PAPER • **OPEN ACCESS**

Analysis of Phylogenetics and Structural Relationships of the Cytochrome b Gene Protein in Mahseer Fish (*Tor* sp.)

To cite this article: E Rochmatika *et al* 2024 *IOP Conf. Ser.: Earth Environ. Sci.* **1392** 012008

IOP Conf. Series: Earth and Environmental Science **1392** (2024) 012008 doi:10.1088/1755-1315/1392/1/012008

Analysis of Phylogenetics and Structural Relationships of the Cytochrome b Gene Protein in Mahseer Fish (*Tor* sp.)

E Rochmatika[1], A S Mubarak[2], S Q Tartila[3], Mufasirin[4], T F Haser[5]

[1]Department of Fisheries and Marine Biotechnology, Faculty of Fisheries and Marine Affairs, Airlangga University, 60115 Surabaya, East Java, Indonesia.
[2]Department of Maritime Affairs, Faculty of Fisheries and Marine Affairs, Airlangga University, 60115 Surabaya, East Java, Indonesia.
[3]Department of Aquaculture, Faculty of Agriculture, Tidar University, 59155 Magelang, Central Java, Indonesia.
[4]Department of Parasitology, Faculty of Veterinary Medicine, Airlangga University, 60155 Surabaya, East Java, Indonesia.
[5]Department of Aquaculture, Faculty of Agriculture, Samudra University, 24354 Langsa, Aceh, Indonesia.

Corresponding author: mubarak.as@fpk.unair.ac.id

Abstract. The Cytochrome b gene from mitochondrial DNA encodes a protein that can be used as a genetic marker to identify species. This study aims to determine the bioinformatic characteristics of the gene encoding Cyt b isolated from various Mahseer species, *Tor tambroides*, *Tor douronensis*, *Tor khudree*, and *Tor putitora*. Mahseer Cyt b sequences can be used for detailed computational investigation of the properties of 3D protein models and phylogenetics. The data analysis technique uses nucleotides with a sequence length of 330 bp obtained from NCBI. The MEGA X program was used for phylogenetic analysis. 3D protein structure predictions were obtained through the Swiss model and protein homology program using PYMOL software. The results of phylogenetics show close kinship relationships with the closest genetic distance of 0.000, while the farthest distance is 0.068. The nucleotide composition contains Thymine (30.23%), Cytosine (25.81%), Arginine (28.05%), and Guanine (15.92%). 3D modeling of the Cyt b protein Mahseer referring to the species *Neolissochilus hexagonolepis* (McClelland, 1839). The 3D protein structure prediction was assessed via a Ramachandran plot of 99.07%. This study can be used as a reference to support sustainable conservation efforts by observing the populations and habitat conditions.

1. Introduction

Mahseer fish, also known as *Ikan Dewa*, are among the most important freshwater fish in Southeast Asia. This species boasts high nutritional value, a delicious taste, and a meaty texture, which makes it quite expensive [1,2]. Mahseer fish are benthopelagic species that live in fast-flowing, clear-water rivers with high dissolved oxygen content [3]. Mahseer fish belong to the *Tor* sp. genus, which comprises several species with different morphologies. The diversity of mahseer fish can be observed through phylogenetics, which provides a comparative picture of the evolution and distribution of characteristics among lineages, such as morphology, biology, and molecular traits [4]. The description of phylogeny is represented in the form of a genealogy of living things that have branches resembling a tree [5]. Morphological and molecular data can indicate the evolutionary relationships of the observed species.

The dendrogram of the phylogenetic tree formed based on genetic distance shows that the fish sample groups [6]. Relationships between species of several taxa use sequence or protein data to compare [7,8].

Cyt b is a protein in eukaryotic cells that helps with electron transport during cellular respiration [9]. It is one of the genes in mitochondrial DNA and can be used as a molecular marker for analyzing the broodstock potential for *Tor* sp. domestication. The cyt b gene is a group of mitochondrial genes inherited from the mother in the intact form that play a role in the bioenergetic cycle at the cellular level [10,11,12]. Genetic distance analysis is a matrix calculation of the distance between base pairs between sequences closely related in evolutionary terms [13,14]. This analysis can be used to determine the relationship between species. cyt b protein modeling or homology modeling is used to create an atomic resolution protein model from the amino acid sequence. Using a protein sequence with a similar amino acid sequence, a 3D protein structure can be determined through a template representing the formation of structural genes for protein folding. This can be useful in bioinformatics to determine the 3D structure of proteins [15]. In summary, cyt b plays an important role in cellular respiration and can be used as a molecular marker for analyzing broodstock potential. Genetic distance analysis can help determine the relationship between species, and homology modeling can be used to determine the 3D structure of proteins.

