Borneo Journal of Resource Science and Technology (2025), 15(2): 80-93
DOI: https:doi.org/10.33736/bjrst.8507.2025

Molecular Characterisation of Upland and Lowland Rice from Sarawak, Malaysian Borneo

FREDDY KUOK SAN YEO[1*], RENEE PRISCILLA TRAWAS SYLVESTER EMBUAS[1], ZAZEVIA FRANK CLIFTON[1], MEEKIONG KALU[1], ZINNIRAH SHABDIN[1] & LEE SAN LAI[2]

[1]Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Jalan Datuk Mohammad Musa, 94300, Kota Samarahan, Sarawak, Malaysia; [2]Agriculture Research Centre Semongok, KM20, Jalan Puncak Borneo, 93250, Kuching, Sarawak
*Corresponding author: yksfreddy@unimas.my
Received: 18 December 2024 Accepted: 7 July 2025 Published: 31 December 2025

ABSTRACT

A total of 39 Simple Sequence Repeat (SSR) markers distributed across 12 chromosomes were screened to assess genetic polymorphism among rice accessions. From the 39 tested SSR primer pairs, eight markers (RM1, RM489, RM552, RM444, RM257, RM481, RM166 and RM164) exhibited clear polymorphic banding patterns and strong amplification, while the remaining were excluded due to monomorphism or unclear bands. Among the selected primers, RM257 recorded the highest polymorphic information content (PIC) value (0.9316), while RM552 had the lowest (0.4029), with an average PIC of 0.6891. The observed number of effective alleles (*Ne*) was 1.543 in the lowland population and 1.566 in the upland population. Nei's gene diversity (*h*) was 0.329 for lowland and 0.318 for upland populations, suggesting a low level of genetic divergence. Similarly, Shannon's Information Index (*I*) values were 0.475 and 0.490 for lowland and upland populations, respectively. Analysis of Molecular Variance (AMOVA) revealed that 20% of genetic variation existed among populations, while 80% occurred within populations. Clustering analysis using a UPGMA dendrogram based on SSR genotypes grouped the 44 rice accessions into two major clusters, each further divided into sub-clusters. However, no clear grouping was observed based on morphological traits or geographical origin, likely due to the higher sensitivity of molecular markers in detecting underlying genetic differences that are not always reflected in visible traits. Unlike conventional morphological classification which relies on observable characteristics such as grain shape, plant height or habitat, genome-based clustering captures deeper genetic relationships that offer a more accurate picture of population structure. Simultaneously, the *Maturase-K* barcoding gene marker grouped the accessions into a single large cluster, which also included 94 accessions from diverse origins and three accessions of *Oryza rufipogon*, indicating genome-level similarities. These findings underscore the utility of molecular markers in revealing genetic diversity and population structure. The insights gained from this study can support the development of targeted rice breeding programs essential for advancing sustainable agriculture.

Keywords: *Maturase-K*, Rice, Sarawak, Simple Sequence Repeat

Copyright: This is an open access article distributed under the terms of the CC-BY-NC-SA (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License) which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original work of the author(s) is properly cited.

INTRODUCTION

Rice *(Oryza sativa* L.) is a staple food for more than half of the world's population, serving as a crucial source of nutrition and sustenance (Awika, 2011). Rice is one of the few crop species endowed with the richest genetic diversity and has one of the largest ex situ germplasm collections in the world (Dempewolf *et al.*, 2023). Sarawak is well known as a rich biodiversity centre with diverse types of rice, from commercial high yielding cultivars to indigenous traditional varieties or landraces, which will make great contributions to rice breeding if accessible (Lee *et al*., 2011; Yeo *et al*., 2018; Frank Clifton *et al*., 2024).

The Agriculture Research Centre Semongok currently conserves 2011 rice landrace accessions collected from across Sarawak (Department of Agriculture Sarawak, 2020). Although these accessions display morphological variation, comprehensive molecular characterization is lacking. Such information is crucial for optimizing the use of these genetic resources in breeding, conservation and crop improvement.

To address this gap, molecular tools such as Simple Sequence Repeat (SSR) markers and DNA barcoding genes like *Maturase-K* (*matK*) offer powerful and complementary insights. The SSR markers are widely used due to their high polymorphism, codominant inheritance and genome-wide distribution, making them ideal for assessing nuclear genetic variation, population structure and diversity (Akagi *et al*., 1997; Miah *et al*., 2013). In contrast, *matK*, a chloroplast-encoded gene, provides information from the maternal genome and is particularly useful for identifying phylogenetic relationships and evolutionary lineage due to its conserved sequences and relatively high mutation rate (Yao *et al.*, 2019; Adriansyah *et al.*, 2021).

By combining these two markers; SSR for fine-scale nuclear diversity and *matK* for broader plastid-based phylogeny, this study adopts an integrated approach to analyze the genetic variation of Sarawak rice landraces more comprehensively. This dual-marker strategy enhances resolution, facilitates more accurate classification and provides insights that would not be possible through a single marker system alone.

While genetic research on Sarawak rice is still at an early stage, the previous work such as a 2006 study using six SSR markers on Bario varieties (Tan *et al*., 2006) has highlighted the potential of molecular tools in variety certification. However, broader genetic profiling across diverse landraces remains limited.

Therefore, the objective of this study was to evaluate the genetic diversity of rice landraces collected from the North-West region of Sarawak using both SSR markers and *matK* sequences. This research aims to fill a critical gap in molecular data for local rice germplasm, ultimately contributing to more informed conservation strategies and targeted breeding programs that support sustainable agriculture.

MATERIALS & METHODS

Plant Materials Preparation and DNA Extraction

A total of 44 rice accessions (22 upland and 22 rice accessions) from Frank Clifton *et al*. (2024) were used in this study. The 44 rice accessions may have been obtained from the same seed source or different seed sources from farmers, as reported in Frank Clifton *et al*. (2024). The 44 rice accessions were designated as UNIMAS-01 to UNIMAS-44 (Supplementary Table 1). Upland rice is grown in rainfed and non-irrigated areas, often on elevated or hilly terrains, whereas lowland rice is cultivated in paddy fields with irrigated or rainfed conditions, where the soil remains waterlogged for most of the growth cycle (Sohrabi *et al*., 2012).

The accessions were planted in soil mixture of topsoil, compost and sand (3:2:1 ratio). For upland accessions, each of the accessions were transplanted into pots with the drainage holes at the bottom to drain the excess water. Lowland accessions were transplanted into pots without holes at the bottom but having holes at the top of the pots to maintain the water level about 2 cm above soil surface. Nitrogen, phosphorus, and potassium (NPK) fertilizer (17.5:15.5:10) was used for fertilization, with each plant pot receiving 10 grams per application. The first application was conducted around the time of transplanting.

At one month old, young leaf samples of about 4-5 cm were collected. DNA was extracted using cetyltrimethylammonium bromide protocol following Doyle and Doyle (1987). The concentration of extracted DNA was measured using Nano-drop machine (Model: ND-2800-ODJ Nano DOT Nucleic Acid Analyzer; Hercuvan), where the quality and quantity of genomic DNA were measured. The integrity of the DNA was visually checked by electrophoresis on 1% agarose gel. The DNA stocks were diluted to 20 ng/ul using distilled water and stored at -20ºC.

Microsatellite Marker Amplification

A total of 39 Simple Sequence Repeat (SSR) primer pairs were selected randomly from Gramene Markers Database (https://archive.gramene.org/markers/microsat/ssr.html), representing 12 rice chromosomes (Supplementary Table 2).

Polymerase Chain Reaction (PCR) amplification was performed following Zhu *et al*. (2012) in a 20 µL reaction containing 5 ng of genomic DNA, 0.2 µM of each forward and reverse primer, 0.2 mM dNTPs, 1X PCR buffer (First Base Laboratories), 2.5 mM $MgCl_2$ and 1

unit of Taq DNA polymerase (First Base Laboratories). The PCR started with initial denaturation at 94°C for 4 minutes followed by 40 cycles of denaturation at 94°C for 45 seconds, annealing at optimized temperature (Supplementary Table 2) for 45 seconds, extension at 72°C for 1 minute, and with a final extension at 72°C for 10 minutes. The amplified PCR products were electrophoresed on 2% agarose gel stained with ethidium bromide (0.1mg/ml; EtBr). The gel was visualized under an ultraviolet transilluminator.

The final SSR dataset was reviewed for scoring consistency and unclear bands were re-scored to minimize genotyping errors.

SSR Marker Analysis

The amplified bands were visually scored as present (1) or absence (0) separately for each primer pair. Polymorphic Information Content (PIC) was calculated following Anderson *et al*. (1993), Eq. (1):

$$PIC_i = 1 - \Sigma P_{ij}^2 \qquad \text{Eq. (1)}$$

Where i is the total number of alleles detected for SSR marker and P_{ij} is the frequency of the jth allele for marker i.

