**Research Article**

Check for updates          OPEN ACCESS

# Genome-Wide SNP Discovery and Population Structure Analysis of 244 Indonesian Local Rice Accessions

Rizky Dwi Satrio[1], Miftahul Huda Fendiyanto[1], Satya Nugroho[3], Aris Hairmansis[4], Miftahudin[2*]

[1]Department of Biology, Faculty of Military Mathematics and Natural Science, The Republic of Indonesia Defense University, Komplek Indonesia Peace and Security Center (IPSC) Sentul, Bogor 16810, Indonesia
[2]Department of Biology, Faculty of Mathematics and Natural Science, IPB University, Kampus IPB Dramaga, Bogor 16680, Indonesia
[3]Research Center for Genetic Engineering, Research Organization of Life Sciences and Environment, National Research and Innovation Agency, Science Center Jl. Raya Jakarta-Bogor KM 46 Cibinong, Bogor 16911, Indonesia
[4]Research Center for Food Crops, Research Organization of Agriculture and Food, National Research and Innovation Agency, Jl. Raya Jakarta-Bogor KM 46 Cibinong, Bogor 16911, Indonesia

## ABSTRACT

Single Nucleotide Polymorphism (SNP) markers are essential tools for the genetic management and breeding of rice, enabling precise characterization of genetic diversity. However, SNP datasets specifically developed for analyzing the population structure of Indonesian local rice remain limited. This study aimed to identify polymorphic SNP markers and analyze the lineage and population structure of 244 Indonesian local rice accessions. BAM files of these accessions were obtained from the 3K Rice Genome Project via the International Rice Informatics Consortium. High-quality SNPs were generated through variant calling and stringent filtering, followed by annotation and population analyses using IR64, DJ123, and Nipponbare as references. A total of 3,839,961 high-quality SNP loci were identified across the 12 rice chromosomes. Hierarchical clustering and principal coordinate analysis consistently grouped the accessions into three major clusters, corresponding to *Indica*, *Japonica*, and admixed groups. Population structure analysis confirmed two main subpopulations: *Indica* and *Japonica*, with *Japonica* further divided into Temperate and Tropical subgroups. Eleven accessions showed admixed ancestry. This study provides essential insights into the genetic diversity and population structure of Indonesian local rice, supporting genetic resource conservation, trait discovery, and breeding efforts.

## 1. Introduction

Rice (*Oryza sativa* L.) is one of the most widely cultivated and consumed staple crops globally, sustaining the livelihood and food security of more than half of the world's population. Its domestication dates back more than 9,000 years, primarily in the Yangtze River Valley for *Japonica* and in the Ganges basin for *Indica*, resulting in two major subspecies that differ in morphology, physiology, and ecological adaptation (Kovach *et al.* 2007; Choi & Purugganan 2018). The process of domestication and human-mediated selection has led to

significant genetic bottlenecks, particularly in modern elite cultivars (Sweeney & McCouch 2007). However, traditional landraces, often cultivated by smallholder farmers under diverse environmental conditions, retain considerable genetic diversity and represent reservoirs of unique alleles that may be absent in modern varieties (Civáň & Brown 2017; Kou *et al.* 2020). These landraces are essential for understanding the evolutionary history of rice and for identifying novel traits such as stress tolerance, grain quality, and disease resistance (Chen *et al.* 2019). In the era of climate change and increasing food demands, the conservation and utilization of such genetic resources have become a global priority (Fornasiero *et al.* 2022). Therefore, characterizing the genetic diversity and population structure of rice

---

\* Corresponding Author
    E-mail Address: miftahudin@apps.ipb.ac.id

landraces is a critical step for both in situ conservation and targeted breeding.

Indonesia, a megadiverse archipelago situated within the center of rice domestication and diversification in Southeast Asia, possesses a rich repository of local rice accessions that are poorly characterized at the genomic level. These local varieties have been cultivated across highly heterogeneous agroecological zones, ranging from terraced uplands in Java and Bali to flood-prone lowlands in Kalimantan and Papua, each posing unique selection pressures (Mursyidin *et al.* 2018). As a result, Indonesian landraces exhibit tremendous phenotypic and genotypic variation that is valuable for dissecting complex traits and for developing climate-resilient cultivars (Muhamad *et al.* 2017; Karimah *et al.* 2021). Unfortunately, many of these accessions remain underrepresented in global genomic databases, and few studies have explored their population structure in detail. Most of the research has focused on phenotypic characterization or limited molecular markers such as SSRs (Thomson *et al.* 2007), which offer lower resolution compared to high-density SNP markers. Given the rapid development of high-throughput sequencing technologies, it is now feasible to comprehensively analyze these local genomes using robust bioinformatics pipelines. A population-scale genomic investigation of Indonesian rice landraces is therefore both timely and necessary.

The 3K Rice Genome Project (3K-RGP), launched by the International Rice Research Institute (IRRI) and partners, has revolutionized rice genomics by sequencing of and releasing high-quality data over 3,000 diverse rice accessions worldwide (Project 2014; Sun *et al.* 2017). This project has produced a unified reference SNP dataset and extensive genotype information, enabling detailed studies on phylogenetics, domestication history, and genomic selection (Alexandrov *et al.* 2015; Mansueto *et al.* 2016). The dataset includes accessions from major rice-growing regions, but representation from Indonesia remains limited or fragmented. Nevertheless, the availability of aligned sequencing data (e.g., BAM files) through the International Rice Informatics Consortium (IRIC) allows researchers to reanalyze and extract population-specific SNPs for focused regional studies. With proper variant calling and stringent filtering, high-confidence SNP datasets can be developed to assess diversity, ancestry, and gene flow within and across rice populations (Li *et al.* 2014). This capacity has led to novel insights into subpopulation structures in regions like South Asia, East Asia, and Africa, but has yet to be fully applied to Indonesian landraces. Thus, the 3K-RGP

provides an underutilized but powerful platform to explore the evolutionary and breeding potential of Indonesian rice.