The DNA barcoding method is a quick and accurate way to identify species. By analyzing the cyt b gene marker in the mitochondrial segment, we can trace variations in nucleotide bases and determine the relationship between mahseer fish species in the phylogenetic tree. This study builds upon previous research, which included genetic distance, nucleotide composition, and 3D protein structure analyses. As a result, this study provides the latest information on diversity and relationships, supports sustainable conservation, and serves as a reference from a protein perspective.

2. Material and methods

2.1. Sample collection

This study used comparative data from the GenBank National Centre for Biotechnology Information (www.ncbi.nml.nih.gov). Complete mitochondrial DNA sequences of mahseer fish were 17 partial cytochrome b sequence samples from various species *Tor tambroides, Tor douronensis, Tor khudree*, and *Tor putitora* [16]. The following table 1 presents sample data for cyt b mahseer fish.

Table 1. Partial sequence table of cyt b mitochondrial DNA in mahseer fish.

Species	Cyt b	Sequence Length (pb)	Distribution
Tor tambroides	EF588167	330	Indonesia, Malaysia
	EF588169	330	
	EF588161	330	
	EF588160	330	
	EF588166	330	
	EF588170	330	
Tor douronensis	EF588149	330	Indonesia, Malaysia
	EF588189	330	
	EF588186	330	
Tor khudree	EF588182	330	India
	EF588181	330	
Tor putitora	EF588201	330	
	EF588202	330	Afghanistan,
	EF588203	330	Bangladesh, Bhutan,
	EF588173	330	India, Myanmar, Nepal,
	EF588172	330	Pakistan
	EF588183	330	

Sample site: NCBI, 2023

2.2 Phylogenetics analysis

The phylogenetic tree was designed using the *MEGAX* program using the bootstrap Neighbor-Joining (NJ) method with 1000 repetitions. The Bootstrap Neighbor-Joining method can construct a good phylogenetic tree and select sequences that have the best estimation of the actual distance between the combined sequences. The sequence data for arranging branches on a phylogenetic tree was validated using the bootstrap analysis. Bootstrap values can indicate the stability of branches in the phylogenetic tree. Through the *MEGA version X*, the genetic distance and nucleotide base composition were analyzed [17,18]. Data from NCBI was analyzed, following the bioinformatic analysis. The base sequences were saved in *FASTA* format, then the sequence was aligned/edited with *ClustalW* via *MEGA* to determine the relationship and diversity of Mahseer fish species by reconstructing a phylogenetic tree from the number of differences in nucleotide bases between two or more DNA sequences [19]. In addition, the *MEGA X* analyzed the nucleotide composition and genetic distance between the observed species.

2.3 Predicted protein model

The Swiss model program validated the linkage relationships by observing the protein sequence and 3D protein structure. The protein homology between species can be observed through the *PYMOL* program. The 3D structure prediction of *cyt b* protein is useful for observing differences in protein structure between species. The *cyt b* protein template using the Swiss Model program using *FASTA* data containing amino acids. The Swiss model is an automatic server for comparative modeling of 3D protein structures by inputting the data containing amino acids from target proteins [20,21]. *FASTA* which consists of nucleotide sequences is changed using *BLAST* to obtain the amino acid sequences. Next, the data were downloaded in PDB format to observe the protein homology using *PYMOL* software [22].

3. Result and discussion

3.1. Phylogenetic relationships among samples

Phylogenetics is a method used to determine the kinship relationships among living organisms. Organisms with similar characteristics are considered to be closely related and are believed to have descended from a common female parent [23]. Based on the Neighbor-Joining (NJ) phylogenetic tree analysis, it was observed that *Tor tambroides* and *Tor putitora*, together, form clade I. *Tor khudree*, on the other hand, has a different kinship relationship and is not part of clade I. *Tor douronensis*, with a bootstrap value of 99%, is an out-group having a kinship relationship distinct from the other clades. The phylogenetic analysis of the branch locations of each *Tor* species showed that *Tor tambroides*, *Tor putitora*, and *Tor douronensis* are the closest relatives of each other as they have the shortest branches. According to Zein and Sulandari (2009) [24], the longer the branches of the phylogenetic tree, the greater the evolutionary changes that occur. Despite the evolutionary changes, the analysis still indicates that there are close kinship relations among the species of the genus *Tor*. The phylogenetic tree constructed in this study can be found in Figure 1.

