



Research article

Genetic diversity of upland traditional rice varieties in Malaysian Borneo based on mitochondrial cytochrome c oxidase 3 gene analysis

Eric Tzyy Jiann Chong, Lucky Poh Wah Goh, Mariam Abd. Latip, Zaleha Abdul Aziz, Noumie Surugau and Ping-Chin Lee*

Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

* **Correspondence:** Email: leepc@ums.edu.my; Tel: +6088320000.

Abstract: The origins of upland traditional rice varieties in Sabah, Malaysian Borneo and their domestications are still being debated until today, despite the fact that rice planting has been practiced for centuries in this region. We report the genetic diversity of upland traditional rice varieties in Sabah targeting the mitochondrial cytochrome c oxidase 3 (*COX3*) gene in this study. The upland traditional rice germplasms were collected from three divisions in Sabah, namely the Interior Division (ID), Sandakan Division (SD), and West Coast Division (WCD). Polymerase chain reaction was conducted to amplify the *COX3* gene of these germplasms, cloned into a vector, and subjected to direct sequencing. The genetic diversity of the aligned *COX3* coding sequences among rice varieties were analyzed. This study identified an amino acid variant (S186L) in the *COX3* gene that enables the differentiation between Japonica and Indica rice groups, and almost all the upland traditional rice varieties in this study shared the same ancestor as the Japonica group. The genetic diversity among the upland traditional rice varieties was relatively low and the genetic variations in the *COX3* gene had undergone positive selection. A total of 14 monomorphic amino acid variations were identified. Rice varieties from ID were more genetically similar to those from WCD but genetically diverse to rice varieties from SD. In conclusion, the S186L amino acid variant in the *COX3* gene is a reliable molecular marker to differentiate between Japonica and Indica rice groups in this study. The genetic knowledge of rice varieties in this study could be utilized to broaden the genetic diversity of rice in a local and international scale.

Keywords: cytochrome c oxidase 3; genetic diversity; Sabah division; Sabah traditional rice

1. Introduction

Rice (*Oryza sativa* L.) is cultivated on about 148 million hectares of the world's agricultural land and serves as a main source of food for more than half of the global population [1]. In Sabah, also known as Malaysian Borneo, more than 30,000 hectares of land in three main divisions, namely the Interior Division (ID), Sandakan Division (SD), and West Coast Division (WCD), are occupied by rice plantations [2]. Due to this unique geographical structure, rice plantations in Sabah are divided into highlands that are exposed to long-term lower temperatures and serves as a staple food in rural areas, as well as lowlands that are exposed to higher temperatures for domestic supplies and exports [3].

Rice is widely planted in China, Japan, India, and Southeast Asia. It can be divided into five subgroups, namely temperate Japonica, tropical Japonica (also known as Javanica), Indica, aus/boro, and basmati/sadri. Of these, only the Japonica and Indica rice groups have been extensively studied. Previous studies claim that Japonica and Indica rice were evolutionarily domesticated from common wild rice (*Oryza rufipogon* Griff.) about 10,000 years ago [4,5]. Japonica grains are short and round, do not easily shatter, and have an intermediate amylose content which gives them moisture and stickiness after being cooked. However, the characteristics of Indica grains are the opposite, especially with its higher amylose content [6]. There is a variety of upland rice and this colored or pigmented rice is gaining popularity due to its perceived health-promoting properties [7]. In Sabah, the origins of upland traditional rice varieties are still being debated as there is a lack of proper rice plantation management and hence developing a systematic breeding approach for rice improvement is faced with many difficulties. Local farmers in Sabah usually identify the origins of rice based on grain morphologies without any validation. Since crops planted in different geographical and environmental conditions exhibit changes in their morphologies that prohibit accurate identification [8], it is crucial to utilize a molecular approach to study the characterization of the upland traditional rice varieties and genetic diversity.

Organelles of living organisms such as mitochondria or chloroplast have been widely used in genetic and evolution analyses due to their slow nucleotide substitution rate, uniparental inheritance, and the absence of intermolecular recombination [9,10]. The complete rice mitochondria genome exceeds 490 kilobase pairs that encodes for more than 300 functional proteins to carry out daily biological mechanisms including electron transport chain and tricarboxylic acid cycle respiration, general metabolism, supporting machinery, heat shock and stress response, and some are with unknown function in rice [11,12]. Several complex subunits of cytochrome c oxidase (COX) in the mitochondria are responsible for electron transfer and proton-pumping functions in rice. COX3 is one of the biggest subunits that forms the functional core of the COX complex IV. Knock-down of this gene shows no assembly of COX complex IV [13,14], indicating the important role of COX3 in the daily metabolism of plants. Recently, *COX3* gene has been used to evaluate the domestication of rice in China [15]. To date, there is no report of upland rice domestication based on *COX3* gene in Southeast Asia and this includes Sabah. Hence, this is the first study to describe the genetic diversity of upland traditional rice varieties targeting the mitochondrial *COX3* gene.