The genetic distances as well as Analysis of Molecular Variation where number of alleles, number of effective alleles (*Ne*), Nei's gene diversity (*h*) and Shannon's Information Index (*I*) were calculated in Genetic Analysis in Excel (GenAlEx) version 6.51 (Smouse *et al*., 2017). Dendrogram was constructed using Unweighted Pair Group Method with Arithmetic Means (UPGMA) generated from Molecular Evolutionary Genetic Analysis (MEGA X) v.10.1 software based on genetic distance.

Amplification of *Maturase-K* Barcoding Marker

The *matK* region was amplified by using forward primer: 5'- TAA TTA AGA GGA TTC ACC AG -3' and reverse primer: 5'- ATG CAA CAC CCT GTT CTG AC -3'. The PCR protocols were based on Patil *et al*. (2015) with expected product size of 1500 bp.

The PCR reaction contains 3-5 ng of genomic DNA, 0.5 µM each of forward and reverse primers, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1X PCR buffer (First Base Laboratories), and 2.5 unit of Taq DNA polymerase (First Base Laboratories). The thermocycler profile for amplification of *matK* had initial denaturation at 94°C for 4 minutes, followed by 35 cycles of denaturation at 94°C for 1 minute, 2 minutes of annealing at 57.8°C, extension at 72°C for 2 minutes and ended with final extension at 72°C for 15 min.

The PCR products were electrophoresed in a 1.5% agarose gel, stained with EtBr and visualised under UV transilluminator. The PCR product was purified using QIAquick Gel Extraction Kit (QIAGEN, Germantown, MD, USA) following the provided protocol by the manufacturer. Purified PCR products were subjected for two-way sequencing by Apical Scientific Sdn. Bhd. Malaysia.

Cluster Analysis Based on *Maturase-K*

The *matK* sequences from the 44 accessions of this study were manually edited in BioEdit software. An additional 94 *matK* sequences of *O. sativa* were downloaded from GenBank database at National Centre for Biotechnology Information (NCBI) (Supplementary Table 3). One to three *matK* sequences of *O. rufipogon*, *O. punctata*, *O. officinalis, O. australiensis,* and *O. granulata* were also downloaded (Supplementary Table 3). The *matK* sequences of the 44 accessions of this study were aligned with the sequences from NCBI using ClustalW (MEGA X v. 10.1). Phylogenetic tree was constructed using Maximum Likelihood with Tamura 3-parameter model evolutionary rates among sites with discrete gamma distribution (T92+G) at 1000 bootstrap value.

RESULTS

From 39 tested SSR primer pairs, eight primer pairs displayed polymorphic banding patterns and exhibited strong polymorphism among the accessions. The primers with clear amplification results were RM1, RM489, RM552, RM444, RM257, RM481, RM166 and RM164 (Figure 1). The other primers showing either monomorphic banding patterns or unclear band were excluded from the study.

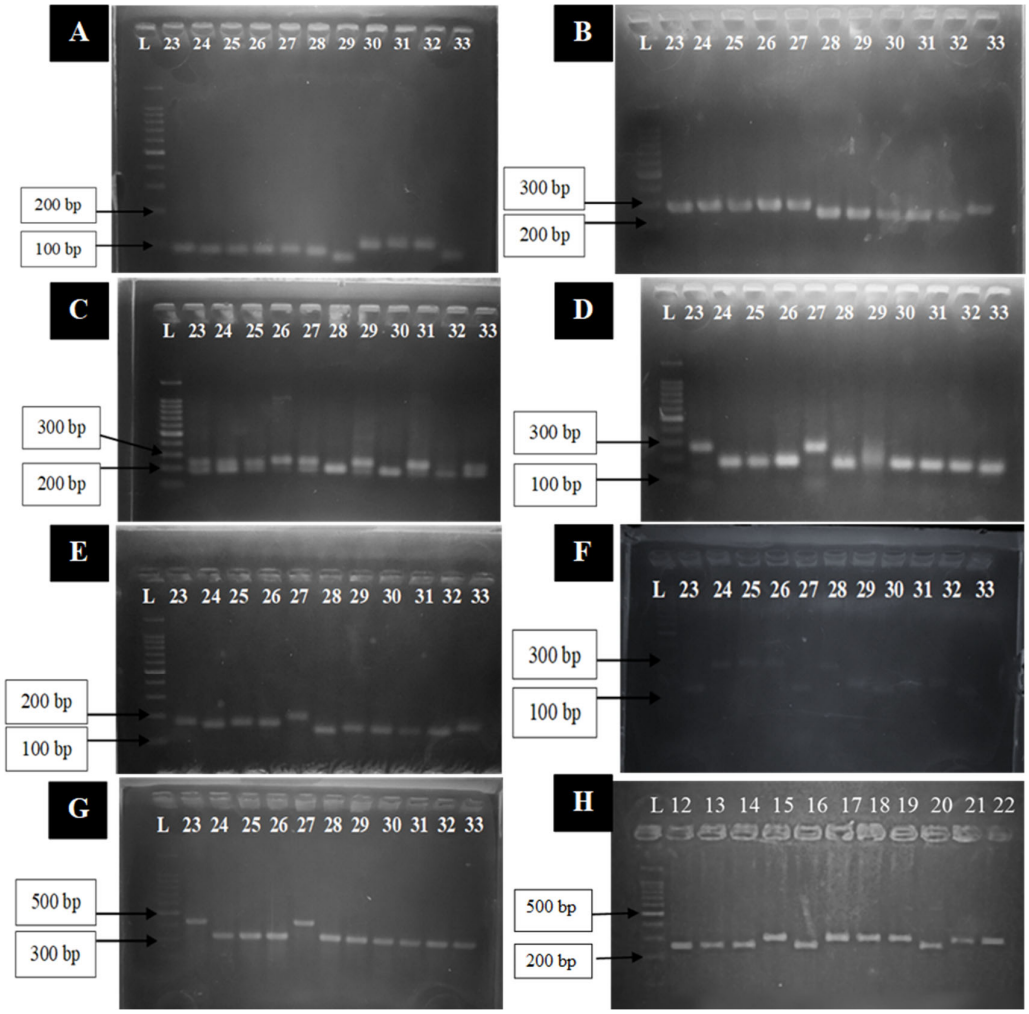

Figure 1. A Representative figure for the banding pattern of the seven polymorphic SSR markers. A: RM1, B: RM489, C: RM552, D: RM444, E: RM257, F: RM481, G: RM166, H: RM164. L= 100 bp ladder. The number represents UNIMAS collection number.

PIC Values and Genetic Diversity

In this study, the highest PIC value (0.9316) was recorded for RM257 (Table 1). The lowest PIC value (0.4029) was recorded for RM552 and the average PIC value for 44 accessions was 0.6891. About 87.5% of the SSR markers (7 out of 8 SSR markers) used in this study showed PIC values higher than 0.5 indicating markers used in the study were highly informative.

Table 2 summarizes the genetic analysis of the two rice populations. The observed number of effective alleles (*Ne*) in upland rice population recorded 1.566 and 1.543 in lowland rice population.

Nei's gene diversity (*h*) in upland rice and lowland rice population was 0.329 and 0.318, respectively. Shannon's Information Index (*I*) recorded 0.490 in upland rice population, and 0.475 in lowland rice population. Analysis of Molecular Variance indicated a 20% of genetic variation among population. Meanwhile, genetic variation within the population was 80%.

Table 1. SSR markers with Polymorphic Information Content.

SSR marker	Number of alleles	PIC
RM1	3	0.5072
RM489	2	0.7371
RM552	2	0.4029
RM444	2	0.7537
RM257	4	0.9316
RM481	3	0.864
RM166	2	0.6984
RM164	2	0.6178
Average	2.5	0.6891

Note. PIC: Polymorphic Information Content

Table 2. Summary of genetic analysis of upland and lowland accessions. The values indicate the means ± standard error of the 22 accessions of the respective upland and lowland rice population.

Population	*Ne*	*h*	*I*
Upland	1.566 ± 0.075	0.329 ± 0.036	0.490
Lowland	1.543 ± 0.077	0.318 ± 0.037	0.475

Note. Ne: Number of effective alleles; *h*: Nei's gene diversity; *I*: Shannon's Information Index

Cluster Analysis Based on SSR Markers

The UPGMA dendrogram was constructed for the 44 accessions using the MEGA X v.10.1 program based on the eight SSR markers genotype. The UPGMA dendrogram classified the 44 rice accessions into two primary clusters, Cluster I and Cluster II. Cluster I was divided into two sub-clusters, sub-cluster IA and sub-cluster IB and Cluster II was splat into sub-cluster IIA and sub-cluster IIB (Figure 2).

As tabulated in Table 3, sub-cluster IA had the highest number of accessions with 19 rice accessions. Sub-cluster IB comprises of seven accessions. Meanwhile, sub-cluster IIA consisted of six accessions and the rest were grouped in sub-cluster IIB. The dendrogram tree did not show clear clustering pattern according to their morphological traits described by Frank Clifton *et al*. (2024), nor according to the division where the seeds were sampled. The clustering was also not clearly according to the rice type (upland vs lowland).