Analyzing the population structure of local rice accessions is crucial for understanding genetic differentiation, migration events, and admixture that have shaped current genotypic landscapes. Recent advances in population genomics tools—such as STRUCTURE, ADMIXTURE, PCA, and hierarchical clustering—allow researchers to detect subtle population stratifications and to distinguish between closely related genetic groups (Ma & Amos 2012; Porras-Hurtado *et al.* 2013 Cheng *et al.* 2017; Francis 2017). These analyses help trace lineage-specific evolution, identify founder events, and uncover hidden subpopulations that may be associated with unique traits. In rice, subpopulation classification has typically followed four major groups: *Indica, Japonica*, Aus, and Aromatic, each with specific adaptation zones and genetic signatures (Wang *et al.* 2018; He *et al.* 2021; Hamilton *et al.* 2025). However, regional studies have shown that local varieties often do not fit neatly into these categories, especially when they have undergone hybridization or local adaptation. Indonesian rice, with its long history of human cultivation and trade, may harbor a complex ancestry that reflects both ancient and recent gene flows. Therefore, a rigorous population structure analysis using dense SNP data is vital to reveal the hidden genomic architecture of these landraces.

Despite the richness of Indonesia's rice germplasm, few studies have focused on its genome-wide diversity using modern population genomics approaches. Most previous efforts have either focused on narrow subsets of accessions or have used molecular markers with low genome coverage, limiting the resolution of subpopulation detection. The use of genome-wide SNP markers offers a more precise and reproducible method for evaluating genetic relationships and identifying distinct population clusters. SNPs are abundant, bi-allelic, and amenable to high-throughput genotyping and statistical modeling, making them ideal for detecting fine-scale population structures. Moreover, SNP-based studies have the advantage of enabling downstream applications such as genetic mapping, marker-assisted selection, and genomic prediction (Satrio *et al.* 2021, 2023). In Indonesia, where rice breeding efforts are increasingly targeting local adaptability and resilience traits, understanding the genomic landscape of landraces is becoming more urgent. Thus, genome-wide SNP analysis provides both a foundation for fundamental

research and a bridge to practical applications in breeding and conservation.

This study aims to investigate the genetic diversity, lineage relationships, and population structure of 244 Indonesian local rice accessions using whole-genome SNP data. By leveraging BAM files from the 3K Rice Genome Project, we applied a robust pipeline for variant calling, filtering, and SNP annotation to obtain high-quality, polymorphic markers. The reference varieties IR64 (*Indica*), DJ123 (Aus-like), and Nipponbare (Temperate *Japonica*) were included to provide context and serve as anchors for population clustering. We employed hierarchical clustering, principal coordinate analysis (PCoA), and population structure inference to elucidate genetic relationships among accessions. Special attention was given to identifying admixed individuals and detecting subgroup patterns within the broader *Indica* and *Japonica* categories. We also explored whether unique subpopulation events are evident among the Indonesian accessions, which may reflect adaptation or historical migration. Ultimately, this study contributes to filling the knowledge gap in Indonesian rice genomics and provides a valuable resource for future breeding and conservation efforts.

## 2. Materials and Methods

### 2.1. Data Collection

This study utilized whole-genome resequencing data of Indonesian local rice accessions obtained from the 3,000 Rice Genomes Project (3K-RGP), a global initiative hosted by the International Rice Informatics Consortium (IRIC) (Li *et al.* 2014). Specifically, we retrieved the Binary Alignment Map (BAM) files of 244 Indonesian accessions from the IRIC database (https://iris.gramene.org/rice/), representing a wide range of local landraces across various islands and agroecosystems in Indonesia. These BAM files were aligned to the Nipponbare reference genome (IRGSP-1.0) using the Burrows-Wheeler Aligner (BWA) (Li & Durbin 2009), as originally processed by the 3K-RGP. Three reference accessions—IR64 (*Indica*), DJ123 (*Aus-like*), and Nipponbare (*Temperate Japonica*)—were included in the dataset for comparative and lineage classification purposes (Schatz *et al.* 2014). The choice of these references was based on their well-established genome assemblies and their use in previous rice population studies.

### 2.2. Sequence Data Processing

Variant calling was conducted using a standardized pipeline to ensure the accuracy and reproducibility of SNP identification. First, BAM files were subjected to duplicate marking and indexing using SAMtools (Li *et al.* 2009), followed by base quality recalibration and local realignment around indels using Genome Analysis Toolkit (GATK) 4.2.3.0. (McKenna *et al.* 2010). High-confidence single-nucleotide polymorphisms (SNPs) were then called across all accessions using GATK HaplotypeCaller in the GVCF mode and subsequently combined into a joint genotyping dataset. To ensure data quality, SNPs were filtered using stringent criteria, including a minimum depth (DP $\geq$ 15), genotype quality (GQ $\geq$ 30), and minor allele frequency (MAF $\geq$ 0.05). Variants with high levels of missing data (>10%) and those located in poorly aligned regions were removed. The resulting high-quality SNP dataset was further annotated using SnpEff (Cingolani *et al.* 2012), which provided information on variant effects and genomic context based on the Nipponbare reference genome. This rigorous filtering and annotation process yielded a total of 3,839,961 SNPs distributed across all 12 rice chromosomes.

### 2.3. Lineage and Population Structure of Indonesian Local Rice

The genetic relationships among Indonesian rice accessions were analyzed through a combination of multivariate and model-based approaches. Hierarchical clustering was performed using a pairwise genetic distance matrix computed from the filtered SNP dataset, employing the Neighbor Joining (NJ) to generate a dendrogram. Principal Coordinate Analysis (PCoA) was conducted using the same matrix to visualize the clustering patterns in reduced dimensions and to explore the continuity or separation among accessions. Both NJ and PCoA analyses were run on TASSEL software (Bradbury *et al.* 2007). To infer population structure, we employed ADMIXTURE 1.3.0. software (Cheng *et al.* 2017), which estimates individual ancestry proportions assuming a predefined number of ancestral populations (K). The optimal number of K clusters was determined using cross-validation error values, tested from K = 1 to K = 10 with 1,000 bootstrap replicates. The most representative number of clusters (K) for population structure was determined using the delta K method using structureHarvester (https://github.com/dentearl/

structureHarvester). Delta K values were calculated based on the rate of change in the log probability of data between successive K values. Each accession was assigned to a subpopulation based on membership coefficients, and those with shared ancestry across clusters were considered admixed. In addition to population structure analysis, a kinship matrix was constructed using the centered identity-by-state (IBS) method implemented in TASSEL 5.0 (Bradbury *et al.* 2007), allowing estimation of pairwise genetic relatedness among individuals. This matrix helped to quantify the degree of kinship and supported clustering results by identifying genetically similar groups within the population. The combined clustering, dimensional reduction, and admixture analyses provided a robust framework for delineating the evolutionary and population structure relationships of Indonesian local rice.