2. Materials and methods

2.1 Rice samples and DNA isolation

Twenty-three upland traditional rice varieties were accumulated from three different divisions of Sabah including the ID (n = 7), SD (n = 4) and WCD (n = 12) (Figure 1). Other unique features of the rice varieties including length, width, thickness, size, shape, color, head rice recovery, moisture content, and amylose content were previously described [16]. Cetyl trimethylammonium bromide (CTAB) method was used to isolate the genomic DNA from the seeds with slight modifications from those previously reported where the polyvinylpyrrolidone and β -mercaptoethanol were replaced with CTAB [17].

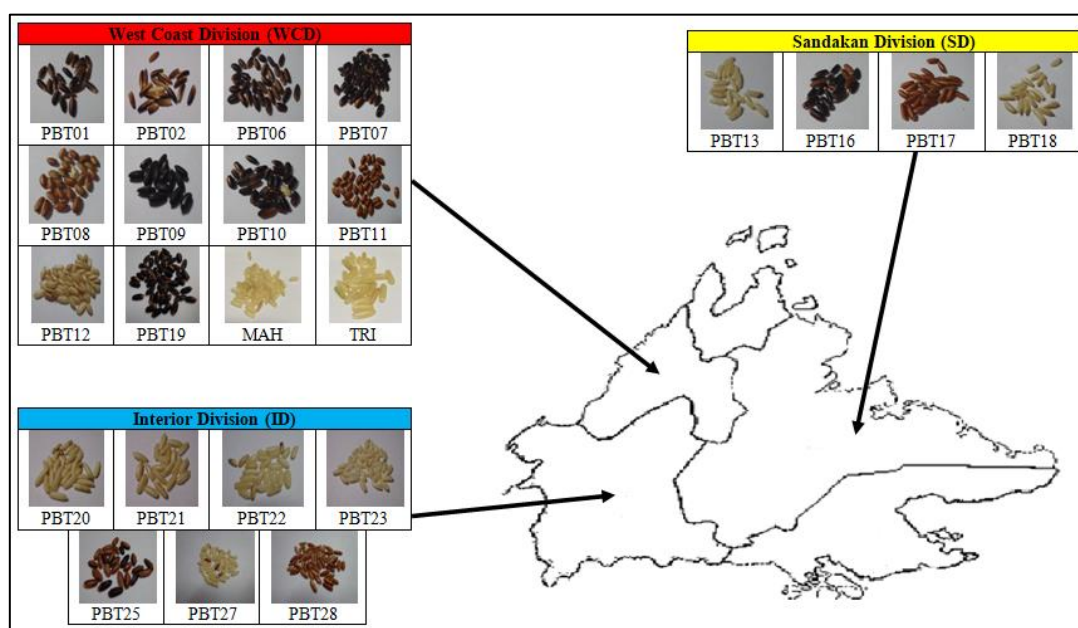


Figure 1. Sabah upland traditional rice varieties collected from different divisions.

2.2 Amplification of *COX3* gene

The full length of the *COX3* gene was amplified using primer pairs *COX3*-F (5'-TAT GAA ATA TCT CAA ACC CAC G-3') and *COX3*-R (5'-GGG CAT GAT AAA GAC CAA TAA-3'). PCR amplification was conducted by preparing a final volume of 20 μ L mixture containing 1 unit of TopTaq DNA polymerase (Qiagen, Germany), 1X of PCR buffer (Qiagen, Germany), 0.2 μ M of each primer, and 0.2 mM of dNTPs. The PCR condition was set at 94 $^{\circ}$ C for 4 min, 35 cycles at 94 $^{\circ}$ C for 30 s, 55 $^{\circ}$ C for 30 s and 72 $^{\circ}$ C for 1.5 min, and final elongation at 72 $^{\circ}$ C for 10 min. A PCR fragment with the estimated size of 1135 bp was electrophoresed and analyzed in 1% agarose gel.