***Maturase-K* Sequence Characteristics and Genetic Diversity**

Bright and clear 1500 bp band of the *matK* gene fragment were successfully amplified from the 44 rice accessions (Figure 3). The aligned sequences resulted in a data matrix of 1470 bp with no alignment gaps. The number of segregating sites showed a value of 140.

Phylogenetic Analysis Based on *Maturase-K*

The phylogenetic tree constructed based on *matK* sequence had one big cluster (Figure 4). All the accessions in this study were grouped in Cluster I with 100% bootstrap value. Also in the same cluster, there were 94 accessions of *O. sativa* which mostly originated from India (88 accessions), followed by Australia (4 accessions), China (1 accession) and one from unknown origin. Other than *O. sativa*, *O. rufipogon* was grouped together in the big cluster (Figure 4; Table 4). The outgroup consists of four species, *viz O. punctata*, *O. granulata*, O. *australiensis*, and *O. officinalis* (Figure 4; Table 4).

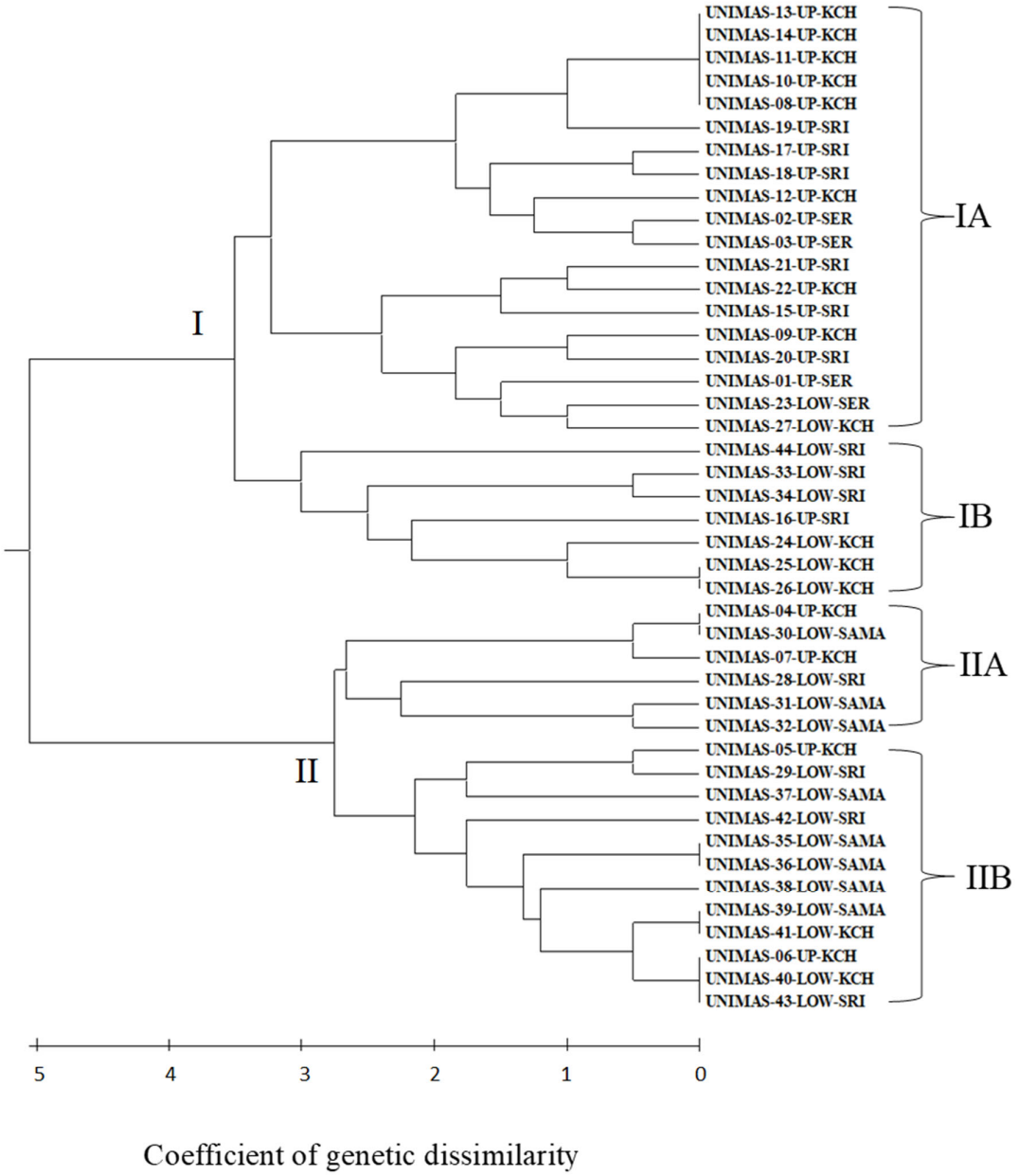

Figure 2. The cluster analysis of 44 rice accessions collected from North-West region of Sarawak based on eight SSR markers. I: Cluster I; IA: Sub-cluster IA; IB: Sub-cluster IB; II: Cluster II; IIA: Sub-cluster IIA; IIB: Sub-cluster IIB. The uppercase abbreviation after the accession number indicates the division where seeds were collected. KCH: Kuching; SER: Serian; SAMA: Kota Samarahan; SRI: Sri Aman.

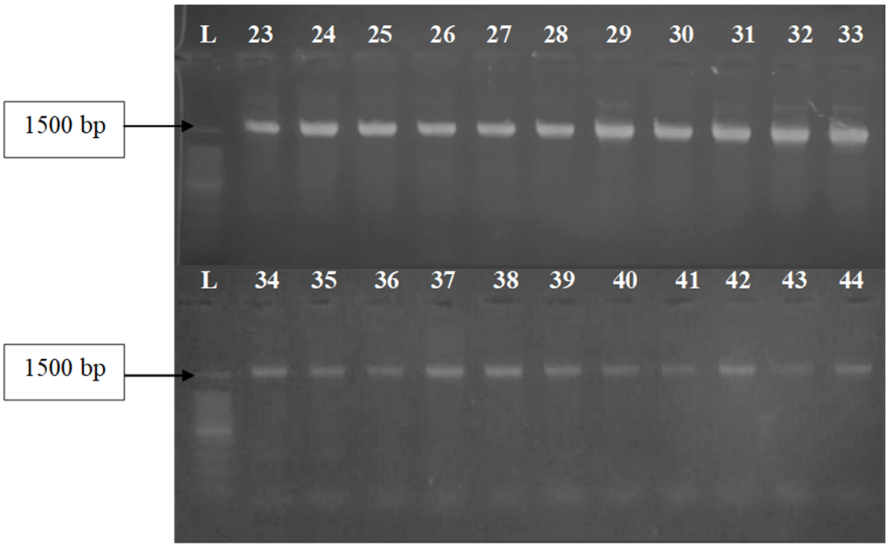

Figure 3. Representative amplification of 22 rice accessions employing *matK* gene specific primer. L= 100 bp ladder. The number represents the 22 UNIMAS collections from UNIMAS-23 to UNIMAS-44.

