Molecular Identification and Genetic Characteristics of Genus *Mystacoleucus* Based on Gene Cytochrome Oxidase C Subunit I (COI) in Sengguruh Dam

Fitri Sil Valen¹, Maheno Sri Widodo², Yuni Kilawati³

¹Graduate School of Fisheries and Marine Sciences, University of Brawijaya, Malang, Indonesia
²Department of Aquaculture, Faculty of Fisheries and Marine Science, University of Brawijaya, Malang, Indonesia
³Department of Aquatic Resources Management, Faculty of Fisheries and Marine Science, University of Brawijaya, Malang, Indonesia

Abstract

In general, the process of naming several species of fish including species from the genus *Mystacoleucus* is still carried out based on morphological characters, which are often found in almost the same morphological characters in different fish species. Therefore, it is necessary to apply a more accurate identification method, which is a DNA-based identification method called DNA barcoding. The purpose of this study was to identify the species of the genus *Mystacoleucus* on the Sengguruh Dam molecularly based on the Cytochrome Oxidase C Subunit I (COI) gene. A caudal fin was taken on the test fish and preserved in 95% ethanol solution for molecular identification. The results of the identification showed that the sample belonged to the *Mystacoleucus marginatus* species with Identity values between 99-100% and E-value 0.0. The data obtained showed that from the calculation of genetic distance presented in the form of data matrix and phylogenetic tree reconstruction, there were 2 species that had a far genetic distance from the *M. marginatus* sample from the Sengguruh Dam namely *M. atridorsalis* with the furthest genetic distance of 0.1932-0.2595 and *M. lepturus* with genetic distance between 0.1117-0.1193. However, there is one species that has the closest distance, *M. padangensis* with genetic distance between 0.0019-0.0038 and identity values up to 99%.

Keywords: DNA barcode, Molecular identification, *Mystacoleucus*, Phylogenetic COI.

INTRODUCTION

Indonesia has a very large diversity of freshwater fish. Found as many as 1218 species from 84 families including 1172 native species of 79 families and 630 species are endemic [1]. One of the endemic fish that lives in Indonesia is the Mystacoleucus genus. The genus Mystacoleucus is a family of Cyprinidae [2-5] with specific characteristics that have horizontal spines in front of the dorsal fin [1]. These fish live in tropical rivers [6,7] and are spread throughout most of Asia [8-24].

Search results in FishBase and Catalog of Fishes found eleven species from the *Mystacoleucus* genus, and two of them were found in Indonesia, namely *Mystacoleucus marginatus* Valenciennes 1842 [25,26,27] and *Mystacoleucus padangensis* Bleeker 1852 [1,24]. Apart from Indonesia, *Mystacoleucus marginatus* is also found in Malaysia [28], Thailand [5] and China [29,30].

In general, the naming of several fish species including the genus *Mystacoleucus* is still based on morphological characters [1]. It is quite difficult to ensure the naming of species because there are so many fish that look visually similar but are genetically different, and vice versa. For this reason, it is necessary to apply a method that is more accurate in identifying species and names. Along with the development of molecular biology, a new method has been found for identification of DNA-based species known as DNA Barcoding [31].

DNA barcoding provides speed and accuracy in species identification with a focus on analysis in small segments of mtDNA [32,33]. DNA barcoding can be a solution to the current taxonomic crisis [34] and has been developed to identify species because it is relatively easy to do compared to other techniques [35]. The use of DNA Barcoding can be practically applied [36] or as a tool to support other studies [37] by selecting one or several loci that can be routinely sequenced and can be relied upon to identify large numbers of samples and diverse and easy compared between species.

The current study showed that the partial COI gene sequence of fin clips and muscle tissue can be used as a diagnostic molecular marker for identification and resolution of taxonomic ambiguity of Ompok species [38]. Other study showed that the genes of cytochrome oxidase
subunit I (COI) contained in mitochondria as DNA. Barcoding globally have been chosen as standard tools for molecular taxonomy and animal identification [31]. The COI gene is a region that encodes a protein with a base length of up to 1551 starting from the start codon “GTG” and ending with the stop codon “TAA” in line 5489-7039 in mitochondrial DNA (mtDNA) [39].

This study aims to identify species and determine the genetic characteristics of the genus Mystacoleucus found in Sengguruh Dam, Malang, with a genetic approach based on the cytochrome oxidase c subunit I (COI) gene, so that it can confirm further molecular naming.

