

Genetic Variation Analysis of Four Local Varieties of Indonesian Black Rice (*Oryza sativa* L.) Based on Partially *rbcl* cpDNA Gene Sequence

Abdul Basith^{1*}, Estri Laras Arumingtyas², Widodo²

¹Master Program of Biology, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

²Department of Biology, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

Abstract

Black rice (*Oryza sativa* L.) varieties i.e. Toraja (South Sulawesi), Cempo Ireng (Yogyakarta), Wojalaka (East Nusa Tenggara), and Manggarai (East Nusa Tenggara) are four local black rice varieties in Indonesia whose character has not been widely studied, especially the character of genetic variation. Research aimed to determine the variation of the *rbcl* gene in the four local black rice varieties. The sample for testing the variation of the *rbcl* gene sequence in the form of black rice leaves six weeks after planting. Dendogram was carried out using the UPGMA method with the Kimura 2-parameter algorithmic calculation model using the MEGA5 version 5.2.2 program. The results showed that partially the *rbcl* gene sequence was successfully amplified on four black rice varieties with a sequence length of 487 bp. The partial *rbcl* sequence of black rice consisted of 26.58% tyrosine, 21.38% cytosine, 28.86% adenine, and 23.18% guanine. The value of G + C content was 0.446, with the frequency of invariable sites of 97.13%. The frequency of informative parsimony sites was 1.43% with a nucleotide diversity (Pi) value of 42-10, the number of haplotypes was 5, and the total number of mutations and polymorphic sites was 14. The ratio between transition and transversion (ts/tv ratio k) for purine bases was 1.741 and pyrimidine was 3.571, with the estimated overall ratio between transition and transversion (R) of 1.31. Based on the dendogram, the farthest genetic distance was found in Wojalaka and Manggarai varieties, which were 0.019 respectively.

Keywords: black rice, genetic variation, local varieties, *rbcl* gen

INTRODUCTION

Black rice (*Oryza sativa* L.) is one type of rice in the world. In addition to white rice, brown rice, and brown rice [1], black rice has become popular and is consumed by some people as a functional food. Petroni and Tenolli [2] suggest that the increase in demand for functional foods is due to the high content of antioxidants, such as anthocyanin pigments. Scientific studies on Indonesian local black rice are few, and information about black rice, which has more potential as a functional food, is also limited.

The *Toraja* variety (Sulawesi), the *Cempo Ireng* variety (Yogyakarta), the *Wojalaka*, and the *Manggarai* varieties (East Nusa Tenggara) are four local black rice varieties in Indonesia that have not been widely studied, particularly concerning genetic variations and anthocyanin levels. Complete information can support the efforts of breeding and conservation of the four black rice varieties.

Genetic variation within a species is often influenced by the reproductive behavior of individuals in that population. Genetic variation arises because each individual has unique gene forms. Genetic variation occurs through mutation and recombination mechanisms [3]. One of the

approaches used to study genetic variation is the DNA barcode. DNA barcode is a sequence or sequence of nucleotide bases of DNA or certain short genes, taken from one or more standardized genomes, used for the fast and practical identification and discovery of species [4]. Most of the genes that were used as DNA barcode in plants are those contained in the chloroplast DNA (cpDNA) genome. The selection of the barcode gene on cpDNA is due to the higher number of nucleotide substitutions compared to the genomic mitochondrial DNA (mtDNA). Besides that, cpDNA in most plant species is uniparentally derived, making it easier for evolutionary studies [5].

The barcode gene used in this study is the *rbcl* gene contained in cpDNA. The *rbcl* gene is a gene that encodes the key protein ribulose-1,5-biphosphate carboxylase-oxygenase (abbreviated as RuBisCO), which participates in carbon fixation in the photosynthesis process [6]. The *rbcl* gene is about 1400 bp (base pairs) in length, so it provides many characters for phylogenetic studies [7]. Cumming *et al.* [8] explained that the database on the *rbcl* gene was owned by many species, making it easier to compare in data analysis. Based on a database that can be accessed through the official NCBI (National Center for Biotechnology Information) website, to date, more than 143 partial nucleotide sequences of the *rbcl* gene from the genus *Oryza* have been collected. The purpose of this study was to reveal the genetic variation in the four

* Correspondence address:

Abdul Basith

Email : golden_bee46@yahoo.com

Address : Dept. Biology, University of Brawijaya, Malang 65145

local black rice varieties based on partial *rbcl* gene sequences.