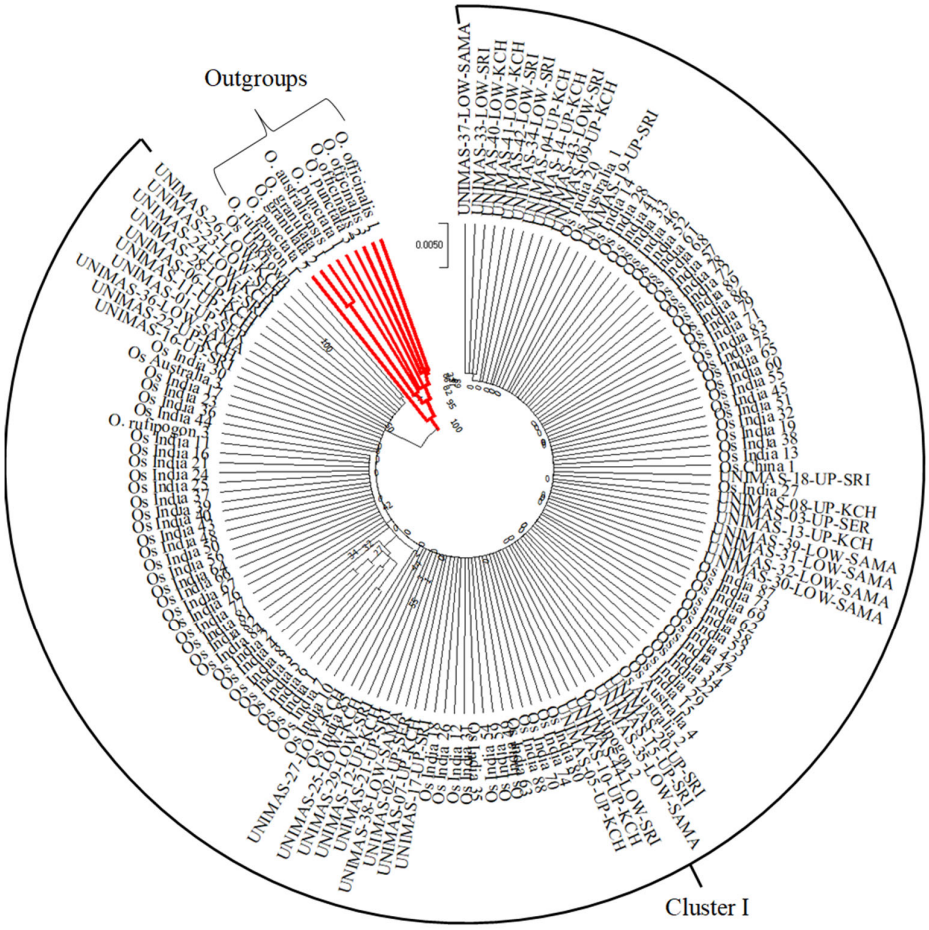

Figure 4. The phylogenetic tree of 44 rice accessions based on *Maturase-K* sequence, constructed using Maximum Likelihood Analysis. Best-fit substitution models: Tamura 3-Parameter Model with discrete gamma distribution evolutionary rates among sites (T92+G) at 1000 replicates bootstrap values. The red line indicates the outgroups.

Table 3. Clustering of the 44 rice accessions based on the eight SSR markers

Cluster	Sub-cluster	Number of accession	Accession
I	IA	19	UNIMAS-13-UP-KCH UNIMAS-14-UP-KCH UNIMAS-11-UP-KCH UNIMAS-10-UP-KCH UNIMAS-08-UP-KCH UNIMAS-19-UP-SRI UNIMAS-17-UP-SRI UNIMAS-18-UP-SRI UNIMAS-12-UP-KCH UNIMAS-02-UP-SER UNIMAS-03-UP-SER UNIMAS-21-UP-SRI UNIMAS-22-UP-KCH UNIMAS-15-UP-SRI UNIMAS-09-UP-KCH UNIMAS-20-UP-SRI UNIMAS-01-UP-SER UNIMAS-23-LOW-SER UNIMAS-27-LOW-KCH
	IB	7	UNIMAS-44-LOW-SRI UNIMAS-33-LOW-SRI UNIMAS-34-LOW-SRI UNIMAS-16-UP-SRI UNIMAS-24-LOW-KCH UNIMAS-25-LOW-KCH UNIMAS-26-LOW-KCH
II	IIA	6	UNIMAS-04-UP-KCH UNIMAS-30-LOW-SAMA UNIMAS-07-UP-KCH UNIMAS-28-LOW-SRI UNIMAS-31-LOW-SAMA UNIMAS-32-LOW-SAMA
	IIB	12	UNIMAS-05-UP-KCH UNIMAS-29-LOW-SRI UNIMAS-37-LOW-SAMA UNIMAS-42-LOW-SRI UNIMAS-35-LOW-SAMA UNIMAS-36-LOW-SAMA UNIMAS-38-LOW-SAMA UNIMAS-39-LOW-SAMA UNIMAS-41-LOW-KCH UNIMAS-06-UP-KCH UNIMAS-40-LOW-KCH UNIMAS-43-LOW-SRI

Note. The uppercase abbreviation after the accession number indicates the location of the collection. KCH: Kuching; SER: Serian; SAMA: Kota Samarahan; SRI: Sri Aman

Table 4. Clusters for 44 upland and lowland rice accessions based on *Maturase-K* marker

Cluster	Species	Country of Origin	No of Accession
I	*Oryza sativa*	India	88
	Oryza sativa	Australia	4
	Orya sativa	China	1
	Oryza sativa	Unknown	1
	*Oryza sativa***	Malaysia	22
	*Oryza sativa**	Malaysia	22
	Oryza rufipogon	Unknown	1
	Oryza rufipogon	China	1
	Oryza rufipogon	India	1
Outgroups	*Oryza punctata*	Unknown	1
	Oryza puntata	Unknown	1
	Oryza granulata	Unknown	1
	Oryza granulata	China	1
	Oryza australiensis	Vietnam	1
	Oryza officinalis	Australia	1
	Oryza officinalis	Phillipines	1
	Oryza officinalis	China	1
	Oryza punctata	Papua New Guinea	1

Note. **Upland rice accessions from UNIMAS collections, *Lowland rice accessions from UNIMAS collections.

DISCUSSION

Characterisation Based on Simple Sequence Repeat

Characterizing rice cultivar using a highly polymorphic microsatellite markers may provide a deeper information on rice germplasm and serves as powerful tools to detect genetic variation and genetic relationship within and among species (Al-Musawi *et al*., 2019). The SSR are markers of choice because SSR demonstrate a higher per-locus information content (García *et al*., 2018). Hodel *et al*. (2016) stated the utilisation of a number of SSR markers are inexpensive and can deliver good results.

A total of 20 alleles were successfully amplified by eight SSR markers in the present study, ranging from two (RM489, RM552, RM444, RM166 and RM164) to four (RM257) alleles per primer with an average of 2.5 alleles per marker. Similarly, Supari *et al*. (2019) and Wong *et al*. (2009) reported an average of 2.26 and 2.60 alleles per SSR locus, respectively, for their rice genotyping. The number of alleles amplified in the present study is, however, relatively lower compared to Sohrabi *et al*. (2013) and Aljumaili *et al*. (2018) which reported two to eleven alleles from the SSR markers used for genotyping their rice samples.

The PIC value serves as an indicator of allele diversity and frequency (Shete *et al*., 2000). In this study, the SSR markers displayed a range of PIC values from 0.4029 (RM552) to 0.9316 (RM257), averaging 0.6891. Among the eight polymorphic markers examined, only RM552 had a PIC value below 0.5, suggesting lower reliability in assessing genetic diversity according to Dalimunthe *et al*. (2020). The SSR markers with a PIC value above 0.5 are considered highly polymorphic and valuable for genetic research (Melaku *et al*., 2018). The high PIC value of RM257 suggests that this marker is particularly effective in evaluating genetic variability among rice accessions. Overall, the eight SSR markers used in this study had an average PIC value greater than 0.5, indicating

they were highly informative and capable of distinguishing between genotypes.

Following the assessment of PIC values, the study also examined the number of effective alleles (*Ne*), which reflects the number of alleles that actively contribute to genetic variation within a population. The upland rice population recorded a higher *Ne* of 1.566 compared to 1.543 in the lowland population. These *Ne* values are greater than those reported for aromatic rice accessions from Malaysia genotyped using SSR markers by Aljumaili *et al.* (2018), indicating a relatively higher effective allele diversity in the current populations.

The Nei's gene diversity (*h*) for markers in upland population was 0.329 and 0.318 in lowland population. A study done by Sohrabi *et al*. (2013) showed a much higher value of *h* in the author's upland rice populations originated from Peninsular Malaysia and Sabah, which was 0.62 and 0.58, respectively. According to Bobokashvili (2016), the *h* of value 0.95 to 1.00 indicates high divergence in a population. The *h* value ranging from 0.50 to 0.60 indicates two species were having similar features, however, unable to cross each other. Two subspecies having very similar features in range of one species were likely to have *h* value ranging from 0.17 to 0.22. This indicates low level of divergence in the 44 accessions of the current study.

The Shannon's Information Index (*I*) based on SSR markers in upland and lowland rice population were 0.490 and 0.475, respectively. Sohrabi *et al*. (2013) reported much higher value of *I* with an average of 1.1653 for the author's rice populations from Peninsular Malaysia and 1.0754 for Sabah. The recorded *I* value in both populations of the current study indicates a low diversity in comparison with the rice accessions from Peninsular Malaysia and Sabah by Sohrabi *et al*. (2013). Higher number of SSR markers were tested in Sohrabi *et al*. (2013) compared to the present study. Several studies highlight the effect of the number of markers on *I* in rice, emphasizing that a higher number of markers often results in better detection of genetic diversity. Study of rice landraces, such as those from the Honghe Hani Rice Terraces (Ma *et al*., 2023), used over 200 SSR markers and reported significant genetic diversity, with Shannon's Index values averaging around 1.08. This indicates that a higher number of markers can capture a broader range of allelic variation, contributing to a more robust Shannon Index value.