## 3. Results

### 3.1. Single Nucleotide Polymorphism Markers

Variant calling and stringent filtering from the 244 Indonesian local rice accessions resulted in a total of 3,839,961 high-quality SNP loci distributed across all 12 rice chromosomes (Table 1). Chromosome 1 harbored the highest number of SNPs (410,164), followed by chromosome 2 (353,873) and chromosome 11 (345,469), while chromosome 9 contained the fewest SNPs (243,301). The SNP density, expressed as the average number of base pairs per SNP, ranged from 84 bp/SNP on chromosome 11 to 110 bp/SNP on chromosome 4, with a genome-wide average of 97 bp/SNP. This high resolution of SNP distribution reflects

the effectiveness of the variant calling and filtering pipeline and highlights the rich genetic variability present in Indonesian rice germplasm. The even coverage of SNPs across all chromosomes provides a robust foundation for downstream genetic diversity and population structure analyses. These findings also demonstrate the capacity of resequencing-based SNP discovery to reveal fine-scale genomic variation across a wide set of local accessions. The large number and broad genomic distribution of SNP markers generated in this study surpass those used in many previous reports on local rice populations.

Functional annotation of the SNPs using SnpEff revealed that the majority of the variants were located in non-coding genomic regions, particularly in the upstream (31.3%), downstream (30.1%), and intergenic (20.0%) regions (Figure 1). A smaller fraction of SNPs was located within gene, including exons (6.6%), introns (5.7%), and untranslated regions (UTRs), such as the 5′UTR (0.1%) and 3′UTR (0.3%). Only a very small proportion of SNPs were found in splice site regions, including splice donors, acceptors, and splice site regulatory regions, each comprising less than 0.1% of the total SNPs. The dominance of SNPs in regulatory and non-coding regions is consistent with previous whole-genome SNP distribution patterns observed in rice and other plant species. Nevertheless, the SNPs within exons and UTRs may include functional variants associated with phenotypic traits of agronomic interest. The structural annotation also provides critical information for prioritizing candidate loci in genome-wide association studies (GWAS) or selection mapping. These results underscore the utility of high-density SNP markers not only for population-

Table 1. SNP markers generated from 244 Indonesian local rice varieties

| Chromosome | Chromosome length (bp) | Number of SNPs | SNPs rate* (bp per SNP) |
|---|---|---|---|
| 1 | 43,270,923 | 410,164 | 105 |
| 2 | 35,937,250 | 353,873 | 101 |
| 3 | 36,413,819 | 340,334 | 106 |
| 4 | 35,502,694 | 320,773 | 110 |
| 5 | 29,958,434 | 301,770 | 99 |
| 6 | 31,248,787 | 338,863 | 92 |
| 7 | 29,697,621 | 321,077 | 92 |
| 8 | 28,443,022 | 329,239 | 86 |
| 9 | 23,012,720 | 243,301 | 94 |
| 10 | 23,207,287 | 265,077 | 87 |
| 11 | 29,021,106 | 345,469 | 84 |
| 12 | 27,531,856 | 270,021 | 101 |
| Total | 373,245,519 | 3,839,961 | 97 |

*SNP rate (bp/SNP) indicates the average nucleotide distance between two adjacent SNPs on each chromosome, providing an estimate of SNP density across the rice genome. Lower SNP rate values reflect higher SNP densities, suggesting regions of greater genetic variation among the analyzed Indonesian local rice accessions
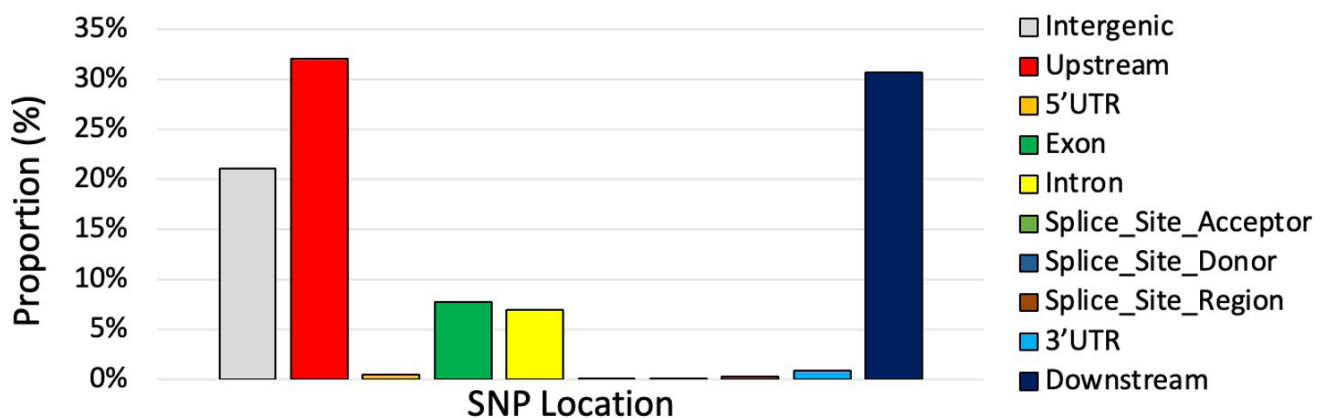
Figure 1. Structural annotation of 3,839,961 SNPs derived from 244 Indonesian local rice varieties

level analysis but also for identifying potential genetic elements underlying adaptation and domestication.

## 3.2. Lineage of Indonesian Local Rice

The lineage analysis of the 244 Indonesian local rice accessions was carried out using hierarchical clustering and principal coordinate analysis (PCoA), both of which revealed consistent grouping patterns. The hierarchical tree showed three major clusters (Figure 2), each aligning with known rice subspecies: *Indica, Japonica*, and *Tropical Japonica*, with reference accessions IR64, Nipponbare, and DJ123 clearly marking these groups. The largest cluster was dominated by accessions grouped with IR64, *Indica*ting a predominance of *Indica*-type rice in Indonesian local varieties. The *Japonica* group, anchored by Nipponbare, formed a separate cluster, while a distinct third cluster represented the Tropical *Japonica* subgroup. PCoA further confirmed this structure (Figure 3), where the first principal component (PC1) explained 91.38% of the genetic variance, distinctly separating *Indica* and *Japonica* accessions. The second component (PC2), accounting for 4.42% of the variance, helped to differentiate Tropical *Japonica* from temperate *Japonica*. Eleven accessions, such as 'Gogo Rajapan', 'Sukamandi 1005', 'Sukaradja', 'Andel Welut', 'Pare Maromba', and 'Telurikan', 'Poenoet Hitam', 'Tjantajan', 'Ketan Urang', 'Kwatik Putih', 'Larondjawi', were positioned between the main clusters, along with DJ123, suggesting admixed ancestry resulting from gene flow or historical hybridization. These admixed accessions likely represent transitional genotypes with alleles from both *Indica* and *Japonica* lineages. This evolutionary pattern underscores the influence of both natural selection and farmer-mediated selection in shaping the genetic structure of local rice germplasm. Together, the phylogenetic and multivariate analyses

provide strong evidence of the complex evolutionary trajectories and domestication processes in Indonesian local rice.