MATERIAL AND METHOD

Samples and DNA Extraction

The samples of the seeds of four black rice varieties used in this study were obtained from local farmers in each region. Partial analysis of *rbcl* gene sequences was carried out on extracted DNA from black rice leaf tissues six weeks after planting.

The extraction of black rice genomic DNA in this study used the Wizard® Genomic DNA Purification Kit (Promega Corporation) following standard protocols. Measurement of the concentration and purity of isolated DNA was carried out qualitatively and quantitatively. Qualitative measurements used 1% agarose gel electrophoresis with a 1 kb DNA ladder comparison (VC 1kb DNA ladder) using a Nanodrop2000 UV spectrophotometer machine. Quantitative measurements of DNA purity and concentration were carried out by taking 2 µl of DNA samples then added with 995 µl of Tris-EDTA (TE) buffer and placed in the vortex machine until the solution was homogeneous. Then the solution was put in a cuvette and its absorbance was measured at a wavelength of 230 nm, 260 nm, and 280 nm.

Amplification of partially *rbcl* gene sequences.

The primers used in this study were universal primers of the *rbcl* gene. The forward primer used was *rbcl*a-F with the 5'-ATGTCACCACAAACAGAAAC-3' arrangement developed by Kress and Erickson [9] and the reverse primer used was *rbcl*-724R by Fay et al. [10] with the 5'-TCGCATGTACTGCAGTAGC-3'. The PCR cycles performed were pre-denaturation at 95°C for 5 minutes, denaturation at 95°C for 45 seconds, annealing at 60.8°C for 45 seconds, extension at 72°C for 45 seconds, and post extension at 72°C

for 10 minutes. The main cycle (denaturation, annealing, and extension) of amplification was repeated 35 times. Confirmation of isolated DNA and *rbcl* gene amplicon was carried out using 2% agarose gel electrophoresis.

Sequencing of partially *rbcl* gene sequences.

Purification and sequencing were carried out by 1st base (Selangor, Malaysia) through distributor PT. Genetics Science Indonesia. Sequencing was performed using the ABI PRISM 3730xl Genetic Analyzer (Biosystem, USA). The standard protocol used was the BigDye® Terminator v3.1 Cycle Sequencing Kit. The sequencing results read by BioEdit program.

Sequence Alignment and Data Analysis.

The validity test of DNA sequences was carried out using the BLAST program, which can be accessed online on the NCBI website. Meanwhile, the partial reliability test of the *rbcl* gene sequences was carried out using the BioEdit program. Next, sequence alignment was carried out to determine the homology of a DNA sequence with other DNA sequences. The sequence alignment used in this study was multiple alignments because it involves many partial homologous gene sequences. Multiple alignments were done using ClustalW. Dendrogram tree construction used the UPGMA (Unweighted Pair Group Method with Arithmetic Means) method with the algorithmic calculation of the 2-parameter Kimura substitution model. A total of 12 accessions to the partial comparison of *rbcl* gene sequences were selected from Genbank. The list of accessions for comparisons was described in Table 1. Evaluation of the phenogram tree was carried out using a bootstrap test with 1000 replications. Multiple alignments and construction of phenogram trees were carried out using the MEGA5 (Molecular Evolutionary Genetic Analysis) program 5.2.2.