2.3 TA-cloning and direct sequencing

QIAquick Gel Extraction Kit (Qiagen, Germany) was utilized to purify the PCR products from gel and the purified products were cloned into pCRTM4-TOPO[®] vector using the TOPO[®] TA Cloning

Kit for Sequencing (Thermo Scientific, USA). The ligated products were transformed into One Shot[®] Mach1[™]-T1^R competent cells (Thermo Scientific, USA) using Gene Pulser Xcell[™] Electroporation Systems (Bio-Rad, USA). QIAprep Spin Miniprep Kit (Qiagen, Germany) was utilized to extract the desired plasmid containing the *COX3* gene following the manufacturer's instructions, and the extracted plasmid was subjected to direct sequencing using T3 universal primer.

2.4 Sequence alignment of the *COX3* gene and phylogenetic analysis

CLUSTAL-W tool in the Molecular Evolutionary Genetic Analysis 6 (MEGA6) Software was applied to align the 29 *COX3* gene nucleotide sequences [18]. These 29 *COX3* gene nucleotide sequences were made up of 23 sequences from this study, three reference sequences of Japonica rice with GenBank accession ID: BA000029, DQ167400 and NC_011033, and three sequences of Indica rice as outgroup with GenBank accession ID: NC_007886, DQ167399, and JN861112 (Supplementary File 1). The aligned sequences were trimmed and only the coding nucleotide sequences (843 bp) of *COX3* gene were included for analysis. Neighbour Joining method was utilized to construct a phylogenetic tree as previously described [19]. Bootstrap replicates of 1000 were applied in constructing the phylogenetic tree to assess its reliability and robustness.

2.5 *COX3* sequence variation analysis

The aligned *COX3* sequences (except for PBT23 and Indica rice outgroups) were subjected to diversity and natural selection analyses. DnaSP ver. 5.10.01 was used to analyze genetic polymorphism in the *COX3* sequences [20]. Data including the number of haplotype (h), haplotype diversity (Hd), nucleotide diversity (π), and the average number of pairwise nucleotide differences (K) were obtained. For step-wise diversity of *COX3* estimation, π was computed on a sliding window of 100 bases with a step size of 25 bp. Z-test in MEGA6 was utilized to compare the rates of synonymous (d_s) and non-synonymous (d_N) variations by using the Nei and Gojobori approach with Jukes and Cantor's correction [21]. To evaluate the neutral theory of evolution, Tajima's D together with Fu and Li's D and F tests were calculated using DnaSP ver.5.10.01 [22,23]. The median-joining method in NETWORK ver.4.6.1.3 software was applied to generate haplotype linkage of *COX3* for rice varieties in this study [24]. The genetic difference of *COX3* between the different divisions in Sabah was calculated using Wright's F_{ST} fixation index in DnaSP ver. 5.10.01 [25].

3. Results

3.1 Phylogenetic analysis

Phylogenetic tree analysis in the present study demonstrated that one of the examined rice varieties from ID (PBT23) was fell under the Indica rice group (4.35%) and all other upland rice varieties were under the Japonica rice group (95.65%) (Figure 2).