The Simple Sequence Repeat Cluster Analysis

The clustering analysis did not reveal clear grouping of the lowland and upland rice accessions based on their morphological traits described by Frank Clifton *et al*. (2024), geographical origin or rice type. This lack of clear clustering may be due in part to the small number of SSR markers used in this study, which was only eight markers. Previous studies have used between 32 and 140 markers (Zhang *et al*., 2010; Ma *et al*., 2016; Aljumaili *et al*., 2018; Hassan and Hama-Ali, 2022). Using fewer markers limits the ability to detect small genetic differences among rice accessions, which can reduce clustering accuracy and weaken conclusions about genetic diversity and relationships (Wang *et al*., 2021). In addition, seed mixing in local farming practices may have contributed to the genetic overlap observed, as seeds from different accessions can be unintentionally combined and this blurs distinct groupings. The observed low genetic diversity is also consistent with morphological findings reported by Frank Clifton *et al*. (2024), where some rice accessions displayed similar traits despite differences in their names or origins. Taken together, these factors of limited marker coverage, possible seed mixing and genuinely low genetic diversity help to explain why genetic clustering in this study did not correspond clearly with morphological traits or geographic origin. Future research should use a larger number of SSR markers or combine different genetic markers to improve resolution and better capture the genetic variation among rice varieties. This could include not only more SSR markers but also different types of genetic markers such as single nucleotide polymorphisms (SNPs), amplified fragment length polymorphisms (AFLPs) or insertions/deletions (InDels), which together can capture a wider range of genetic differences among rice varieties.

Accessions that shared the same name (Padi Merjat) but collected from different villages were found to be genetically different based on SSR genotype as shown in upland rice

accessions namely UNIMAS-04, UNIMAS-05, UNIMAS-06, and UNIMAS-07 (Kampung Mambong) and UNIMAS-22 (Kampung Tabuan Rabak; Table 3; Supplemetary Table 1). Moreover, it can be seen that UNIMAS-17 and UNIMAS-18 from upland rice accessions and UNIMAS-30 from lowland rice accessions were grouped in two different clusters (Cluster IA and Cluster IIA) although both had the same name (Padi Bario).

From the dendrogram obtained using UPGMA cluster analysis of SSR (Figure 2), three lowland rice accessions namely UNIMAS-42, UNIMAS-43 and UNIMAS-44, which derived from a same seed source named Padi Kanowit (Supplemetary Table 1), were separated in different cluster. UNIMAS-42 and UNIMAS-43 were grouped together in Cluster IIB but UNIMAS-44 was separated into Cluster IB. This indicates UNIMAS-44 was genetically different in comparison with UNIMAS-42 and UNIMAS-43 (Table 3). This might be a result from seed impurity or mixture of seed provided by the farmers due to common practice by planting several rice varieties in one field or planting them close to each other without a clear boundary which contribute to diversity in the landrace (Yeo *et al*., 2018).

According to Yeo *et al*. (2018), the registration of a variety based on the name given by farmers may not serve as a reliable basis for registration. There is a possibility for a variety with a similar name, exhibiting distinct morphological characteristics. To support this, the upland accessions despite being given the same name by local farmers, there were noticeable differences in seed and grain color (Frank Clifton *et al*., 2024). Meanwhile the lowland rice varieties in Frank Clifton *et al*. (2024) showed various morphological variations. These include differences in panicle type, ligule shape, seed color, secondary branching, and grain shape. Some accessions exhibited variation in both panicle type and ligule, while others showed differences in secondary branching and grain color.

Genetic Relatedness Based on *Maturase-K*

Phylogenetic analysis based on *matK* has clustered the 22 upland accessions and 22 lowland accessions along with 94 *O. sativa* accessions from different origins and three *O. rufipogon* accessions into one major cluster with 100% bootstrap. Accessions in Cluster I were seen not clustered together with four other *Oryza* species (*O. punctata, O. granulata, O. officinalis* and *O. australiensis*) which marked as outgroups.

The clustering of the upland 22 rice accessions and 22 lowland rice accessions were random. The clustering based on *matK* gene sequence of this study was based on genome rather than geographical or morphological characteristics. *Oryza sativa* and *O. rufipogon* which have AA genome were clustered in one clade but separated from *Oryza* of different genomes [*O. punctata* (BB genome), *O. officinalis* (CC genome), *O. australiensis* (EE genome) and *O. granulata* (GG genome)]. Similar results were obtained by Zodinpuii *et al*. (2013) and Patil *et al*. (2015) in their study by utilizing *matK* gene for the phylogenetic relationships of the *Oryza* genus. *Oryza sativa* and *O. rufipogon* in their study were also grouped in one clade, supporting the previous hypothesis of an Asian origin of *O. sativa*, and *O. rufipogon* as the progenitor (Khush, 1997). This indicates that *matK* marker alone was unable to distinguish closely related species.

The *matK* gene, while commonly used for plant DNA barcoding, often lacks sufficient resolution to distinguish closely related species within the *Oryza* genus. This limitation arises because *matK* is a relatively conserved plastid gene, meaning its DNA sequence evolves slowly and shows little variation among species that have recently diverged (Fazekas *et al*., 2008; Hollingsworth *et al*., 2009). As a result, *matK* may not capture enough genetic differences to separate closely related *Oryza* species effectively. Several studies by Zodinpuii *et al*. (2013) and Patil *et al*. (2015) have reported similar findings, where *matK*-based phylogenetic trees clustered *Oryza* species broadly but failed to discriminate subspecies or varieties clearly. To overcome this, combining *matK* with other more variable genetic markers such as intergenic spacers or nuclear genes is recommended, as this approach can increase discriminatory power and provide a more detailed resolution of species relationships.

A combination of barcoding markers from the plastid genome and intergenic spacers should be considered to improve the resolution of the

phylogenetic tree for a higher discrimination power. Such proposition is based on De Mattia *et al*. (2011), where the author manages to discriminate *Ocimum basilicum* from other *Ocimum* species up to cultivar level by using the combination of *matK* and *trnH-psbA* genes. The hypothesis is supported by Ho *et al*. (2021) on phylogenetic analysis of jewel orchid accessions by using *matK* and *ribulose 1,5-biphosphate carboxylase*, where the analysis revealed the combination of the two barcoding markers have higher discrimination power in comparison to analysis based on *matK* gene alone.

CONCLUSION

The 44 upland and lowland rice accessions collected from the North-West region of Sarawak were successfully genotyped using SSR markers and analyzed with the *matK* barcoding marker. The SSR analysis revealed low genetic diversity among the accessions, which may reflect either the limited number of markers used or a narrow genetic base arising from shared ancestry among local landraces. Phylogenetic analysis based on the *matK* gene grouped accessions primarily by genome type rather than by geographic origin or morphological traits.

The low genetic diversity observed among these local landraces carries important implications for both conservation and rice breeding efforts in Sarawak. On one hand, this genetic uniformity may reflect long term adaptation to local environmental conditions, suggesting a certain degree of genetic stability. On the other hand, it also indicates a limited capacity to respond to emerging challenges such as climate change, pest outbreaks and shifting agronomic needs. This potential vulnerability highlights the importance of conserving the remaining genetic variation through both in-situ and *ex situ* approaches to safeguard against further genetic erosion. Furthermore, the narrow genetic base underscores the need to introduce new, diverse germplasm into local breeding programs. Expanding the genetic pool can support the development of rice varieties with greater resilience, adaptability and productivity for sustainable cultivation in Sarawak.

To enhance the resolution and accuracy of genetic diversity assessments, future studies should consider expanding the SSR marker set or incorporating SNP markers. These approaches will provide finer genetic resolution, improve classification accuracy and strengthen molecular-assisted selection strategies in breeding programs. Additionally, combining *matK* with other barcoding markers such as *trnL-trnF* and *rbcL* can improve species discrimination and provide a more comprehensive understanding of the genetic relationships among accessions. Collectively, these strategies can support the development of effective germplasm conservation plans and guide the selection of parental lines for sustainable rice cultivation in Sarawak.

ACKNOWLEDGEMENTS

This study was funded by Ministry of Higher Education, Malaysia under Fundamental Research Grant Scheme (FRGS/1/2017/STG03/UNIMAS/03/2). The authors would like to acknowledge Universiti Malaysia Sarawak, for the facilities provided.

REFERENCES

Adriansyah, F., Hasmeda, M., Suwignyo, R.A., Halimi, E.S. & Sarimana, U. (2021). Genetic diversity and relationship of South Sumatran local rice and its backcrossed lines based on the *matK* gene. *Journal of Breeding & Genetics*, 53(3): 499-509.

Akagi, H., Yokozeki, Y., Inagaki, A. & Fujimura, T. (1997). Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. *Theoretical and Applied Genetics*, 94: 61-67. DOI: 10.1007/s001220050382

Aljumaili, S.J., Rafii, M.Y., Latif, M.A., Sakimin, S.Z., Arolu, I.W. & Miah, G. (2018). Genetic diversity of aromatic rice germplasm revealed by SSR markers. *BioMed Research International*, 2018: 1-11. DOI: 10.1155/2018/7658032

Al-Musawi, B.H., Al-Bdairi, N.A.H. & Al-Anbari, M.A. (2019). Sequence variation and phylogenetic relationship among rice (*Oryza sativa* L.) genotypes in Iraq. *American Institute of Physics Conference Proceedings,* 2144(1): 040001. DOI: 10.1063/1.5123102

Anderson, A., Churchill, G.A., Autrique, J.E., Tanksley, S.D. & Sorrells, M.E. (1993). Optimizing parental selection for genetic linkage

maps. *Genome*, 36(1): 181-186. DOI: 10.1139/g93-024

Awika, J.M. (2011). *Major cereal grains production and use around the world (Advances in cereal science: implications to food processing and health promotion)*. United States, American Chemical Society. DOI: 10.1021/bk-2011-1089.ch001

Bobokashvili, Z. (2016). *Re: How to interpret or infer Nei's phenotypic diversity index?* Retrieved December 10, 2022, from https://www.researchgate.net/post/How_to_interprete_or_infer_Neis_phenotypic_diversity_index.