## 3.3. Population Structure and Kinship of Indonesian Local Rice

The population structure analysis of 244 Indonesian local rice accessions revealed clear genetic stratification when analyzed using ADMIXTURE software. Based on the cross-validation error and ΔK method, the most representative number of clusters was found to be K = 2 (Figure 4A), *Indica*ting the primary genetic divergence between the *Indica* and *Japonica* subspecies. At K = 2, most accessions were distinctly assigned to either the *Indica* or *Japonica* group, with a few individuals showing mixed ancestry, *Indica*ting possible gene flow or historical admixture between the groups. When the number of clusters increased to K = 3, an additional cluster emerged representing Tropical *Japonica*, effectively separating it from the temperate *Japonica* group and refining the structure of the *Japonica* lineage (Figure 4B). Several accessions showed patterns of admixture, highlighting the hybrid nature or introgression events in local breeding history. These admixtures reflect complex evolutionary processes in Indonesian rice domestication and cultivation, likely influenced by both natural gene flow and farmer selection. The structure plot also confirmed the correspondence between the genetic clustering and the lineage assignments previously observed in the phylogenetic tree and PCoA.

The kinship analysis using the centered IBS method implemented in TASSEL provided a detailed view of genetic relatedness among the accessions. The kinship heatmap revealed distinct blocks of strong relatedness (Figure 5), with *Indica* and *Japonica* accessions forming two major kinship groups showing high internal genetic
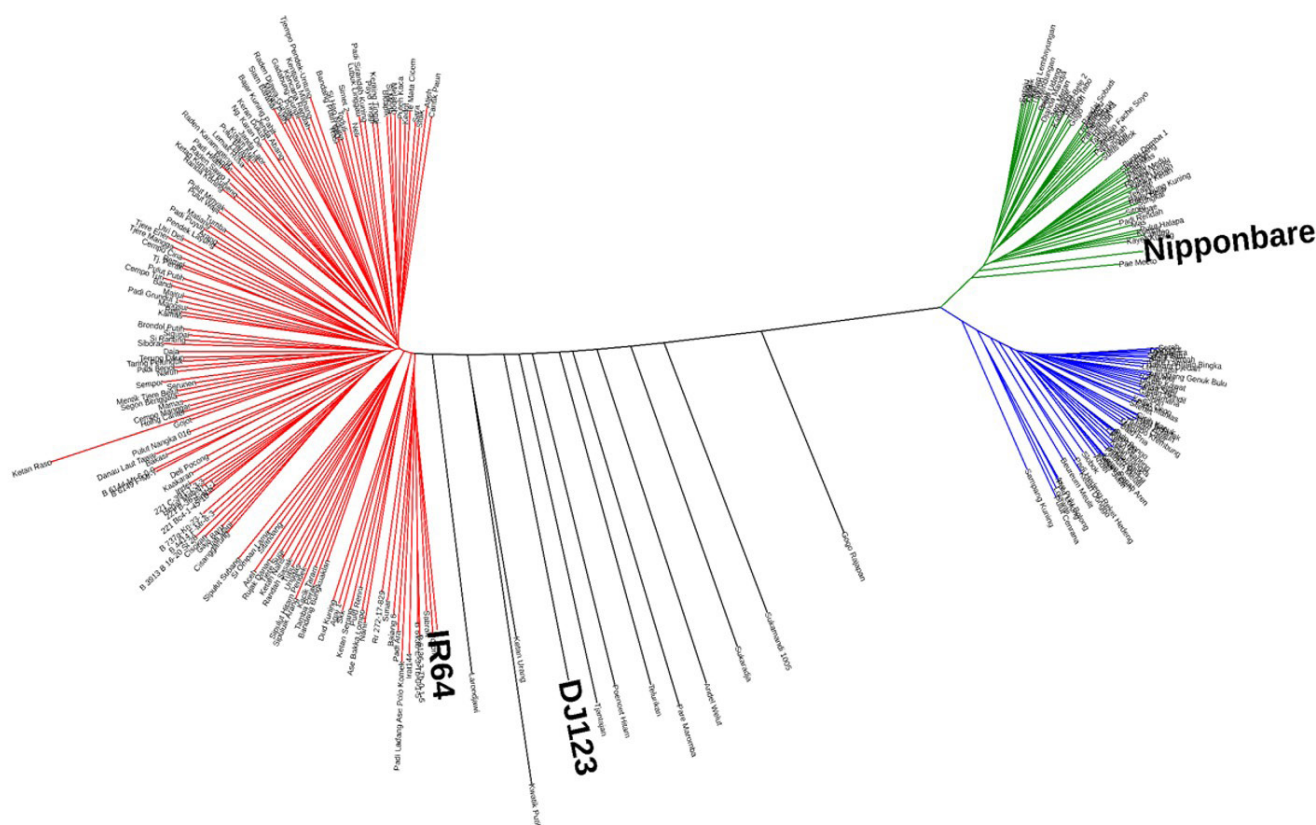
Figure 2. Hierarchical clustering tree of 244 Indonesian local rice accessions and three reference genomes (IR64, Nipponbare, and DJ123) showing three major lineages: *Indica* (red), *Japonica* (green), and Tropical *Japonica* (blue).
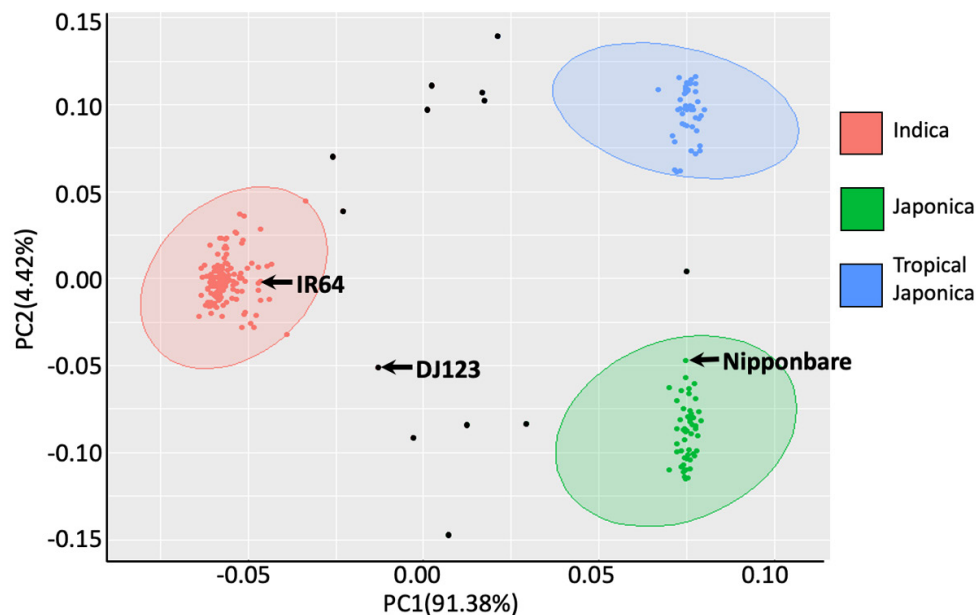