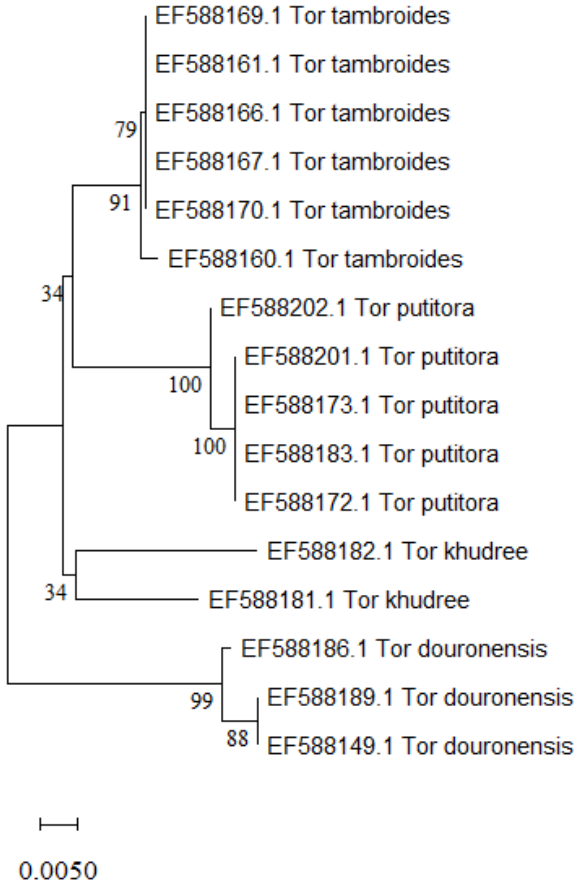

Figure 1. Construction of the cyt b phylogenetic tree of mahseer fish (*Tor* sp.) using the Neighbor-Joining method.

According to Nguyen et al. [36], *Tor* sp. consists of *Tor tambroides, Tor douronensis*, *Tor khudree*, and *Tor putitora*. Every species has its distinctive holotype codes, but they all share a close relationship within a phylogenetic tree. The *Tor douronensis, Tor tambroides, Tor putitora, Tor khudree* still have the same distribution in the Sunda shelf. Tang et al., [26] stated that the ancient river network areas on the Sunda Shelf show morphological and genetic similarities. The *Tor douronensis* and *Tor tambroides* which are distributed in Indonesia are also part of the Sunda shelf. Ancient rivers were formed during the Pleistocene era, with the East Sunda ancient river emptying into the Java Sea, while the North Sunda River flows into the Natuna Sea. The separation of Java and Sumatra occurred around the mid-Miocene era when the polar ice melted, causing ancient rivers in the Sunda Shelf area to be covered and rising seawater conditions causing land formation [26,27,28].

Santoso et al. (2021) [16] explained that mahseer fish produced Random Amplified Polymorphic DNA with a heterozygosity value of 0.92 and a polymorphic percentage of 21.3%. This shows that the mitochondrial DNA in the Mahseer fish species is still in the category of having low DNA variation. However, mahseer fish have a unique form of mitochondrial DNA. Phylogenetic analysis is strengthened by the calculated value of genetic distance (pairwise distance). This study resulted in the closest genetic distance of mahseer fish being 0.00 and the furthest being 0.068 in the *Tor khudree* species with the code EF588182. Gusmiaty et al. [29] stated that species in the same area will have a small genetic distance and have a high genetic similarity value. This long genetic distance is possible due to the geographical barrier between the two [30]. The results of study on the composition of nucleotide bases between species show A=28.05%, C=25.81%, G=15.92%, T=30.23% with total

nucleotide bases A+T=58.28%, while G +C=41.73%. This is by the statement of Rahayu et al. [31] that the nucleotide base composition of *Tor* sp. is A=26.79%, C=23.16%, G=19.17% and T=30.93%. The total nucleotide bases A+T are 57.73%, while G+C is 42.33%, the GC < AT value is relatively balanced, and generally the GC content in vertebrates is 40-45%. The genetic distance and nucleotide composition are presented in tables 2 and 3.