Dalimunthe, S.R., Siregar, L.A.M., Putri, L.A.P., Cirunnisa, T. & Hairmansis, A. (2020). Polymorphism levels of some SSR markers (Simple Sequence Repeat) for parental line identification on low temperature tolerance. *Institute of Physics Conference Series: Earth and Environmental Science*, *454*(1): 012165. DOI: 10.1088/1755-1315/454/1/012165

De Mattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M. & Labra, M. (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. *Food Research International*, 44(3): 693-702. DOI: 10.1016/j.foodres.2010.12.032

Dempewolf, H., Krishnan, S. & Guarino, L. (2023). Our shared global responsibility: Safeguarding crop diversity for future generations. *Proceedings of the National Academy of Sciences,* 120(14): e2205768119. DOI: 10.1073/pnas.2205768119

Department of Agriculture Sarawak. (2020). Agriculture Research Centre: Programmes. Retrieved November 20, 2021, from https://doa.sarawak.gov.my/page-0-0-415-AGRICULTURE-RESEARCH-CENTRE-PROGRAMMES.html

Doyle, J.J. & Doyle, D.J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry*, 19: 1115.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M. & Barrett, S.C.H. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE,* 3(7): e2802. DOI: 10.1371/journal.pone.0002802

Frank Clifton, Z., Yeo F.K.S., Sylvester Embuas, R.P.T., Kalu, M., Shabdin, Z. & Lai, L.S. (2024). Preliminary characterisation of lowland and upland rice from Sarawak, Malaysian Borneo. *Borneo Journal of Resource Science and Technology*, 14(1): 123-138. DOI: 10.33736/bjrst.5653.2024

García, C., Guichoux, E. & Hampe, A. (2018). A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. turbinata). *Tree Genetics & Genomes*, 14: 1-9. DOI: 10.1007/s11295-018-1301-x

Hassan, D.A. & Hama-Ali, E.O. (2022). Evaluation of gene flow and genetic diversty in rice accessions across Kurdistan region-Iraq using SSR Markers. *Molecular Biology Reports*, 49: 1007-1016.

Ho, V.T., Tran, T.K.P., Vu, T.T.T. & Widiarsih, S. (2021). Comparison of *matK* and *rbcL* DNA barcodes for genetic classification of jewel orchid accessions in Vietnam. *Journal of Genetic Engineering and Biotechnology*, 19(1): 93. DOI: 10.1186/s43141-021-00188-1

Hodel, R.G., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Liu, X. & Soltis, P.S. (2016). The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*, 4(6): 1600025. DOI: 10.3732/apps.1600025

Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L., Fazekas, A.J., Graham, S.W., James, K.E., Kim, K.J., Kress, W.J., Schneider, H., van Alphen Stahl, J., Barrett, S.C.H., van den Berg, C., Bogarin, D. & Little, D.P. (2009). A DNA barcode for land plants. *Botanical Journal of the Linnean Society,* 159(1): 1-16. DOI: 10.1073/pnas.0905845106

Khush, G.S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology*, 35: 25-34. DOI: 10.1023/A:1005810616885

Lee, H.H., Neoh, P.P.N., Bong, W.S.T., Puvaneswaran, J., Wong, S.C., Yiu, P.H. & Rajan, A. (2011). Genotyping of Sarawak rice cultivars using microsatellite markers. *Pertanika Journal of Tropical Agriculture Science*, 34: 123-136.

Ma, M., Liu, Y., Wang, T. & Lu, B. (2016). Cluster analysis of red rice based on SSR markers from Hani's terraced fields in Yunnan Province.

Advances in Computer Science Research, 63. DOI: 10.2991/aiea-16.2016.29

Ma, M., Lei, E., Wang, T., Meng, H., Zhang, W. & Lu, B. (2023). Genetic diversity and association mapping of grain-size traits in rice landraces from the Honghe Hani Rice Terraces system in Yunnan Province. *Plants,* 12(8): 1678.

Melaku, G., Zhang, S. & Haileselassie, T. (2018). Comparative evaluation of rice SSR markers on different *Oryza* species. *Journal of Rice Research and Developments*, 1(1): 38-48. DOI: 10.36959/973/418

Miah, G., Rafii, M.Y, Ismail, M.R., Puteh, A.B., Rahim, H.A., Islam, K.H. & Latif, M.A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International Journal of Molecular Sciences*, 14: 22499-22528.

Patil, R.G., Jadhao, K.R., Samal, K.C. & Rout, G.R. (2015). Molecular phylogeny of Indian indigenous aromatic rice based on sequence diversity of the chloroplast-encoded *matK* gene. *Rice Genomics and Genetics*, 6(8): 1-8. DOI: 10.5376/rgg.2015.06.0008

Shete, S., Tiwari, H. & Elston, R.C. (2000). On estimating the heterozygosity and polymorphism information content value. *Theoretical Population Biology*, 57(3): 265-271. DOI: 10.1006/tpbi.2000.1452

Smouse, P.E., Banks, S.C. & Peakall, R. (2017). Converting quadratic entropy to diversity: Both animals and alleles are diverse, but some are more diverse than others. *The Public Library of Science One*, 12(10): e0185499. DOI: 10.1371/journal.pone.0185499

Sohrabi, M., Rafii, M.Y., Hanafi, M.M. & Latif, M.A. (2013). Genetic divergence of Malaysian upland rice revealed by microsatellite markers. *Plant Omics Journal*, *6*(3): 175-182.

Sohrabi, M., Rafii, M.Y., Hanafi, M.M., Siti Nor Akmar, A. & Latif, M.A. (2012). Genetic diversity of upland rice germplasm in Malaysia based on quantitative traits. *The Scientific World Journal*, 2012(1), 1-9. DOI: 10.1100/2012/416291

Supari, N., Kaya, Y., Biroudian, M. & Javed, M. (2019). Molecular characterization of Malaysian rice cultivars using SSR markers. *American Institute of Physics Conference Proceedings,* 2155(1): 020016. DOI: 10.1063/1.5125520

Tan, C.S., Teo, G.K., Jamadon, B. & Tan P.H. (2006). *Microsatellite markers for differentiation of local Bario rice varieties*. Retrieved August 15, 2022, from: https://doa.sarawak.gov.my/web/attachment/show/?docid=https://doa.sarawak.gov.my/web/attachment/show/?docid=bGlYVDZsV3hTU1NkVUQyTHBSOG9EQT09Ojr-_vyzGFP5599ncaaN5uaQ

Wang, H., Yang, B., Wang, H. & Xiao, H. (2021). Impact of different numbers of microsatellite markers on population genetic results using SLAF-seq data for *Rhododendron* species. *Scientific Reports*, 11(1): 8597. DOI:10.1038/s41598-021-87945-x

Wong, S.C., Yiu, P.H., Bong, S.T.W., Lee, H.H., Neoh, P.N.P. & Rajan, A. (2009). Analysis of Sarawak Bario rice diversity using microsatellite markers. *American Journal of Agricultural and Biological Sciences*, 4(4): 298-304.

Yao, X., Tan, Y.H., Liu, Y.Y. & Song, Y. (2019). Molecular markers for species identification of medicinal plants: A review. *Evidence-Based Complementary and Alternative Medicine*, 2019: 8746802. DOI: 10.1155/2019/8746802

Yeo, F.K.S., Meekiong, K., Shabdin, Z., Mohamad, N.K., Hussin, N.A. & Chung, H.H. (2018). Diversity of rice in Kampung Lebor, Serian – A first insight. In Yeo, F.K.S., Chong, Y.L. & Khan, F.A.A. (eds.). *Glimpses of Bornean Biodiversity.* Kuching, Malaysia, Universiti Malaysia Sarawak Publisher. pp. 155-167.

Zhang, C.H., Li, J.Z., Zhu, Z., Zhang, Y.D., Zhao, L. & Wang, C.L. (2010). Cluster analysis on japonica rice (*Oryza sativa* L.) with good eating quality based on SSR markers and phenotypic traits. *Rice Science*, 17(2): 111-121. DOI: 10.1016/S1672-6308(08)60113-4

Zhu, Y.F., Qin, G.C., Yang, W., Wang, J.C. & Zhu, S.J. (2012). Fingerprinting and variety identification of rice (*Oryza sativa* L.) based on Simple Sequence Repeat markers. *Plant Omics Journal*, 5(4): 421-426.

