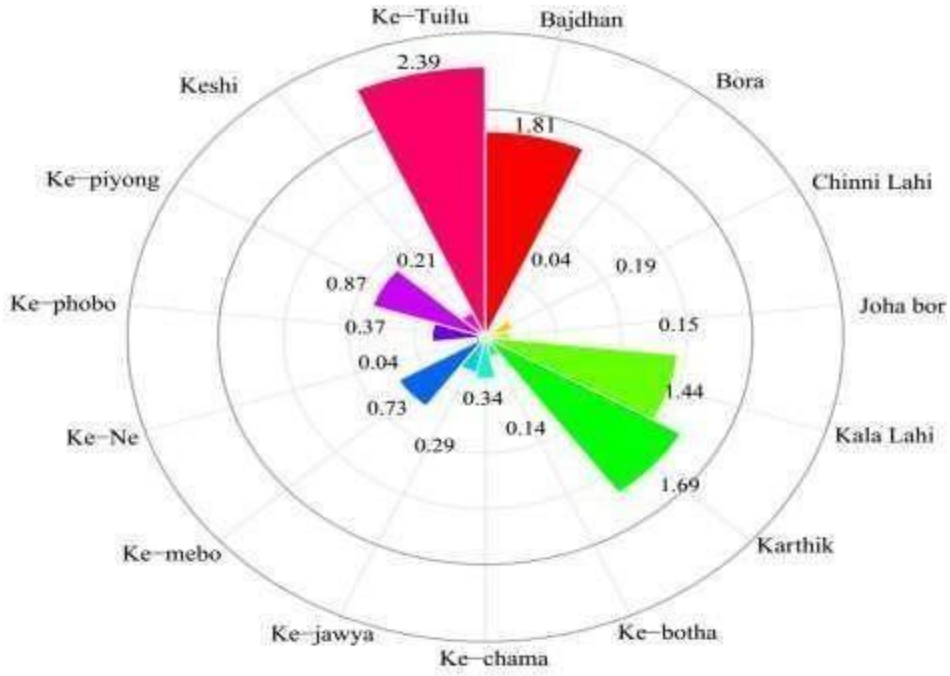


Figure 29. Cultural importance (CI) scores of rice landraces summarizing their overall significance in the community's ethnobotanical knowledge and cultural practices.

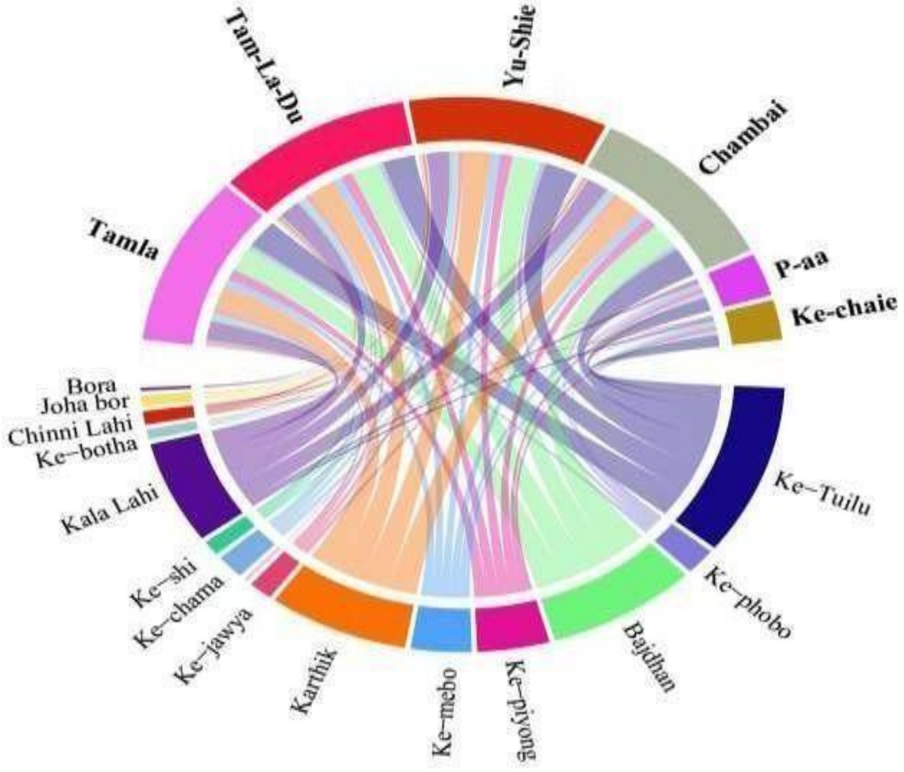


Figure 30. EthnoChord plot illustrating the connections between the rice landraces of Lohit and their documented traditional uses highlighting the multifunctionality and cultural embedding of each landrace.