Table 1. List of 12 accessions used for comparison

No.	Accession codes	Species description	Code for analysis
1.	KF731215.1	<i>Oryza sativa</i> Japonica Group bio-material RGCB#CR145	Os Japonica Group 1
2.	KF731214.1	<i>Oryza sativa</i> Japonica Group bio-material RGCB#CR383	Os Japonica Group 2
3.	KF731204.1	<i>Oryza sativa</i> Japonica Group bio-material IRGC:9147	Os Japonica Group 3
4.	KF731201.1	<i>Oryza sativa</i> Japonica Group bio-material RGCB#CR380	Os Japonica Group 4
5.	KF731220.1	<i>Oryza sativa</i> Indica Group bio-material RGCB#CR487	Os Indica Group 1
6.	KF731218.1	<i>Oryza sativa</i> Indica Group bio-material RGCB#CR495	Os Indica Group 2
7.	KF731217.1	<i>Oryza sativa</i> Indica Group bio-material RGCB#CR666	Os Indica Group 3
8.	KF731212.1	<i>Oryza sativa</i> Indica Group bio-material RGCB#CR40	Os Indica Group 4
9.	KF731190.1	<i>Oryza rufipogon</i> bio-material RGCB#Kor138	<i>Oryza rufipogon</i> 1
10.	KF731189.1	<i>Oryza rufipogon</i> bio-material RGCB#Kor124	<i>Oryza rufipogon</i> 2
11.	JQ593378.1	<i>Oryza latifolia</i> voucher BioBot01092	<i>Oryza latifolia</i> 1
12.	JQ593377.1	<i>Oryza latifolia</i> voucher BioBot01091	<i>Oryza latifolia</i> 2

(Source: NCBI, 2014)

RESULT AND DISCUSSION

Partial Extraction and Amplification of the *rbcl* gene sequence

Partial amplification of *rbcl* gene sequences with universal primers was successfully carried out with *rbcl*a-F forward primers and *rbcl*-724R reverse primers. The results of electrophoresis on 2% agarose gel showed that the partial length of the amplified *rbcl* gene sequence was 600 bp (Fig. 1).

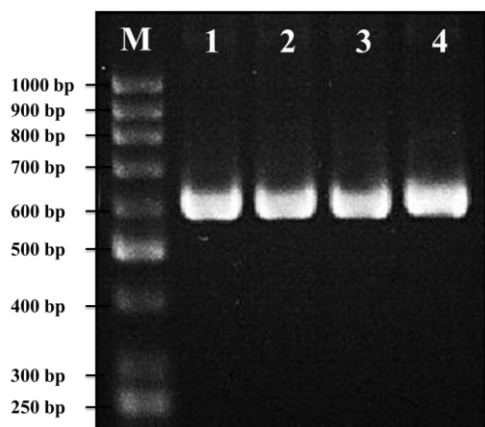


Figure 1. Visualization of 1% agarose gel electrophoresis as the result of genomic DNA isolation of four black rice varieties; M: Marker 1 kb, 1: Toraja variety, 2: Cempo Ireng variety, 3: Manggarai variety, and 4: Wojalaka variety

Sequence's characteristics.

Partial characteristics of the 487 bp *rbcl* gene sequence obtained from black rice varieties Cempo Ireng, Toraja, Wojalaka, and Manggarai were compared to 12 accessions of Genus *Oryza* obtained from Genbank. It showed the characteristics of nucleotides composed of 26.58% tyrosine, 21.38% cytosine, 28.86% adenine, and 23.18% guanine (Table 2). The value of G + C content is 0.446, with the frequency of

invariable sites of 97.13%. The frequency of informative parsimony sites is 1.43% with a nucleotide diversity (Pi) value of 42-10, the number of haplotypes is 5, and the total number of mutations and polymorphic sites is 14. The ratio between transition and transversion (ts/tv ratio k) for purine bases is 1.741, and pyrimidine is 3.571, with the estimated overall ratio of transition and transversion (ts/tv R) of 1.31.