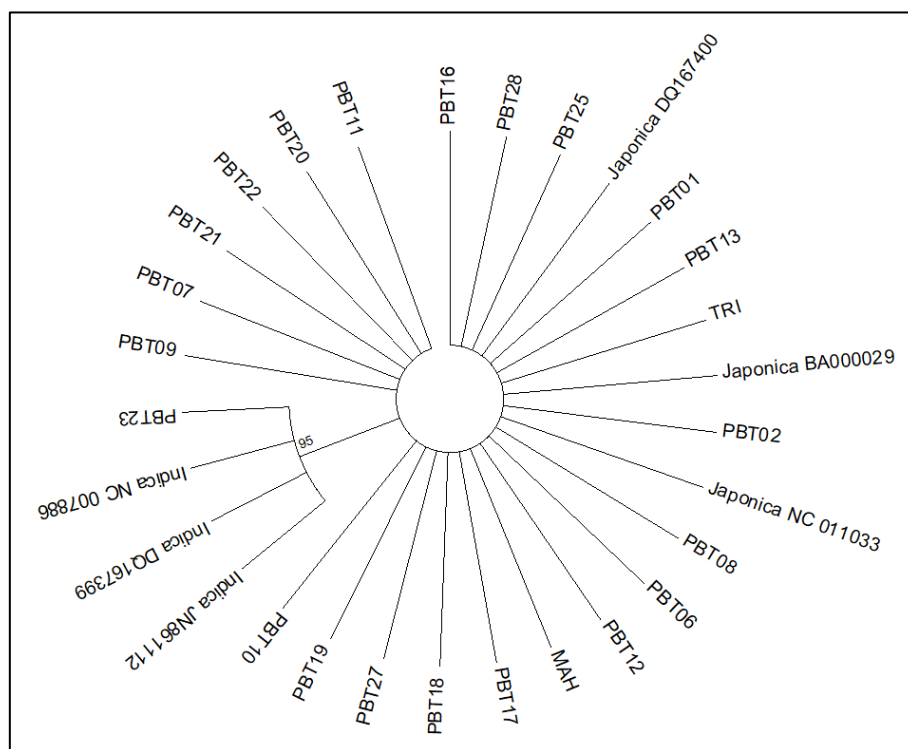


Figure 2. Phylogenetic tree of *COX3* gene sequences constructed using the Neighbour Joining method in MEGA6. The number at nodes indicates the percentage support of 1000 bootstrap replicates.

3.2 Nucleotide diversity

The analysis at the nucleotide level of *COX3* in this study revealed that the overall nucleotide diversity (π) and haplotype diversity (Hd) were 0.002 ± 0.001 and 0.737 ± 0.096 respectively. The average number of pairwise nucleotide differences (K) was 1.820. The advanced analysis of π with sliding window length of 100 bp and step size of 25 bp showed that between 201–300, 476–625 and 701–842 bp were the most preserved region while nucleotide position at 26–125 bp had the greatest peak of nucleotide diversity of the *COX3* gene among these rice varieties (Figure 3).

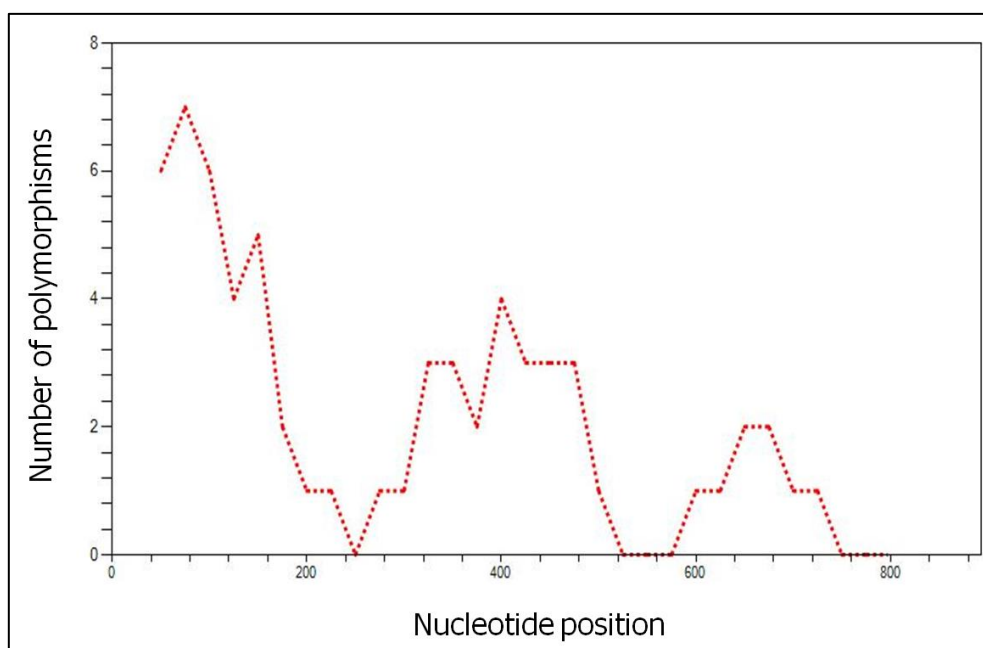


Figure 3. Nucleotide segregating in the *COX3* of upland traditional rice varieties. Sliding window plot with a window length of 100 bp and a step size of 25 bp for number of segregating sites within the aligned *COX3* sequences was generated using DnaSP ver5.10.01.