Figure 3. Principal Coordinate Analysis (PCoA) of 244 Indonesian local rice accessions based on SNP data, illustrating genetic clustering into *Indica, Japonica*, *Tropical Japonica*, and admixed groups
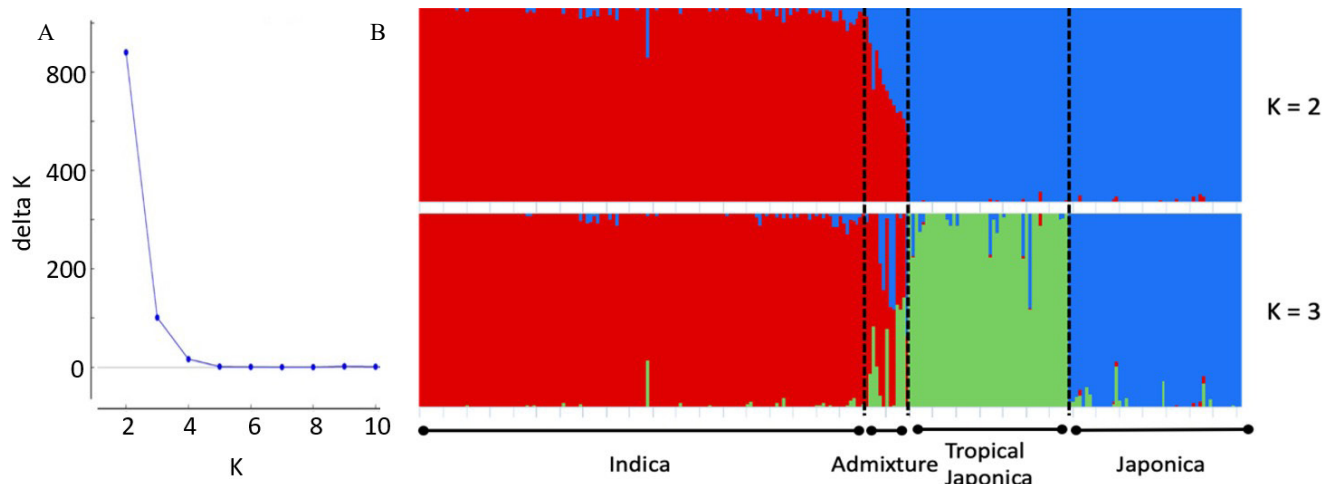
Figure 4. Population structure of 244 Indonesian local rice accessions based on ADMIXTURE analysis. (A) ΔK values indicating the most probable number of clusters (K), (B) ancestry proportions of each accession at K = 2 and K = 3.
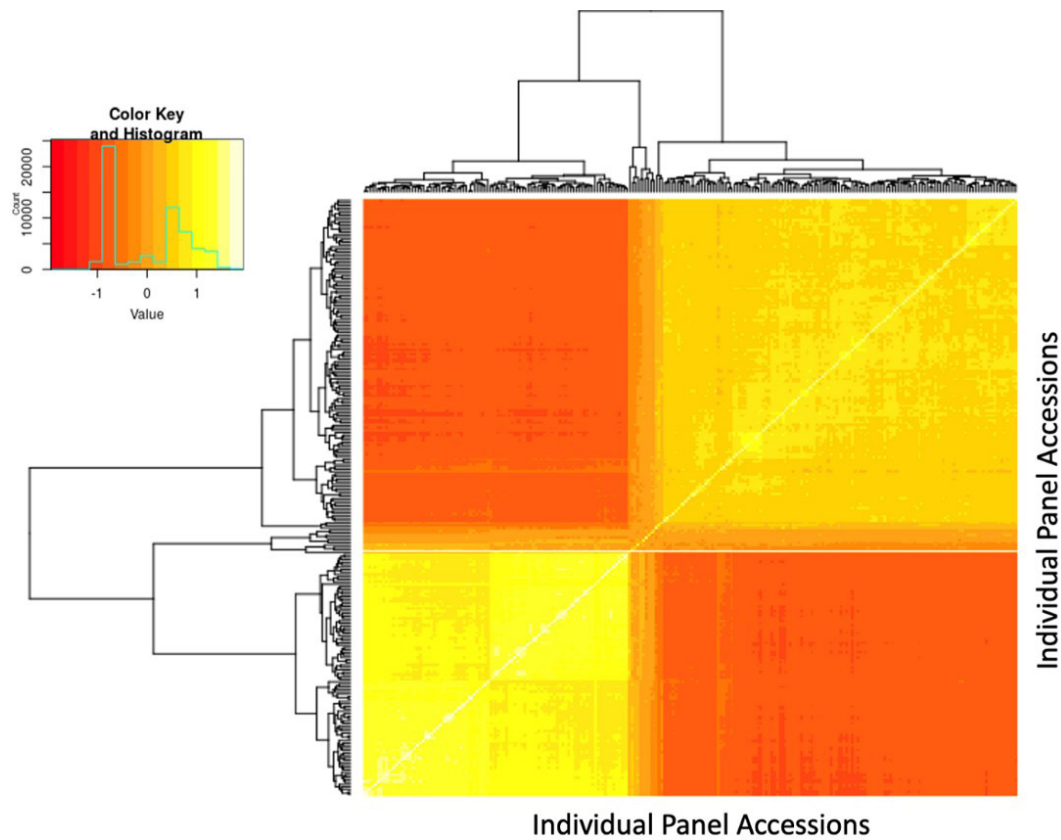


Figure 5. Kinship matrix based on centered Identity-by-State (IBS) analysis among 244 Indonesian local rice accessions

similarity. The *Indica* group, where most accessions were densely clustered, exhibited stronger kinship coefficients, *Indica*ting a more homogenous genetic background likely due to shared ancestry or directional selection. In contrast, *Japonica* accessions showed a more diverse kinship pattern, consistent with their broader geographical origin and genetic variability.