Partial comparison of *rbcl* gene sequences of black rice with 12 accessions from GenBank showed 473 monomorphic sites (invariable) and 14 polymorphic sites (variable). These 14 polymorphic sites show that 14 nucleotide base points have mutations. Singleton variations are found in the positions of the nucleotide base sequences 131, 174, 206, 212, 216, 220, and 419, while the informative parsimony variations are in the positions of the nucleotide base sequences 35, 128, 154, 162, 296, 424, and 440. According to Yingzhi *et al.* [11], the single tone variation is a site where only one taxon is different. Meanwhile, the informative parsimony site is an informative site with a minimum characteristic of being composed of two nucleotides, and both must appear at least twice on the site.

Singleton variations occur because of mutations, both transitional mutations, and transversion mutations. According to Pierce [6], a transition mutation is the replacement of a purine nucleotide base with a purine or pyrimidine nucleotide base with a pyrimidine, while a transversion mutation is the replacement of a purine nucleotide base with a pyrimidine or vice versa. In the Manggarai variety, variations at positions 174, 206, 212, and 220 were caused by a transfer mutation, while 131 and 419 were due to a transitional mutation.

Table 2. Characteristics of partially *rbcl* gene sequences.

Parameter	Results	Total
Frequency of tyrosine (T)	26.58%	487
Frequency of cytosine (C)	21.38%	487
Frequency of adenine (A)	28.86%	487
Frequency of guanin (G)	23.18%	487
G + C content	0.446	
Invariable site frequency	97.13%	
Frequency of informative parsimony sites	1.43%	
Diversity of nucleotide (Pi)	0.0000042	
Number of haplotypes	5	
Total number of mutations	14	
Polymorphic site	14	
Estimated overall ts/tv ratio (k)	Purine = 1.741 Pyrimidine = 3.571	
Estimated overall ts/tv ratio (R)	1.31	

Note: frequency analysis of nucleotide bases based solely on 487 bp

Variations of informative parsimony at positions 35 and 440 occurred transitional mutations, while at positions 154 and 296 there were transversion mutations in comparison accessions of *Oryza latifolia* 1 and 2. At position 128 there were transitional mutations in Manggarai and Cempo Ireng, position 162 there were transition mutations in Toraja varieties and Manggarai, position 424, there are transitional mutations in the Toraja and Cempo Ireng varieties. The total number of mutations for the parsimony informative variation of the Manggarai variety is 2 and the overall number of single nucleotide variations is 7, this is thought to be the cause of the position of the Manggarai variety being in the outermost branch of the dendrogram.

Other molecular markers have also been used to describe genetic variation in local Indonesian black rice, namely microsatellites or SSR (short tandem repeat) [1]. Based on the use of microsatellite markers, it is known that black rice varieties are genetically different from white rice and that there are intraspecies variations regardless of geographical origin. In line with the results of this study, the use of partial markers for the *rbcl* gene sequence also showed that there were genetic variations in these four black rice varieties regardless of their geographic origin. This kind of genetic variation analysis still needs to be carried out as an initial step for the identification of black rice varieties and the selection of parent crosses in assisting black rice breeding programs in Indonesia.

Genetic Distance and Dendrogram.

Dendrogram construction was carried out using the UPGMA method with the Kimura 2-parameter algorithmic calculation model (Fig. 2). The UPGMA dendrogram was constructed based on a 487 bp long sequence.

The farthest genetic distance was found in Wojalaka and Manggarai varieties, which were 0.019. This genetic distance showed that the kinship of the two varieties is quite far, even separated and isolated even though they are in one province. The Toraja and Cempo Ireng varieties were grouped into one node with a genetic distance of 0.04. There are 14 nucleotide base points that have mutations. Single nucleotide variations (single tone) are found at the positions of the nucleotide sequence 131, 174, 206, 212, 216, 220, and 419, while the informative parsimony sites are at positions of

nucleotide base sequences 35, 128, 154, 162, 296, 424, and 440.