3.3 Natural selection

The average rates of non-synonymous mutation (d_N) and synonymous mutation (d_S) were 0.0023 and 0.0015 respectively, and the d_N/d_S ratio was 1.533 ($d_S < d_N$; $P < 0.05$ in Z-test). In addition, Tajima's D was -2.295 ($P < 0.05$) whereas Fu and Li's D and F were -3.881 and -3.971 (both $P < 0.02$) respectively in testing the neutral theory of evolution.

3.4 Amino acid variation and haplotype sharing of the *COX3* gene

The amino acid analysis showed 14 monomorphic mutations (M1L, T9P, W14G, S19F, S23F, G36V, G39R, T43I, V106A, H121D, P126S, N151S, G212S, and F232S) (Figure 4). The *COX3* amino acid sequences were categorized into 12 different haplotypes and a total of 13 rice varieties belonged to haplotype 1 (including three Japonica rice varieties from the NCBI database). When the median-joining network analysis was performed, an observation was made of the haplotype sharing of upland traditional rice varieties between different divisions in Sabah and Japonica rice (Figure 5).

	0	0	0	0	0	0	0	0	1	1	1	1	2	2	
	0	0	1	1	2	3	3	4	0	2	2	5	1	3	
<u>Haplotype</u>	1	9	4	9	3	6	9	3	6	1	6	1	2	2	<u>Total</u>
1	M	T	W	S	S	G	G	T	V	H	P	N	G	F	13
2	.	.	.	F	2
3	S	.	1
4	L	P	G	F	F	1
5	S	.	.	1
6	I	1
7	1
8	S	.	.	.	1
9	.	.	.	F	.	.	.	A	1
10	.	.	.	F	D	1
11	.	.	.	F	.	V	R	1
12	S	1

Figure 4. Amino acid sequence polymorphism in *COX3* from the upland traditional rice varieties. Polymorphic amino acid residues are listed for each haplotype and total number of sequence for each haplotype is showed in the right end column. Monomorphic amino acid changes are highlighted in grey shading.

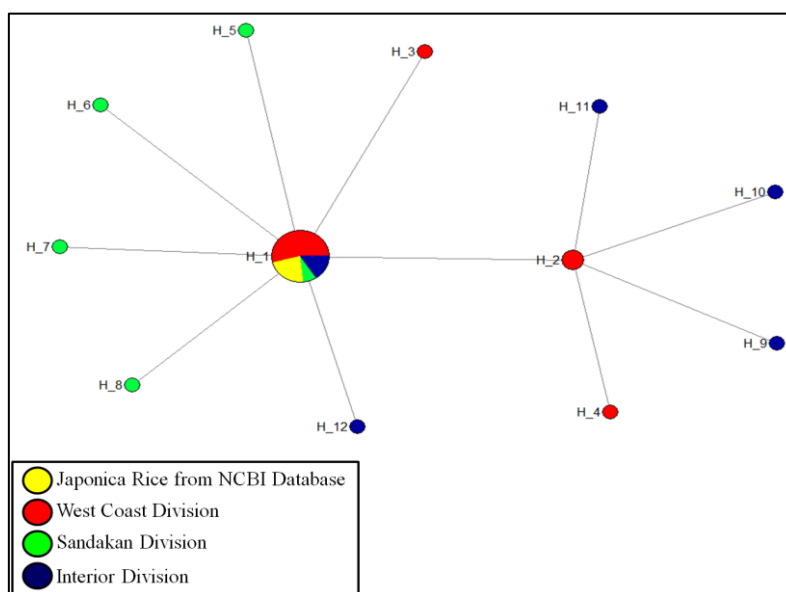


Figure 5. Haplotype (h = 12) networking for *COX3* among the upland traditional rice varieties in different divisions of Sabah when compared to Japonica rice from the NCBI database (GenBank ID: BA000029, DQ167400, and NC_011033). Each division is represented by different colors as indicated. The radius of the circle is corresponding to the total number of samples for each haplotype.

3.5 Wright's F_{ST} fixation index

The Wright's F_{ST} fixation index of *COX3* between different divisions of Sabah revealed that rice varieties from WCD and ID had the smallest F_{ST} index (Wright's $F_{ST} = 0.0003$) whereas rice varieties in SD and ID had the largest F_{ST} index (Wright's $F_{ST} = 0.1171$) in this study (Table 1).

Table 1. Wright's F_{ST} fixation index of *COX3* for upland traditional rice varieties in different divisions.