Zodinpuii, D., Ghatak, S., Mukherjee, S. & Kumar, N.S. (2013). Genetic relatedness of genus *Oryza* from Eastern Himalayan region as revealed by chloroplast *matK* gene. *Asian Journal of Conservation Biology*, 2(2): 144-151.

Published in Borneo Journal of Resource Science and Technology (2025), 15(2): 80-93

SUPPLEMENTARY MATERIALS

Molecular Characterisation of Upland and Lowland Rice from Sarawak, Malaysian Borneo

FREDDY KUOK SAN YEO[1*], RENEE PRISCILLA TRAWAS SYLVESTER EMBUAS[1], ZAZEVIA FRANK CLIFTON[1], MEEKIONG KALU[1], ZINNIRAH SHABDIN[1] & LEE SAN LAI[2]

[1]Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Jalan Datuk Mohammad Musa, 94300, Kota Samarahan, Sarawak, Malaysia; [2]Agriculture Research Centre Semongok, KM20, Jalan Puncak Borneo, 93250, Kuching, Sarawak
*Corresponding author: yksfreddy@unimas.my

SUPPLEMENTARY DATA

Supplementary Table 1.
List of landraces with its locality and GPS

Upland Landraces					Lowland Landraces				
Name	Accession	Location	GPS	Division	Name	Accession	Location	GPS	Division
Padi Belawi Pandan	UNIMAS-01, UNIMAS-02	Kg Paon Gahat	0° 57' 0" North, 110° 39' 0" East	Serian	Padi Pandan	UNIMAS-23	Kg Paon Gahat	0° 57' 0" North, 110° 39' 0" East	Serian
Padi Belawi	UNIMAS-03	Kg Paon Gahat	0° 57' 0" North, 110° 39' 0" East	Serian	Padi Hitam	UNIMAS-24, UNIMAS-25, UNIMAS-26	Kg Pueh	1° 51' 0" North, 109° 41' 0" East	Kuching
Padi Merjat	UNIMAS-04, UNIMAS-05, UNIMAS-06, UNIMAS-07	Kg Mambong	1° 22' 0" North, 110° 21' 0" East	Kuching	Padi Merah	UNIMAS-27	Kg Pesak	1° 45' 0" North, 110° 46' 0" East	Kuching
Padi Pandan Wangi	UNIMAS-08, UNIMAS-09, UNIMAS-10, UNIMAS-11, UNIMAS-12, UNIMAS-13, UNIMAS-14	Kg Pesak	1° 45' 0" North, 110° 46' 0" East	Kuching	Padi Bajong	UNIMAS-28	Kg Panggil	1°05'48" North, 111°23'23" East	Sri Aman
Padi Limbang	UNIMAS-15	Kg Sg Tenggang	1° 04' 48" North, 111° 03' 28" East	Sri Aman	Padi Arang	UNIMAS-29	Kg Sg Tenggang	1° 04' 48" North, 111° 03' 28" East	Sri Aman
Padi Chelum	UNIMAS-16	Kg Sg Tenggang	1° 04' 48" North, 111° 03' 28" East	Sri Aman	Padi Bario	UNIMAS-30	Kg Baru	1° 33' 28" North, 110° 22' 22" East	Kota Samarahan
Padi Bario	UNIMAS-17, UNIMAS-18	Kg Sg Tenggang	1° 04' 48" North, 111° 03' 28" East	Sri Aman	Padi Sabak Hitam	UNIMAS-31, UNIMAS-32	Kg Baru	1° 33' 28" North, 110° 22' 22" East	Kota Samarahan
Padi Badawi	UNIMAS-19	Kg Sg Tenggang	1° 04' 48" North, 111° 03' 28" East	Sri Aman	Padi Selasih	UNIMAS-33, UNIMAS-34	Kg Melugu Skim	1°07'10" North, 111°25'33" East	Sri Aman

Supplementary Table 1. *(continue)*

Padi Sempang	UNIMAS-20	Kg Melugu Tengah	1° 06'14" North, 111°23'57" East	Sri Aman	Padi Sabak Angin	UNIMAS-35, UNIMAS-36, UNIMAS-37, UNIMAS-38, UNIMAS-39	Kg Baru	1° 33' 28" North, 110° 22' 22" East	Kota Samarahan
Padi Mawang	UNIMAS-21	Kg Melugu Tengah	1° 06' 14" North, 111° 23' 57" East	Sri Aman	Padi Kihuai	UNIMAS-40, UNIMAS-41	Kg Mambong	1° 22' 0" North, 110° 21' 0" East	Kuching
Padi Merjat	UNIMAS-22	Kg Tabuan Rabak	1° 18' 39" North, 110° 20' 59" East	Kuching	Padi Kanowit	UNIMAS-42, UNIMAS-43, UNIMAS-44	Kg Melugu Skim	1° 07' 10" North, 111° 25' 33" East	Sri Aman

Supplementary Table 2.
The 39 sets of SSR markers with sequences and details

SSR	Primer sequence	Ch	EPS (bp)	Ta (°C)	Source
RM1	F: GCGAAAACACAATGCAAAAA R: GCGTTGGTTGGACCTGAC	1	90-120	55	Gramene (https://archive.gramene.org/markers/microsat/all-ssr.html)
RM226	F: GAAGCTAAGGTCTGGGAGAAACC R: AATGGCCTTAACCAAGTAGGATGG	1	280-350	55	
RM279	F: GCGGGAGAGGGATCTCCT R: GGCTAGGAGTTAACCTCGCG	2	170-190	55	
RM207	F: CCATTCGTGAGAAGATCTGA R: CACCTCATCCTCGTAACGCC	2	N/A	55	
RM489	F: ACTTGAGACGATCGGACACC R: TCACCCATGGATGTTGTCAG	3	250-270	55	
RM514	F: AGATTGATCTCCCATTCCCC R: CACGAGCATATTACTAGTGG	3	N/A	55	
RM335	F: GTACACACCCACATCGAGAAG R: GCTCTATGCGAGTATCCATGG	4	100-200	55	
RM567	F: ATCAGGGAAATCCTGAAGGG R: GGAAGGAGCAATCACCACTG	4	280-400	55	
RM317	F: CATACTTACCAGTTCACCGCC R: CTGGAGAGTGTCAGCTAGTTGA	4	160	55	
RM229	F: CACTCACACGAACGACTGAC R: CGCAGGTTCTTGTGAAATGT	4	100-120	55	
RM166	F: GGTCCTGGGTCAATAATTGGGTTACC R: TTGCTGCATGATCCTAAACCGG	4	321-410	61	
RM108	F: TCTCTTGCGCGCACACTGGCAC R: CGTGCACCACCACCACCACCAC	4	N/A	67	
RM413	F: GGCGATTCTTGGATGAAGAG R: TCCCCACCAATCTTGTCTTC	5	N/A	53	
RM480	F: GCTCAAGCATTCTGCAGTTG R: GCGCTTCTGCTTATTGGAAG	5	N/A	58	

Supplementary Table 2. *(continue)*

SSR	Primer sequence	Ch	EPS (bp)	Ta (°C)	Source
RM164	F: TCTTGCCCGTCACTGCAGATATCC R: GCAGCCCTAATGCTACAATTCTTC	5	246-300	58	Gramene (https://archive.gramene.org/markers/microsat/all-ssr.html)
RM250	F: GGTTCAAACCAAGCTGATCA R: GATGAAGGCCTTCCACGCAG	5	170-190	57	
RM469	F: AGCTGAACAAGCCCTGAAAG R: GACTTGGGCAGTGTGACATG	6	100-120	55	
RM454	F: CTCAAGCTTAGCTGCTGCTG R: GTGATCAGTGCACCATAGCG	6	N/A	55	
RM276	F: CTCAACGTTGACACCTCGTG R: TCCTCCATCGAGCAGTATCA	6	100-180	59	
RM125	F: ATCAGCAGCCATGGCAGCGACC R: AGGGGATCATGTGCCGAAGGCC	7	130-600	63	
RM455	F: AACAACCCACCACCTGTCTC R: AGAAGGAAAAGGGCTCGATC	7	140	55	
RM481	F: TAGCTAGCCGATTGAATGGC R: CTCCACCTCCTATGTTGTTG	7	190-300	59	
RM47	F: ACTCCACTCCACTCCCCAC R: GTCAGCAGGTCGGACGTC	7	N/A	59	
RM152	F: GAAACCACCACACCTCACCG R: CCGTAGACCTTCTTGAAGTAG	8	140	53	
RM447	F: CCCTTGTGCTGTCTCCTCTC R: ACGGGCTTCTTCTCCTTCTC	8	100-600	53	
RM210	F: TCACATTCGGTGGCATTG R: CGAGGATGGTTGTTCACTTG	8	120-180	57	
RM444	F: GCTCCACCTGCTTAAGCATC R: TGAAGACCATGTTCTGCAGG	9	190-280	55	
RM215	F: CAAAATGGAGCAGCAAGAGC R: TGAGCACCTCCTTCTCTGTAG	9	150	56.	