**MATERIALS AND METHODS**

**Sample Collection**

A total of five samples were collected from the Sengguruh Dam, Malang, East Java Province, from June to August 2018. For the purposes of molecular identification, a caudal fin was taken on the test fish and preserved in 95% ethanol solution. Whereas as a reference for identification, a total of nine genus Mystacoleucus from various regions were downloaded from GenBank with accession numbers which can be seen in Table 1.

**Table 1.** DNA sequence of Genus Mystacoleucus from Genbank

<table>
<thead>
<tr>
<th>Species</th>
<th>Location</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. marginatus</em></td>
<td>Mojokerto</td>
<td>KU692639 [27]</td>
</tr>
<tr>
<td><em>M. marginatus</em></td>
<td>Jawa barat</td>
<td>KU692643 [27]</td>
</tr>
<tr>
<td><em>M. marginatus</em></td>
<td>Lumajang</td>
<td>KU692636 [27]</td>
</tr>
<tr>
<td><em>M. marginatus</em></td>
<td>Malaysia</td>
<td>KT001017 [28]</td>
</tr>
<tr>
<td><em>M. marginatus</em></td>
<td>Laos</td>
<td>JX346560 [40]</td>
</tr>
<tr>
<td><em>M. marginatus</em></td>
<td>China</td>
<td>KJ994653 [30]</td>
</tr>
<tr>
<td><em>M. atridorsalis</em></td>
<td>Singapura</td>
<td>JF915630.1 [41]</td>
</tr>
<tr>
<td><em>M. lepturus</em></td>
<td>China</td>
<td>KJ994654.1 [30]</td>
</tr>
<tr>
<td><em>M. padangensis</em></td>
<td>Medan</td>
<td>MF062176.1 [42]</td>
</tr>
</tbody>
</table>

**DNA Extraction, Isolation, and Amplification**

Genomic DNA extraction from all samples was carried out using the KIT method: Genomic DNA Mini Tissue Animal Kit (GENEAID). The Geneaid™ DNA Isolation Kit offers a simple and gentle reagent DNA precipitation method for isolating high molecular weight genomic, mitochondrial or viral DNA suitable for archiving or sensitive downstream applications. This highly versatile solution based system can be scaled proportionately in order to satisfy larger sample volumes providing a convenient sample-storage procedure with minimal hands on time.

Initially, cells are lysed in the presence of detergents and a proprietary DNA stabilization solution followed by RNase A treatment. Once proteins and other contaminants are removed DNA is precipitated then rehydrated. The high quality extracted DNA is ready for use in a variety of downstream applications. Amplification (PCR) of the mitochondrial locus cytochrome c oxidase subunit I (COI) gene was carried out using the GO TAQ Green PCR Mix method with universal primer pairs LCO1490: 5’-ggtcaaaatataagatgg-3’ and HCO2198: 5’-taaaactcagggatatgaata-3’ [43]. Making mastermix (Go Taq Green) was carried out by adding 14 µL ddH2O, forward and reverse primers of 2.5 µL, DMSO 1 µL, Go Taq Green 25 µL and DNA extraction of 5 µL. Amplification is carried out at a 50µl final volume.

The PCR process includes pre-denaturation at 94°C for three minutes, followed by 35 cycles consisting of denaturation at 94°C for 30 seconds, annealing at 50°C for 30 seconds and extension stage at 72°C for 45 seconds. Furthermore, the PCR results were carried out by an electrophoresis process to separate, identify and purify DNA fragments, using 1% agarose gel with 50 mLTris Borate EDTA (TBE). The PCR results that were successfully amplified were then sent to First Base CO (Malaysia) using Big Dye® terminator chemistry (Perkin Elmer), to obtain base arrangements that form DNA or nucleotide sequences.

**Data Analysis**

Sequencing results are aligned using the Clustal W method found in MEGA 6.06 software [44]. The sequential data is then matched with the data available in the NCBI GenBank (National Center for Biotechnology Information) online using the BLAST (Basic Local Alignment Search Tool-nucleotide) method. Genetic distance analysis of COI gene sequences was calculated using the pairwise distance method found in the MEGA 6.06 program [44]. The results of the calculation of genetic distance are presented in the form of a data matrix that can be used to analyze kinship relationships between species based on the tree of phylogeny. Phylogenetic reconstruction using the Maximum Likelihood Trees method [45], Kimura-2 parameter model and 10000 × bootstrap value using MEGA 6.06 software [44].