Division		1	2	3
West Coast Division (WCD)	1	-		
Sandakan Division (SD)	2	0.0202	-	
Interior Division (ID)	3	0.0003	0.1171	-

4. Discussion

Upland traditional rice varieties used in the present study were initially recorded as originating from the Japonica group based on their morphologies by local breeders. Since plant morphology can be influenced by environmental conditions and consequently can lead to inaccuracy, molecular markers such as *COX3* from the rice's mitochondrial DNA is useful to assist the characterization of an accession. Besides *COX1* and *COX2*, *COX3* is one of the largest subunits that forms the COX complex IV that functional in oxidizes cytochrome c and transfers electrons to molecular oxygen to form molecular water in plants, including rice, in a process that is coupled to H^+ translocation for ATP production [26]. The COX complexes are essential for plant growth and development, including embryogenesis, germination, vegetative growth, and senescence [26].

In this study, we unexpectedly found that one of the Sabah upland traditional rice varieties in ID, namely PBT23, was genetically similar to the Indica rice group. This was revealed in the phylogenetic tree, which underlines the sensitivity of using the *COX3* gene to differentiate between Japonica and Indica rice groups. A detailed amino acid analysis showed that the Japonica rice carried a serine amino acid while Indica rice inherited a leucine amino acid at codon 186 in the *COX3* gene (Figure 6). This finding shows that the S186L amino acid variation in the *COX3* gene is a reliable molecular marker to differentiate between Japonica and Indica rice groups in this study, and further supports the importance of utilizing a biomarker in accession characterization. To avoid statistical bias, PBT23 was excluded in the subsequent genetic diversity analyses in this study.

Our genetic diversity analysis revealed extremely low nucleotide diversity (π) and a moderately low average number of pairwise nucleotide differences (K) of the aligned *COX3* sequences, which indicates a low genetic diversity among the upland traditional rice varieties in this study. The sliding window *COX3* gene analysis revealed that the gene is fairly conserved among rice varieties with a relatively low polymorphic rate. In the natural selection analysis, the d_N/d_S ratio was 1.533, which suggested that the *COX3* variations of the upland traditional rice varieties have undergone positive selection. All the negative values in Tajima's D, as well as Fu and Li's D and F tests also supported positive selection with an expansion in population size, or selective sweep. The geographical selection has been reported to generate new mutations in *COX* coding genes in other species [27]. Therefore, highland and lowland rice plantations attributed to the unique geographical structure in Sabah might

The upland traditional rice varieties from ID had the lowest Wright's F_{ST} fixation index when compared to those from WCD, but this was relatively high when compared to those from SD. This indicates that rice varieties from ID were more genetically similar to those from WCD but genetically diverse to rice varieties from SD. Previous studies report that geographical barriers such as mountains and geographic distances may have an effect on genetic diversity and organism differentiation [31,32]. ID and WCD have a smaller geographic distance (about 171 km by road) when compared to ID and SD (about 319 km by road). Additionally, there are mountains and forest barriers between ID and SD, such as the Crocker Range. Taken together, these existing geographical barriers explain the genetic similarity between the traditional rice varieties from ID to those from WCD, but the diversity to rice varieties from SD. Therefore, it is recommended to cross-breed rice varieties from ID and SD to broaden the genetic diversity of upland traditional rice varieties in Sabah.

5. Conclusions

In summary, this study revealed that the S186L amino acid variant in the *COX3* gene is a reliable molecular marker in differentiating between Japonica and Indica rice groups in the studied samples. The genetic diversity among the traditional rice varieties was relatively low. Genetic variations in the *COX3* gene have undergone positive selection, potentially through the conventional practice of rice cultivation from local breeders. A number of monomorphic amino acid variations were obtained and future studies will need to be conducted to investigate the influence of these variations in *COX3* expression and COX complex IV assembly. As rice is the essential food for more than half of the global population and given the growing demand of the Earth's ever-increasing population, the genetic knowledge of rice varieties in this study could be utilized to broaden the genetic diversity of rice in a local and global scale.

Acknowledgments

We would like to thank Sabah Agriculture Department, Tongod Agriculture Department, Sipitang Agriculture Department, and local farmers for providing the upland traditional rice varieties for this study. This study was financially supported by the Ministry of Higher Education, Malaysia (ERGS0031-STG-1/2013).

Conflict of interest

The authors declare no conflict of interest.