Some admixed accessions occupied intermediate positions in the kinship matrix, supporting their hybrid ancestry as inferred from the structure analysis. This combined interpretation of ancestry and kinship confirmed that Indonesian local rice varieties comprise at least two primary genetic pools with varying degrees of admixture. These results underscore the importance

of both population structure and kinship information in informing future conservation, breeding, and genetic studies of rice in Indonesia.

## 4. Discussion

The present study generated a comprehensive set of 3,839,961 high-quality SNP markers from 244 Indonesian local rice accessions, offering a valuable genomic resource for understanding the genetic structure of rice in the region. These markers, distributed across all 12 chromosomes, demonstrated dense and uniform coverage with an average SNP density of one SNP every 97 bp, which reflects a high degree of polymorphism in the germplasm analyzed (Table 1). Hierarchical clustering, principal coordinate analysis (PCoA), and population structure analysis consistently revealed three major genetic groupings within Indonesian local rice: *Indica*, *Japonica* (divided into Temperate and Tropical subgroups), and a group of admixed individuals (Figures 2-4). The identification of 11 accessions with admixed ancestry highlights the complex historical gene flow and breeding events within the Indonesian archipelago. The use of IR64, DJ123, and Nipponbare as reference genotypes enabled accurate lineage classification. Furthermore, the kinship matrix derived from centered Identity-by-State (IBS) analysis provided additional support for the population structure observed, showing clear genetic relatedness patterns corresponding to the ancestry groups (Figure 5). These findings not only validate the SNP dataset but also emphasize the rich and diverse genetic legacy preserved in Indonesia's local rice varieties, representing a significant contribution to the current understanding of rice population genomics.

The grouping results of Indonesian local rice accessions in this study (Figures 2 and 3) are largely consistent with global classification schemes such as those proposed by the 3K RGP (Wang *et al.* 2018). The majority of accessions fell into the *Indica* group, with possible correspondence to subgroups like ind1a, ind1b, and ind2, which are known to represent lowland tropical varieties widely cultivated in South and Southeast Asia. A smaller subset of accessions was clustered into groups similar to trop (tropical *Japonica*) and temp, suggesting that some Indonesian varieties share ancestry with upland and temperate *Japonica* ecotypes. Interestingly, no clear representatives of aus, aro, or subtrop were observed, reinforcing the notion that these subpopulations have limited distribution in

the Indonesian archipelago. Some accessions showing admixture patterns might correspond to the admix group described in 3K RGP (Sun *et al.* 2017), particularly in areas where *Japonica* and *Indica* varieties have historically coexisted and exchanged genetic material (Londo *et al.* 2006). This pattern of genetic blending is consistent with Indonesia's geographical position as a biodiversity hotspot and a corridor for rice domestication and dispersal (Von Rintelen *et al.* 2017; Deng *et al.* 2020). Thus, the lineage structure observed in this study reflects both historical introduction events and local adaptation processes that have shaped Indonesian rice diversity.

Compared to previous studies, which employed genome-wide SNP data to explore rice population structure at continental or regional levels (Lu *et al.* 2005; Zhang *et al.* 2011; Courtois *et al.* 2012), our findings underscore the importance of high-resolution analysis within specific ecological and cultural contexts. While global studies often prioritize broad categorizations, this study reveals intricate genetic subdivisions and admixture patterns within a single country, pointing to the complex interplay between traditional farming systems, environmental gradients, and human cultural practices. In this context, the Indonesian archipelago offers a unique evolutionary landscape due to its geographic fragmentation, ecological and ethnological diversity (Geertz 1963; Iskandar *et al.* 2018). The SNP dataset used in our study provides a level of granularity that enables the detection of substructure even within *Japonica* groups—distinguishing between Tropical and Temperate *Japonica* types (Figures 2 and 3). These results are not only consistent with the proposed origins and dispersal of rice as outlined by Fuller (Fuller 2011) but also suggest that Indonesia may harbor intermediate forms or transitional genotypes resulting from ancient introgressions. The identification of distinct genetic patterns within widely cultivated and culturally significant varieties further demonstrates the relevance of local adaptation and selection in shaping rice diversity at the population level.

The findings of this study have important implications for rice breeding, conservation, and genetic resource management in Indonesia and beyond. The availability of a high-density, genome-wide SNP dataset specific to Indonesian rice opens opportunities for trait mapping, marker-assisted selection, and genomic prediction in local breeding programs (Miftahudin *et al.* 2021a; Satrio *et al.* 2024). We also used the SNP data obtained in the study for association

mapping on seed characteristics, and have successfully revealed two QTL hotspots, specifically related to the size, shape, weight, and color of rice seeds (Satrio *et al.* 2025). In the future, these SNP markers could be utilised to map traits associated with productivity and environmental stress tolerance, as previously demonstrated with aluminium toxicity (Miftahudin *et al.* 2007; Fendiyanto *et al.* 2019; Miftahudin *et al.* 2021b). Furthermore, understanding the population structure and kinship among these accessions can help breeders avoid inbreeding depression, exploit heterosis, and design effective crossing strategies. In addition, the genomic characterization of these accessions supports the development of rice varieties tailored to diverse agroecological zones and farmer preferences, thereby enhancing their agricultural utility and adaptability. From a conservation perspective, our results provide a framework for prioritizing accessions for in situ and ex situ preservation based on their genetic uniqueness and ancestral background. At a broader scale, this study contributes to the global effort of characterizing rice diversity by filling a significant geographic and genetic gap in current rice genomics datasets, specifically at the country level (Yamasaki & Ideta 2013; Islam *et al.* 2018; Peringottillam *et al.* 2022). The insights gained here may also inform evolutionary studies on rice domestication and adaptation, particularly in tropical island ecosystems. Finally, the identification of admixed individuals highlights the ongoing dynamics of gene flow and offers a unique model system for studying hybridization and genome evolution in crops. Taken together, this research demonstrates the scientific and practical value of exploring rice genetic diversity at both local and global scales.