**RESULTS AND DISCUSSION**

**Species Identification**

The length of the sequence results of the COI gene genus Mystacoleucus sequencing from the Sengguruh Dam using primers LCO1490 and HCO2198 is 683-686 bp (base pairs). The primary use of LCO1490 and HCO2198 [43]. The results
show that the primary pairs of LCO1490 and HCO2198 consistently strengthen the fragments of approximately 700 bp. This result is also in accordance with the research conducted before [27-30,40,41] on the genus Mystacoleucus using the mitochondrial COI gene resulting in a sequence length of 549-859 bp. Some length differences in DNA sequences amplified due to the type of primer used, primary base composition, primary length, DNA quality found, food, offspring and environment [46]. A previous study suggested that 658 bp fragments using the COI gene could be used as a differentiating basis between animals [31].

Samples from the Sengguruh Dam were then identified in Genbank using the BLAST method. Samples were identified as Mystacoleucus marginatus with query cover values between 95-98%, Identity values between 99-100% and E-value 0.0. Based on the results of the BLAST analysis, it can be concluded that the sample DNA sequences have a very high level of similarity with the DNA sequences in GenBank. Wahyuningsih [47] suggested that with a similarity rate of 99-100% it could be said that species were identical and could be identified as that species. Claverie and Notredame [48] suggest that DNA sequences can be said to have homology if the E-value is smaller than e-0.4. The results of the identification are presented in Table 2.

**Genetic Characteristics**

Nucleotide analysis is very much needed in research because it is a compiler of DNA sequences. Nucleotides are DNA monomers containing three different parts, namely pentose sugar, nitrogen bases (A, T, G, C) and phosphate groups [49]. The results of the analysis of nucleotide composition showed that the average amount of adenine and thymine was found to be the highest. The average nucleotide composition (Table 4) found in the COI M. marginatus gene in Sengguruh Dam is C (Cytosine) of 28.6%, T (Thymine) of 27.0%, A (Adenine) of 27.6% and G (Guanine) of 16.8%. The G + C content of all samples is 45.4% and has a lower number than the number of A + T which amounts to 54.6%.

Based on the nucleotide base composition, the nucleotide base of M. marginatus is dominated by alkaline A (adenine) and T (thymine) bonds so that the COI gene of both species is categorized as a rich A-T (A-T rich) group. The hydrogen bond A-T consists of 2 hydrogen bonds which are weaker than the G-C hydrogen bonds which have 3 hydrogen bonds. The nucleotide base composition of Mystacoleucus marginatus is an easily separated bond so the possibility of mutation of species is higher. The composition of the nucleotide bases is presented in Table 3.

**Genetic distance and phylogenetic analysis**

The genetic distance of the COI gene fragment between Mystacoleucus spp. based on research results and GenBank is presented in the form of a data matrix (Table 4). The genetic distance of interspecific M. marginatus ranges from 0.0000 - 0.0625. While the highest genetic distance of intraspecies Mystacoleucus spp was found in M. lepturus and M. padangensis for 0.1193 and 0.0038 respectively.

Data from the genetic distance matrix on COI gene fragments were used to analyze kinship relationships based on the phylogeny tree. The phylogenetic reconstruction based on the COI gene using the Maximum Likelihood Trees method of the Kimura-2 parameter and the bootstrap value 10000 is shown in Figure 1.

---

**Table 2.** The results of the identification of the Genus Mystacoleucus sequences in the Sengguruh Dam through BLAST Analysis

<table>
<thead>
<tr>
<th>Sample code</th>
<th>Species outcome</th>
<th>Access code of NCBI</th>
<th>Query Cover(%)</th>
<th>E-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M. marginatus</td>
<td>KU692642.1</td>
<td>95</td>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>M. marginatus</td>
<td>KU692641.1</td>
<td>95</td>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>M. marginatus</td>
<td>KT001063.1</td>
<td>98</td>
<td>0.0</td>
<td>99</td>
</tr>
</tbody>
</table>