Supplementary Table 2. *(continue)*

SSR marker	Primer sequence	Ch	EPS (bp)	Ta (°C)	Source
RM244	F: CCGACTGTTCGTCCTTATCA R: CTGCTCTCGGGTGAACGT	10	N/A	55	Gramene (https://archive.gramene.org/markers/microsat/all-ssr.html)
RM228	F: CTGGCCATTAGTCCTTGG R: GCTTGCGGCTCTGCTTAC	10	N/A	55	
RM552	F: CGCAGTTGTGGATTTCAGTG R: TGCTCAACGTTTGACTGTCC	11	200-210	55	
RM206	F: CCCATGCGTTTAACTATTCT R: CGTTCCATCGATCCGTATGG	11	N/A	58.1	
RM257	F: CAGTTCCGAGCAAGAGTACTC R: GGATCGGACGTGGCATATG	11	150-200	55	
RM224	F: ATCGATCGATCTTCACGAGG R: TGCTATAAAAGGCATTCGGG	11	150-200	55	
RM21	F: ACAGTATTCCGTAGGCACGG R: GCTCCATGAGGGTGGTAGAG	11	140-180	51	
RM247	F: TAGTGCCGATCGATGTAACG R: CATATGGTTTTGACAAAGCG	12	140-190	55	
RM17	F: TGCCCTGTTATTTTCTTCTCTC R: GGTGATCCTTTCCCATTTCA	12	190-220	55	
RM519	F: AGAGAGCCCCTAAATTTCCG R: AGGTACGCTCACCTGTGGAC	12	120-140	55	
RM561	F: GAGCTGTTTTGGACTACGGC R: GAGTAGCTTTCTCCCACCCC	N/A	N/A	55	

Note. SSR: Simple Sequence Repeat; F: Forward primer; R: Reverse primer; Ch: Chromosome; EPS: Expected product size in base pair; Ta: Annealing temperature in degree Celsius; N/A indicates data not available.

Supplementary Table 3.
The accession number for *matK* sequences downloaded from NCBI

Species	Identification Code	Country of Origin	GenBank Accession Number
Oryza sativa	OS India 1	India	KT894765.1
Oryza sativa	OS India 2	India	KT894764.1
Oryza sativa	OS India 3	India	KT894763.1
Oryza sativa	OS India 4	India	KT894762.1
Oryza sativa	OS India 5	India	KT894761.1
Oryza sativa	OS India 6	India	KT894760.1
Oryza sativa	OS India 7	India	KT894759.1
Oryza sativa	OS India 8	India	KT894758.1
Oryza sativa	OS India 10	India	KT894756.1
Oryza sativa	OS India 11	India	KF731111.1
Oryza sativa	OS India 12	India	KF731110.1
Oryza sativa	OS India 13	India	KF731109.1
Oryza sativa	OS India 14	India	KF731108.1
Oryza sativa	OS India 15	India	KF731107.1
Oryza sativa	OS India 16	India	KF731106.1
Oryza sativa	OS India 17	India	KF731105.1
Oryza sativa	OS India 18	India	KF731104.1
Oryza sativa	OS India 19	India	KF731103.1
Oryza sativa	OS India 20	India	KF731102.1
Oryza sativa	OS India 21	India	KF731101.1
Oryza sativa	OS India 22	India	KF731100.1
Oryza sativa	OS India 23	India	KF731099.1
Oryza sativa	OS India 24	India	KF731098.1
Oryza sativa	OS India 25	India	KF731097.1
Oryza sativa	OS India 26	India	KF731096.1
Oryza sativa	OS India 27	India	KF731095.1
Oryza sativa	OS India 28	India	KF731094.1
Oryza sativa	OS India 29	India	KF731093.1
Oryza sativa	OS India 30	India	KF731092.1

Supplementary Table 3. *(continue)*

Species	Identification Code	Country of Origin	GenBank Accession Number
Oryza sativa	OS India 31	India	KF731091.1
Oryza sativa	OS India 32	India	KF731090.1
Oryza sativa	OS India 33	India	KF731089.1
Oryza sativa	OS India 34	India	KF731088.1
Oryza sativa	OS India 35	India	KF731087.1
Oryza sativa	OS India 36	India	KF731086.1
Oryza sativa	OS India 37	India	KF731085.1
Oryza sativa	OS India 38	India	KF731084.1
Oryza sativa	OS India 39	India	KF731083.1
Oryza sativa	OS India 40	India	KF731082.1
Oryza sativa	OS India 41	India	KF731081.1
Oryza sativa	OS India 42	India	KF731080.1
Oryza sativa	OS India 43	India	KF731079.1
Oryza sativa	OS India 44	India	KF731078.1
Oryza sativa	OS India 45	India	KF731077.1
Oryza sativa	OS India 46	India	KF731074.1
Oryza sativa	OS India 47	India	KF731073.1
Oryza sativa	OS India 48	India	KF731072.1
Oryza sativa	OS India 49	India	KF731071.1
Oryza sativa	OS India 50	India	KF731070.1
Oryza sativa	OS India 51	India	KF731069.1
Oryza sativa	OS India 52	India	KF731068.1
Oryza sativa	OS India 53	India	KF731067.1
Oryza sativa	OS India 54	India	KF731066.1
Oryza sativa	OS India 55	India	KF731065.1
Oryza sativa	OS India 56	India	KF731064.1
Oryza sativa	OS India 57	India	KF731063.1
Oryza sativa	OS India 58	India	KF731062.1
Oryza sativa	OS India 59	India	KF731061.1
Oryza sativa	OS India 60	India	KF731060.1
Oryza sativa	OS India 61	India	KF731059.1

Supplementary Table 3. *(continue)*

Species	Identification Code	Country of Origin	GenBank Accession Number
Oryza sativa	OS India 62	India	KF731058.1
Oryza sativa	OS India 63	India	KF731057.1
Oryza sativa	OS India 64	India	KF731056.1
Oryza sativa	OS India 65	India	KF731055.1
Oryza sativa	OS India 66	India	KF731054.1
Oryza sativa	OS India 67	India	KF731053.1
Oryza sativa	OS India 68	India	KF731052.1
Oryza sativa	OS India 69	India	KF731051.1
Oryza sativa	OS India 70	India	KF731050.1
Oryza sativa	OS India 71	India	KF731049.1
Oryza sativa	OS India 72	India	KF731048.1
Oryza sativa	OS India 73	India	KF731047.1
Oryza sativa	OS India 74	India	KF731046.1
Oryza sativa	OS India 75	India	KF731045.1
Oryza sativa	OS India 76	India	KF731044.1
Oryza sativa	OS India 77	India	KF731043.1
Oryza sativa	OS India 78	India	KF731042.1
Oryza sativa	OS India 79	India	KF731041.1
Oryza sativa	OS India 80	India	KF731040.1
Oryza sativa	OS India 81	India	KF731039.1
Oryza sativa	OS India 82	India	KF731038.1
Oryza sativa	OS India 83	India	KF731037.1
Oryza sativa	OS India 84	India	KF731036.1
Oryza sativa	OS India 85	India	KF731035.1
Oryza sativa	OS India 86	India	KF731034.1
Oryza sativa	OS India 87	India	KF731033.1
Oryza sativa	OS India 88	India	KF731032.1
Oryza sativa	OS India 89	India	HG794000.1
Oryza sativa	OS Australia 1	Australia	KU923990.1
Oryza sativa	OS Australia 2	Australia	KU923989.1
Oryza sativa	OS Australia 3	Australia	KU923988.1

Supplementary Table 3. *(continue)*

Species	Identification Code	Country of Origin	GenBank Accession Number
Oryza sativa	OS Australia 4	Australia	KU923987.1
Oryza sativa	OS China 1	China	AF148650.1
O. sativa	OS Unknown	Unknown	EU434287.1
Oryza rufipogon	*O. rufipogon 1*	Unknown	EU434286.1
Oryza rufipogon	*O. rufipogon 2*	India	FJ908261.1
Oryza rufipogon	*O. rufipogon 3*	China	AF148651.1
Oryza punctata	*O. punctata 1*	Unknown	KP864529.1
Oryza punctata	*O. punctata 2*	Unknown	KP864528.1
Oryza punctata	*O. punctata 3*	Unknown	KP864527.1
Oryza officinalis	*O. officinalis 1*	Philippines	AF148658.1
Oryza officinalis	*O. officinalis 2*	China	KP121859.1
Oryza officinalis	*O. officinalis 3*	Papua New Guinea	KP121858.1
Oryza australiensis	*O. australiensis 1*	Australia	AF148667
Oryza granulata	*O. granulata 1*	China	AF148674.1
Oryza granulata	*O. granulata 2*	Vietnam	KP121857.1