References

- 1 Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 35: 25–34.
- 2 Department of Agriculture Sabah (2017) Department of Agriculture Sabah: Laporan Tahunan 2017. Available from: <https://tani.sabah.gov.my/wp-content/uploads/2019/11/4-PEMBANGUNAN-1.pdf>.
- 3 Low KO, Lee YF (2012) Investigation the relationship between Kadazandusun beliefs about paddy sprits, riddling in harvest-time and paddy related sundait. *SEA J Gen Stud* 13: 71–97.

- 4 Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. *Curr Opin Plant Biol* 6: 139–146.
- 5 Jiang L, Liu L (2006) New evidence for the origins of sedentism and rice domestication in the Lower YangZi River, China. *Antiquity* 80: 335–361.
- 6 Lang NT, Buu BC (2004) Quantitative analysis on amylose content by DNA markers through backcross populations of rice (*Oryza sativa* L.). *Omonrice* 12: 13–18.
- 7 Kong S, Lee J (2010) Antioxidants in milling fractions of black rice cultivars. *Food Chem* 120: 278–281.
- 8 Bitá CF, Gerats T (2013) Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front Plant Sci* 4: 273.
- 9 Tong W, He Q, Park YJ (2017) Genetic variation architecture of mitochondrial genome reveals the differentiation in Korean landrace and weedy rice. *Sci Rep* 7: 43327.
- 10 Skuza L, Szućko I, Filip E, et al. (2019) Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. *PLoS One* 14: e0213023.
- 11 Tian XJ, Zheng J, Hu SN, et al. (2006) The rice mitochondrial genomes and their variations. *Plant Physiol* 140: 401–410.
- 12 Huang SB, Shingaki-Wells RN, Taylor NL, et al. (2013) The rice mitochondria proteome and its response during development and to the environment. *Front Plant Sci* 4: 16.
- 13 Capaldi RA (1990) Structure and function of cytochrome c oxidase. *Annu Rev Biochem* 59: 569–596.
- 14 Remacle C, Coosemans N, Jans F, et al. (2010) Knock-down of the *COX3* and *COX17* gene expression of cytochrome c oxidase in the unicellular green alga *Chlamydomonas reinhardtii*. *Plant Mol Biol* 74: 223–233.
- 15 Wei X, Wang R, Cao L, et al. (2012) Origin of *Oryza sativa* in China inferred by nucleotide polymorphisms of organelle DNA. *PLoS One* 7: e49546.
- 16 Chong ETJ, Goh LPW, Wong JJ, et al. (2018) Genetic diversity and relationship of Sabah traditional rice varieties as revealed by RAPD markers. *Pertanika J Trop Agri Sci* 41: 177–190.
- 17 Sharma K, Bhattacharjee R, Sartie A, et al. (2013) An improved method of DNA extraction from plants for pathogen detection and genotyping by polymerase chain reaction. *Afr J Biotechnol* 12: 1894–1901.
- 18 Tamura K, Stecher G, Peterson D, et al. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729.
- 19 Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
- 20 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–452.
- 21 Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426.
- 22 Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
- 23 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- 24 Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.

- 25 Wright S (1951) The genetical structure of populations. *Ann Eugen* 15: 323–354.
- 26 Mansilla N, Racca S, Gras DE, et al. (2018) The complexity of mitochondrial complex IV: an update of cytochrome c oxidase biogenesis in plants. *Int J Mol Sci* 19: 662.
- 27 Luo Y, Gao W, Gao Y, et al. (2008) Mitochondrial genome analysis of *Ochotona curzoniae* and implication of cytochrome c oxidase in hypoxic adaptation. *Mitochondrion* 8: 352–357.
- 28 Scott GR, Schulte PM, Egginton S, et al. (2011) Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol* 28: 351–363.
- 29 Lee JS, Torollo G, Ndayiragije A, et al. (2018) Genetic relationship of tropical region-bred temperate *japonica* rice (*Oryza sativa*) plants and their grain yield variations in three different tropical environments. *Plant Breed* 137: 857–864.
- 30 The 3,000 rice genome project (2014) The 3,000 rice genome projects. *GigaSci* 3: 7.
- 31 Hou Y, Lou AR (2011) Population genetic diversity and structure of a naturally isolated plant species, *Rhodiola dumulosa* (Crassulaceae). *PLoS One* 6: e24497.
- 32 Zhang YX, He CZ, Dudgeon D, et al. (2011) Mountain ridge and sea: geographic-barrier effects on genetic diversity and differentiation of the Hong Kong newt (*Paramesotriton hongkongensis*) revealed by AFLP. *Ann Zool Fenn* 48: 119–127.



AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)