Table 2. Nucleotide composition table of cyt b mitochondrial DNA in mahseer fish

No	Sample	Gen cyt b (%)				Total
		T(U)	C	A	G	
1	EF588202.1 Tor putitora	29,70	26,36	28,18	15,76	330
2	EF588201.1 Tor putitora	29,70	26,36	27,88	16,06	330
3	EF588189.1 Tor douronensis	29,09	26,97	28,18	15,76	330
4	EF588186.1 Tor douronensis	29,09	26,97	27,58	16,36	330
5	EF588183.1 Tor putitora	29,48	26,44	27,96	16,11	329
6	EF588182.1 Tor khudree	29,70	26,36	27,88	16,06	330
7	EF588181.1 Tor khudree	31,52	24,55	28,18	15,76	330
8	EF588173.1 Tor putitora	29,70	26,36	27,88	16,06	330
9	EF588172.1 Tor putitora	29,70	26,36	27,88	16,06	330
10	EF588170.1 Tor tambroides	31,21	24,85	28,18	15,76	330
11	EF588169.1 Tor tambroides	31,00	24,92	28,27	15,81	329
12	EF588167.1 Tor tambroides	31,21	24,85	28,18	15,76	330
13	EF588166.1 Tor tambroides	31,21	24,85	28,18	15,76	330
14	EF588161.1 Tor tambroides	31,00	24,92	28,27	15,81	329
15	EF588160.1 Tor tambroides	31,21	24,85	27,88	16,06	330
16	EF588149.1 Tor douronensis	29,09	26,97	28,18	15,76	330
Average amount		30,23	25,81	28,05	15,92	30,23

Table 3. Genetic distance table of cyt b mitochondrial DNA in mahseer fish.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.EF588202.1_Tor_putitora																
2.EF588201.1_Tor_putitora	0,003															
3.EF588189.1_Tor_douronensis	0,058	0,061														
4.EF588186.1_Tor_douronensis	0,058	0,061	0,006													
5.EF588183.1_Tor_putitora	0,003	0,000	0,062	0,062												
6.EF588182.1_Tor_khudree	0,041	0,044	0,068	0,061	0,045											
7.EF588181.1_Tor_khudree	0,044	0,048	0,058	0,058	0,048	0,041										
8.EF588173.1_Tor_putitora	0,003	0,000	0,061	0,061	0,000	0,044	0,048									
9.EF588172.1_Tor_putitora	0,003	0,000	0,061	0,061	0,000	0,044	0,048	0,000								
10.EF588170.1_Tor_tambroides	0,028	0,031	0,055	0,048	0,031	0,038	0,028	0,031	0,031							
11.EF588169.1_Tor_tambroides	0,028	0,031	0,055	0,048	0,031	0,038	0,028	0,031	0,031	0,000						
12.EF588167.1_Tor_tambroides	0,028	0,031	0,055	0,048	0,031	0,038	0,028	0,031	0,031	0,000	0,000					
13.EF588166.1_Tor_tambroides	0,028	0,031	0,055	0,048	0,031	0,038	0,028	0,031	0,031	0,000	0,000	0,000				
14.EF588161.1_Tor_tambroides	0,028	0,031	0,055	0,048	0,031	0,038	0,028	0,031	0,031	0,000	0,000	0,000	0,000			
15.EF588160.1_Tor_tambroides	0,031	0,035	0,051	0,044	0,035	0,041	0,031	0,035	0,035	0,003	0,003	0,003	0,003	0,003		
16.EF588149.1_Tor_douronensis	0,058	0,061	0,000	0,006	0,062	0,068	0,058	0,061	0,061	0,055	0,055	0,055	0,055	0,055	0,051	

3.2. Structure of 3D Protein

Proteins are a series of complex biomolecules that play an important role in living things and are formed through the stages of transcription and translation. The resulting protein must go through a translation process from RNA, while the process to obtain RNA is through a transcription process from DNA. 3D model estimation aims to determine protein structure based on amino acid sequences. There are changes in the composition of nucleotide bases from DNA to RNA and then protein [32]. This can be applied to observe the description of protein homology between Mahseer fish species. The 3D structure of proteins provides an overview of the function and mechanism of biological action [33]. The results of the 3D protein model estimation are shown in Figure 2. The 3D model estimation of *Cyt b* of Mahseer fish between species obtained almost the same results. There are 4 different fragments, marked by different colors, red with a combination of yellow, blue, light green, and cyan blue, and is hydrophobic. 3D modeling of the *cyt b* protein in four species of Mahseer fish refers to the species *Neolissochilus hexagonolepis* (Copper masheer) (McClelland, 1839). Based on validation via the Ramachandran plot, a value of 99.07% was obtained, Komari et al. [34] stated that the Ramachandran plot value has very good protein structure quality.