**Table 3.** Composition of Nucleotide bases Mystacoleucus marginatus in the Sengguruh Dam

<table>
<thead>
<tr>
<th>Sample code</th>
<th>T (Thymine)</th>
<th>C (Cytosin)</th>
<th>A (Adenin)</th>
<th>G (Guanine)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.0</td>
<td>28.7</td>
<td>27.6</td>
<td>16.7</td>
</tr>
<tr>
<td>2</td>
<td>26.8</td>
<td>28.7</td>
<td>27.7</td>
<td>16.8</td>
</tr>
<tr>
<td>3</td>
<td>27.3</td>
<td>28.5</td>
<td>27.5</td>
<td>16.7</td>
</tr>
<tr>
<td>Average</td>
<td>27.0</td>
<td>28.6</td>
<td>27.6</td>
<td>16.8</td>
</tr>
</tbody>
</table>
The construction of the phylogeny tree based on the COI gene above shows that the species in the *M. marginatus* group are significantly different from the *M. atridorsalis* species with genetic distances of 0.1932-0.2595 and *M. lepturus* with a genetic distance between 0.1117-0.1193. Genetically *M. marginatus* is very close to *M. padangensis*, the genetic distance of these two species ranges from 0.0019-0.0038. This is also proven by the results of BLAST, where the samples identified as *M. marginatus* from the Sengguruh Dam have an Identity value of 99% and E-value 0.0 with *M. padangensis*, so it can be concluded that the DNA sequences of *M. marginatus* and *M. padangensis* have very high genetic similarities. Wahyuningsih [47] suggested that with a 99-100% similarity rate it could be said that the species was identical and DNA sequences had homology if the E-value was smaller than e-0.4 [48].

Based on the phylogenetic tree reconstruction, it was also seen that by ingroup species, individuals of different *M. marginatus* distribution regions had very close relationships with genetic distances between 0.0625-0.0000, where 0.000 is the closest distance and this value indicates that from 1000 base pairs, no one has a different base pair. The closest distance is obtained by the *M. marginatus* species from Sengguruh Dam, Mojokerto and Lumajang. Mojokerto has a genetic closeness with the Sengguruh Dam because it is still in one water source, namely Brantas waters, from Mount Arjuno springs, Brantas source village, Bumi Aji District, Batu City which then flows to Malang, Blitar, Kediri, Mojokerto, and Sidoarjo. This
pattern of currents results in gene flow or gene exchange between locations resulting in homogeneous genes.

It is also known that the spread of *M. marginatus* is very broad covering the territory of Indonesia (East Java and West Java) to Malaysia, Laos, and China. This spread is possible due to the influence of Sundaland, where the Sunda region has a shallow continental shelf and changes in the eustatic sea repeatedly linking large islands in this region (Sundanese and Southeast Asian exposures) form Sundaland [16,50]. This allows the migration of freshwater fish into and outside Java which may have occurred 10-70 thousand years ago [51].

**CONCLUSION**

The length of fragments from the amplification of the COI gene of the genus *Mystacoleucus* from the Sengguruh Dam using primary LCO1490 and HCO2198 is 681-686 bp (base pairs), which shows that the test sample belongs to the *M. marginatus* species with an identity value of 99-100% and E-value 0.0. The average nucleotide composition in the COI *M. marginatus* gene in Sengguruh Dam is C (Cytosine) of 28.6%, T (Thymine) of 27.0%, A (Adenine) of 27.6% and G (Guanine) of 16.8%. The G + C content of all samples is 45.4% and has a lower number than the number of A + T which amounts to 54.6%.

**REFERENCES**


Genetics of Genus Mystacoleucus Based on Gene COI in Sengguruh Dam (Valen et al)


[40] Pasco-Viel, E., M. Veran, L. Viriot. 2012. Revision of the genus Cyclocheilichthys (Bleeker 1859) and resurrection of the genus Anematichthys (Bleeker 1859).based on morphological and molecular data of
Southeast Asian Cyprininae (Teleostei, Cypriniformes). Zootaxa. 3586. 41-54.
[51] Kusuma, W.E., S. Ratmuangkhwang, Y. Kumazawa. 2016. Department of Information and Basic Science and Research Center for Biological Diversity, Department of Aquaculture, Faculty of Fisheries and Marine Science, University of Brawijaya, Andaman Coastal Research Station for Development, Faculty of Fisheries, KU.