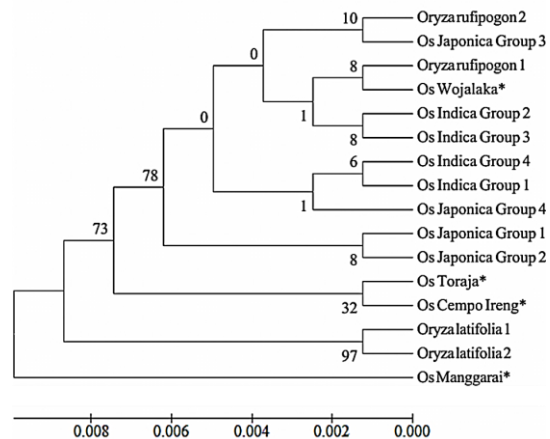


Figure 2. The dendrogram with the UPGMA method (bootstrap 1000 replications) showed the genetic relationship of four black rice varieties; (*) samples of black rice varieties in this research.

CONCLUSION

The 487 bp long nucleotides tested in this study consisted of 26.58% tyrosine, 21.38% cytosine, 28.86% adenine, and 23.18% guanine. The value of G + C content was 0.446, with the frequency of invariable sites of 97.13%. The frequency of informative parsimony sites was 1.43% with a nucleotide diversity (Pi) value of 42-10, the number of haplotypes was 5, and the total number of mutations and polymorphic sites was 14. The ratio between transition and transversion (ts-tv ratio or k) for purine bases was 1.741 and pyrimidine was 3.571, with the estimated overall ratio between transition and transversion (R) of 1.31.

The UPGMA dendrogram showed the farthest genetic distance was found in Wojalaka and Manggarai varieties. The Toraja and Cempo Ireng varieties were grouped into one node with a 0.04 genetic distance. There are 14 nucleotide base points that have mutations. Single nucleotide variations (single tone) found at the nucleotide sequence 131, 174, 206, 212, 216, 220, and 419, while the informative parsimony sites are at nucleotide base sequences 35, 128, 154, 162, 296, 424, and 440.

ACKNOWLEDGEMENT

Thanks to Dr. Endang Arisoelaningsih (Department of Biology, University of Brawijaya) and the staff of the Laboratory of Genetics and Molecular Biology, Maulana Malik Ibrahim State Islamic University of Malang and the Laboratory of Molecular Biology, State University of Malang.

REFERENCES

- [1] Kristamtini, K., T. Taryono, P. Basunanda, R.H. Murti. 2016. Keragaman genetik kultivar padi beras hitam lokal berdasarkan penanda mikrosatelit. *J. AgroBiogen*. 10(2). 69.
- [2] Petroni, K., C. Tonelli. 2011. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci*. 181(3). 219–229.
- [3] Indrawan, M. 2007. Biologi konservasi. Yayasan Obor Indonesia. Jakarta.
- [4] Spooner, D.M. 2009. DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am. J. Bot.* 96(6). 1177–1189.
- [5] Hollingsworth, P.M., S.W. Graham, D.P. Little. 2011. Choosing and using a plant DNA barcode. *PLoS One*. 6(5). e19254.
- [6] Pierce, B.A. 2017. Genetics : a conceptual approach, 6th Ed. W.H. Freeman. New York.
- [7] Hollingsworth, P.M., L.L. Forrest, J.L. Spouge, et al. 2009. A DNA barcode for land plants. *Proc. Natl. Acad. Sci.* 106(31). 12794–12797.
- [8] Cummings, M.P., J.M. Nugent, R.G. Olmstead, J.D. Palmer. 2003. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcl* to the mitochondrial genome in angiosperms. *Curr. Genet.* 43(2). 131–138.
- [9] Kress, W.J., D.L. Erickson. 2007. A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. *Plos One*. DOI: 10.1371/journal.pone.0000508.
- [10] Fay, M.F., S.M. Swensen, M.W. Chase. 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin*. 52. 111–120.
- [11] Yingzhi, L., C. Yunjiang, T. Nengguo, and D. Xiuxin. 2007. Phylogenetic analysis of mandarin landraces, wild mandarins, and related species in China using nuclear LEAFY second intron and plastid *trnL-trnF* sequence. *J. Am. Soc. Hortic. Sci.* 132(6). 796–806.