## Acknowledgements

## References

Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R.R., Ulat, V.J., Chebotarov, D., Zhang, G., Li, Z., 2015. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Research*. 43, D1023-D1027. https://doi.org/10.1093/nar/gku1039

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 23, 2633-2635. https://doi.org/10.1093/bioinformatics/btm308

Chen, E., Huang, X., Tian, Z., Wing, R.A., Han, B., 2019. The genomics of *Oryza* species provides insights into rice domestication and heterosis. *Annual Review of Plant Biology*. 70, 639-665. https://doi.org/10.1146/annurev-arplant-050718-100320

Cheng, J.Y., Mailund, T., Nielsen, R., 2017. Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics*. 33, 2148-2155. https://doi.org/10.1093/bioinformatics/btx098

Choi, J.Y., Purugganan, M.D., 2018. Multiple origin but single domestication led to *Oryza sativa. G3: Genes, Genomes, Genetics*. 8, 797-803. https://doi.org/10.1534/g3.117.300334

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*. 6, 80-92. https://doi.org/10.4161/fly.19695

Civáň, P., Brown, T.A., 2017. Origin of rice (*Oryza sativa* L.) domestication genes. *Genetic Resources and Crop Evolution*. 64, 1125-1132. https://doi.org/10.1007/s10722-017-0518-0

Courtois, B., Frouin, J., Greco, R., Bruschi, G., Droc, G., Hamelin, C., Ruiz, M., Clément, G., Evrard, J., Van Coppenole, S., 2012. Genetic diversity and population structure in a European collection of rice. *Crop Science*. 52, 1663-1675. https://doi.org/10.2135/cropsci2011.11.0588

Deng, Z., Hung, H., Carson, M.T., Oktaviana, A.A., Hakim, B., Simanjuntak, T., 2020. Validating earliest rice farming in the Indonesian Archipelago. *Scientific Reports*. 10, 10984. https://doi.org/10.1038/s41598-020-67747-3

Fendiyanto, M.H., Satrio, R.D., Suharsono, S., Tjahjoleksono, A., Hanarida, I., Miftahudin, M., 2019. QTL for aluminum tolerance on rice chromosome 3 based on root length characters. *Sabrao Journal of Breeding and Genetics*. 51, 451-469.

Fornasiero, A., Wing, R.A., Ronald, P., 2022. Rice domestication. *Current Biology*. 32, 20-24. https://doi.org/10.1016/j.cub.2021.11.025

Francis, R., 2017. Pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*. 17, 27-32. https://doi.org/10.1111/1755-0998.12509

Fuller, D.Q., 2011. Pathways to Asian civilizations: tracing the origins and spread of rice and rice cultures. *Rice*. 4, 78-92. https://doi.org/10.1007/s12284-011-9078-7

Geertz, C., 1963. *Agricultural Involution: the Processes of Ecological Change in Indonesia (Issue 11)*. Univ of California Press. https://doi.org/10.1525/9780520341821

Hamilton, J.P., Li, C., Buell, C.R., 2025. The rice genome annotation project: an updated database for mining the rice genome. *Nucleic Acids Research*. 53, 1614-1622. https://doi.org/10.1093/nar/gkae1061

He, W., Chen, C., Xiang, K., Wang, J., Zheng, P., Tembrock, L.R., Jin, D., Wu, Z., 2021. The history and diversity of rice domestication as resolved from 1464 complete plastid genomes. *Frontiers in Plant Science*. 12, 781793. https://doi.org/10.3389/fpls.2021.781793

Iskandar, J., Iskandar, B.S., Partasasmita, R., 2018. The impact of social and economic change on domesticated plant diversity with special reference to wet rice field and home-garden farming of West Java, Indonesia. *Biodiversitas Journal of Biological Diversity*. 19, 515-527. https://doi.org/10.13057/biodiv/d190227

Islam, M.Z., Khalequzzaman, M., Prince, M., Siddique, M.A., Rashid, E., Ahmed, M.S.U., Pittendrigh, B.R., Ali, M.P., 2018. Diversity and population structure of red rice germplasm in Bangladesh. *PLoS One*. 13, e0196096. https://doi.org/10.1371/journal.pone.0196096

Karimah, A.Z., Siswoyo, T.A., Kim, K.M., Ubaidillah, M., 2021. Genetic diversity of rice germplasm (*Oryza sativa* L.) of java island, Indonesia. *Journal of Crop Science and Biotechnology*. 24, 93-101. https://doi.org/10.1007/s12892-020-00063-4

Kou, Y., Liao, Y., Toivainen, T., Lv, Y., Tian, X., Emerson, J.J., Gaut, B.S., Zhou, Y., 2020. Evolutionary genomics of structural variation in Asian rice (*Oryza sativa*) domestication. *Molecular Biology and Evolution*. 37, 3507-3524. https://doi.org/10.1093/molbev/msaa185

Kovach, M.J., Sweeney, M.T., McCouch, S.R., 2007. New insights into the history of rice domestication. *TRENDS in Genetics*. 23, 578-587. https://doi.org/10.1016/j.tig.2007.08.012

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25, 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J.Y., Wang, J., Zeigler, R.S., 2014. The 3,000 rice genomes project: new opportunities and challenges for future rice research. *Gigascience*. 3, 2047-217X. https://doi.org/10.1186/2047-217X-3-8

Londo, J.P., Chiang, Y.C., Hung, K.H., Chiang, T.Y., Schaal, B.A., 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences*. 103, 9578-9583. https://doi.org/10.1073/pnas.0603152103

Lu, H., Redus, M.A., Coburn, J.R., Rutger, J.N., McCouch, S.R., Tai, T.H., 2005. Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. *Crop Science*. 45, 66-76. https://doi.org/10.2135/cropsci2005.0066

Ma, J., Amos, C.I., 2012. Principal components analysis of population admixture. *PloS One*. 7, e40115. https://doi.org/10.1371/journal.pone.0040115