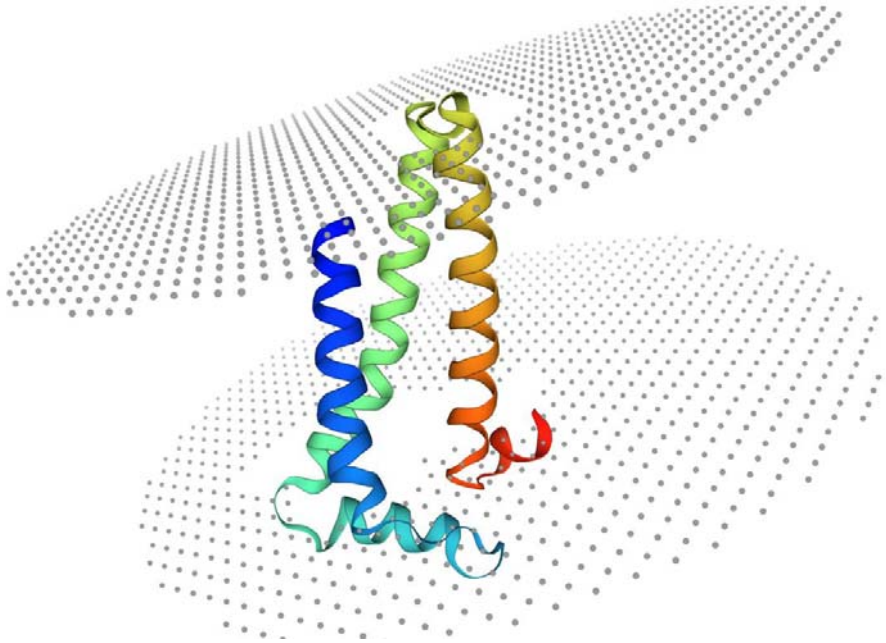

Figure 2. 3D dimensional model of cyt b mahseer fish (*Tor* sp.)

4. Conclusion

The phylogenetics of several Mahseer fish species show a close relationship. The geographical factors that determine the distribution of *Tor douronensis, Tor tambroides, Tor putitora, Tor khudree* are still in the Sunda shelf area, so they have morphological and genetic similarities. Genetic distance shows the closest relationship between species with relatively similar nucleotide composition. 3D protein predictions produced between species show almost the same protein structure. The 3D protein structure refers to an existing protein template, namely *Neolissochilus hexagonolepis* (copper masheer) (McClelland, 1839). The genetic information and molecular taxonomic status that have been identified in this study can be used as a source of the latest information for further studies on Mahseer fish. This study aimed to increase and support the quantity and quality of populations or individuals as a sustainable conservation effort. The use of more varied molecular markers is also needed, accompanied by predictions of the 3D protein structure of each species.