Mansueto, L., Fuentes, R.R., Chebotarov, D., Borja, F.N., Detras, J., Abriol-Santos, J.M., Palis, K., Poliakov, A., Dubchak, I., Solovyev, V., 2016. SNP-Seek II: a resource for allele mining and analysis of big genomic data in *Oryza sativa*. *Current Plant Biology*. 7, 16-25. https://doi.org/10.1016/j.cpb.2016.12.003

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 20, 1297-1303. https://doi.org/10.1101/gr.107524.110

Miftahudin, Fendiyanto, M.H., Satrio, R.D., Turhadi, Chikmawati, T., 2021a. Genomic improvement of rice for drought, aluminum, and iron toxicity stress tolerance, in: Kole, C. (Eds.), *Genomic Designing for Abiotic Stress Resistant Cereal Crops*. Springer, Cham, pp. 1-69. https://doi.org/10.1007/978-3-030-75875-2_1

Miftahudin, M., Roslim, D.I., Fendiyanto, M.H., Satrio, R.D., Zulkifli, A., Umaiyah, E.I., Chikmawati, T., Sulistyaningsih, Y.C., Suharsono, S., Hartana, A., Nguyen, H.T., Gustafson, J.P. 2021b. OsGERLP: a novel aluminum tolerance rice gene isolated from a local cultivar in Indonesia. *Plant Physiology and Biochemistry*. 162, 86-99. https://doi.org/10.1016/j.plaphy.2021.02.019

Miftahudin, Nurlaela, Juliarni, 2007. Uptake and distribution of aluminum in root apices of two rice varieties under aluminum stress. *HAYATI J Biosci*. 14, 110-114. https://doi.org/10.4308/hjb.14.3.110

Muhamad, K., Ebana, K., Fukuoka, S., Okuno, K., (2017). Genetic relationships among improved varieties of rice (*Oryza sativa* L.) in Indonesia over the last 60 years as revealed by morphological traits and DNA markers. *Genetic Resources and Crop Evolution*. 64, 701-715. https://doi.org/10.1007/s10722-016-0392-1

Mursyidin, D.H., Purnomo, P., Sumardi, I., Daryono, B.S., 2018. Molecular diversity of tidal swamp rice (*Oryza sativa* L.) in South Kalimantan, Indonesia. *Diversity*. 10, 22. https://doi.org/10.3390/d10020022

Peringottillam, M., Kunhiraman Vasumathy, S., Selvakumar, H.K.K., Alagu, M., 2022. Genetic diversity and population structure of rice (*Oryza sativa* L.) landraces from Kerala, India analyzed through genotyping-by-sequencing. *Molecular Genetics and Genomics*. 297, 169-182. https://doi.org/10.1007/s00438-021-01844-4

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., Lareu, M.V., 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*. 4, 98. https://doi.org/10.3389/fgene.2013.00098

Project, 3000 Rice Genomes, 2014. The 3000 rice genomes project. *Gigascience*. 3, 2047-217X. https://doi.org/10.1186/2047-217X-3-7

Satrio, R.D., Fendiyanto, M.H., Miftahudin, M., 2024. Tools and techniques used at global scale through genomics, transcriptomics, proteomics, and metabolomics to investigate plant stress responses at the molecular level, in: Shahid, M., Gaur, R. (Eds.), *Molecular Dynamics of Plant Stress and its Management*. Springer, Singapore, pp. 555-607. https://doi.org/10.1007/978-981-97-1699-9_25

Satrio, R.D., Fendiyanto, M.H., Nurrahmah, N., Anofri, N., Ikhsan, M., Nugroho, S., Hairmansis, A., Miftahudin, M., 2025. Rice QTL hotspots related with seed grain size, shape, weight, and color based on genome wide association study and linkage mapping. *Scientific Reports*. 15, 1-16. https://doi.org/10.1038/s41598-025-05814-3

Satrio, R.D., Fendiyanto, M.H., Supena, E.D.J., Suharsono, S., Miftahudin, M., 2021. Genome-wide SNP discovery, linkage mapping, and analysis of QTL for morpho-physiological traits in rice during vegetative stage under drought stress. *Physiology and Molecular Biology of Plants*. 27, 2635-2650. https://doi.org/10.1007/s12298-021-01095-y

Satrio, R.D., Fendiyanto, M.H., Supena, E.D.J., Suharsono, S., Miftahudin, M., 2023. Mapping and identification of QTL for agro-physiological traits in rice (*Oryza sativa* L.) under drought stress. *Plant Gene*. 33, 1-9. https://doi.org/10.1016/j.plgene.2022.100397

Schatz, M.C., Maron, L.G., Stein, J.C., Wences, A.H., Gurtowski, J., Biggers, E., Lee, H., Kramer, M., Antoniou, E., Ghiban, E., 2014. Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and *Indica*. *Genome Biology*. 15, 1-16. https://doi.org/10.1186/s13059-014-0506-z

Sun, C., Hu, Z., Zheng, T., Lu, K., Zhao, Y., Wang, W., Shi, J., Wang, C., Lu, J., Zhang, D., 2017. RPAN: rice pan-genome browser for ~3000 rice genomes. *Nucleic Acids Research*. 45, 597-605. https://doi.org/10.1093/nar/gkw958

Sweeney, M., McCouch, S., 2007. The complex history of the domestication of rice. *Annals of Botany*. 100, 951-957. https://doi.org/10.1093/aob/mcm128

Thomson, M.J., Septiningsih, E.M., Suwardjo, F., Santoso, T.J., Silitonga, T S., McCouch, S.R., 2007. Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. *Theoretical and Applied Genetics*. 114, 559-568. https://doi.org/10.1007/s00122-006-0457-1

Von Rintelen, K., Arida, E., Häuser, C., 2017. A review of biodiversity-related issues and challenges in megadiverse Indonesia and other Southeast Asian countries. *Research Ideas and Outcomes*. 3, e20860. https://doi.org/10.3897/rio.3.e20860

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. 557, 43-49. https://doi.org/10.1038/s41586-018-0063-9

Yamasaki, M., Ideta, O., 2013. Population structure in Japanese rice population. *Breeding Science*. 63, 49-57. https://doi.org/10.1270/jsbbs.63.49

Zhang, P., Li, J., Li, X., Liu, X., Zhao, X., Lu, Y., 2011. Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PloS One*. 6, e27565. https://doi.org/10.1371/journal.pone.0027565