5. References

[1] Lau M M L, Lim W K, Ishak S D M, Munafi A B A, Chung H H 2021 *Proc Zool Soc.* 74 pp 227-237.
[2] Muchlisin Z A, Nur F M, Maulida S, Handayani L S, and Rahayu S R 2022 *E3S Web Conf.* 339 p 03006.
[3] Marson 2013 *J. Fish*. 2 pp 5-7.
[4] Malabarba L R, and Malabarba M C 2019 *J. Elsevier Inc.* 1 pp 1-19.
[5] Lubis K 2014 *J. Pengabdian Kepada Masyarakat* 20 pp 66-69.
[6] Nugraha B, Novianto D, and Barata A 2011 *J. Lit. Perikan. Ind*. 17 pp 277-284.
[7] Akbar N, Aris M, Irfan M, Tahir I, and Baksir A 2018 *J. Indo Mar Sci and Tech.* 11 p 120.
[8] Hall B. G. 2013 *J. Molecular Bio Evol.* 30 pp 1229-1235.
[9] Ghosh R, Upadhyay A D, Roy A K, and Tiwari A 2020 *J. Ento Zoo Stud.* 8 pp *540-550.*
[10] Berry E A, Guergova-Kuras M, Huang L S, and Crofts A R 2000 *Annual Rev Biochem.* 69 pp 1005-1075.
[11] Clark-Walker G D, Rittenberg B, and Lascelles J 1967 J. *Bact.* 94 pp 1648-1655.
[12] Maulid D Y, Nurilmala M, Nurjanah, and Madduppa H 2016 *J. Pengolah Hasil Perikan Indo.* 19 pp 9-16.
[13] Brahmantiyo B, Priyono P, and Rosartio R 2016 *J Vet.* 17 pp 226–234.
[14] Monalisa E, Mantiri F R, and Lengkong H J 2019 *J. MIPA* 8 p 71.
[15] Martí-Renom M A, Stuart A C, Fiser A, Sánchez R, Melo F, and Sali A 2000 *Annual Rev Biophys Biomol Struct.* 29 pp 291-325.
[16] Santosa T A, Yulianti S, Razak A, Ahda Y, and Ferdyan R 2021 *Prosiding Seminar Nasional Biologi (Padang)* 1 pp 312-318.
[17] Abdullah A, Sativa H A, Nurhayati T, and Nurilmala M 2019 *J. Pengolahan Hasil Perikan Indo.* 22 pp 508-519.
[18] Dharmayanti N L P 2011 Wartazoa 21 pp 1-10
[19] Saleky D, Supriyatin F E, and Dailami M 2020 *J. Kel Trop.* 23 pp 305-315.
[20] Schwede T, Kopp J, Guex N, and Peitsch M C 2003 *Nucleic Acids Stud. 31* pp 3381–3385.
[21] Saudale F Z 2020 *Al-Kimia J. Chem. 8* pp 93-103
[22] Kallberg Y, Segerstolpe Å, Lackmann F, Persson B, and Wieslander L 2012 *J. PloS one 7* e43786.
[23] Astarini I A, Ardiana S A, Putra I N G, Pertiwi P D, Sembiring A, Yusmalinda A, Al Malik D 2021 *Musamus Fish and Mar J.* 3 pp 107-115
[24] Zein M S A, and Sulandari S 2009 *J. Vet.* 10 pp 41-49.
[25] Nguyen T T, Na-Nakorn U, Sukmanomon S, and Ziming C 2008 *Mol Phylo Evol. 48* pp 1223-1231.
[26] Tang K L, Agnew M K, Hirt M V, Sado T, Schneider L M, Freyhof J, Sulaiman Z, Swartz E, Vidthayann C, Miya M, Saitoh K, Simons A M, Wood R M, and Mayden R L 2010 *Mol Phyllo Evol.* 57 pp 189-214.
[27] Jaafar F, Na-Nakorn U, Srisapoome P, Amornsakun T, Duong T Y, Gonzales-Plasus M M, Hoang D H, and Parhar I S 2021 *J. Biology* 10 p 286.
[28] Karin B R, Das I, Jackman T R, and Bauer A M 2017 *PeerJ 5* e3762.
[29] Gusmiaty G, Restu M, Asrianny A, and Larekeng S H 2017 *J. Natur Indo.* 16 p 47.
[30] Kurniawan A, Pramono, D Y, Indrayati A, Triswiyana I. 2022 *Acta Aquatica: Aqua Sci J.* 9 pp 164-169.
[31] Rahayu D A, Nugroho E D, Haryono, Kurniawan N, and Azrianingsih R 2015 *Pros Sem Nas Ikan Masy Iktio Indo.* (Bogor) Jilid 1 pp 67-85.
[32] Haryanto T and Ardi B S 2015 *J. Eduk Penelitian Infor.* 1 pp 133-138.
[33] Shi Y 2014 *Cell* 159 pp 995–1014.
[34] Komari N, Hadi S, and Suhartono E 2020 *J. Jejaring Matematika Sains* 2 pp 65-70.

6. Acknowledgement

The author would like to thank the assistance of Mr. Ahmad Shofy Mubarok head of the biotechnology and fisheries master's study program, at the Faculty of Fisheries and Marine Affairs, Airlangga

University, who was willing to join this article, helped finance the publication, and is willing to provide criticism to sharpen the direction of this article. We would also like to thank the editor and reviewer for their suggestions, which improved our manuscript.

CRediT authorship contribution statement

Endah Rochmatika: Analyzed and interpreted the data; Wrote the paper
Ahmad Shofy Mubarok: Contributed funding acquisition; Writing-review
Shobrina Silmi Qori Tartila: Analyzed and interpreted the data; Writing review
Mufasirin: Contributed materials, analysis tools, or data; Interpreted the data
Teuku Fadlon Haser: Contributed materials, analysis tools, or data; Conceptualization

Data availability statement

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no competing interests.

Ethical approval and consent to participate

All the authors consented to the publication of this work.