Analysis of mitochondrial control region sequence of tapah fish (*Wallago leerii*) from Riau Province, Indonesia

Roza Elvyra, Afdizan

Department of Biology, Faculty of Mathematics and Natural Science, Universitas Riau. Jl. HR. Soebrantas Km. 12.5 Panam, 28293 Pekanbaru, Riau, Indonesia. Corresponding author: R. Elvyra, roza.elvyra@gmail.com

Abstract. *Wallago leerii* is a member of the genus *Wallago*, and it is a species of high economic value. This fish is found in the rivers of Riau Province. Identification and characterization of a species is very important in the taxonomic and phylogenetic studies of species. An identification using genetic markers through mitochondrial DNA control region sequence needs to be accomplished. This study aims to determine the characteristics of the control region sequence of *Wallago leerii* from Riau Province. The multiple alignment of nucleotides sequences among *W. leerii* from Riau Province in comparison with GenBank data at the National Center for Biotechnology Information (NCBI 2020) results in sequence of 537 nucleotides. The nucleotide base composition in the control region sequence shows that Adenine (A) > Thymine (T), and Guanine (G) < Cytosine (C). The genetic distance among *W. leerii* with comparison of other species ranged from 0.000-0.299. The average genetic distance in the *W. leerii* population is in the range of 0.000-0.024. Phylogenetic tree reconstruction using the control region sequence of mitochondrial DNA demonstrated that this sequence could differentiate *Wallago leerii* with other species.

Key Words: genetic identification, phylogenetic, taxonomy.

Introduction. *Wallago* is a genus of fish that has a high economic value. One species of important fish as food which belongs to the genus of *Wallago*, is *Wallago leerii*. Belonging to the Siluridae family, *Wallago leerii* has a large body with a length up to 1.5 m, and it is present in Thailand, Peninsular Malaysia, Singapore, Borneo, and Sumatera waters (Ng 1992; Kottelat et al 1993). One of the habitats of the fish in Riau is represented by rivers, specifically the Kampar, Tapung and Indragiri rivers.

The identification and characterization of the diversity of a species is very important in the taxonomic and phylogenetic studies of species. Complemented with molecular data, the morphological study will be more comprehensive (Purnama et al 2019; Elvyra et al 2020). Currently, the concept of identification using genetic markers is an alternative that has been widely used by researchers. Genes in mitochondrial DNA (mtDNA) are often used as target genes in species identification (Kyle & Wilson 2007). Among the DNA markers, mitochondrial DNA sequence has been widely used in population genetic studies (Avise 2000; Kumar et al 2014). Mitochondrial DNA can be used for population genetics and phylogenetic studies, because it has a type of maternal inheritance, compact gene packing with little noncoding intergenic nucleotides, and a considerable number of copies or multicopy in a cell (Pereira 2000).

Control region is formed when there is a single non-coding segment in the mitochondrial DNA (Kocher et al 1989). The control region is one of most interesting to study the vertebrate mitochondrial genome, because of the short, conserved elements and rapid rate of evolution (Guo et al 2003). The control region is the most rapidly evolving region of mitochondrial DNA, with substitution rates 5 to 10 times higher than for other parts (Yu et al 2005). The mitochondrial control region, also called the displacement loop, or D-loop, is responsible for transcription of mitochondrial genes and contains the point of origin of heavy strand replication (Suarez et al 2001). The high

mutation rate in the hypervariable control region has served as a marker for assessing intra-species variation (Kuehn et al 2003; Kumar et al 2017). Therefore, a study of the characteristics of the control region sequence on *Wallago leerii* species needs to be carried out for identification and phylogenetic relationship studies of *Wallago leerii* originating from Riau Province.

Material and Method. The research was conducted from March to July 2020. The *W. leerii* species was used as the sample of this study. Fish samples were obtained from local fishermen in Kampar, Tapung and Indragiri rivers, Riau Province, Indonesia. There are six samples in total consisting of two samples from each river in Kampar, Tapung and Indragiri. Fish identification was carried out based on Kottelat et al (1993) indications. Fish muscle tissue was taken (25-50 mg) and put into a 96% ethanol solution, then stored in the freezer. This research was conducted at the Genetics Laboratory, Department of Biology, Faculty of Mathematics and Natural Science, University of Riau.

Isolation of total DNA. Total genome DNA was extracted from fish muscle tissue using Dneasy Blood and Tissue Kit (Qiagen). The total DNA isolation process was carried out following the kit protocol. The quality and quantity of DNA were determined using electrophoresis techniques on 1.2% agarose gel.

DNA amplification and sequencing. Total DNA was amplified using the Polymerase Chain Reaction (PCR) technique. The PCR reaction followed the Dream Taq DNA Polymerase Kit (Thermo Scientific) protocol. The primer pairs used in this case are designed based on the D-Loop sequence available in the GenBank database (NCBI 2020) at www.ncbi.nlm.nih.gov, namely: CTRG1_F (5'AAGCACCGGTCTTGTAATCC3') and CTRG2_R (5'GCCTTTAAATGAACGCCTTG3'). The PCR process includes pre-PCR at a temperature of 95^{o}C for 5 minutes. The 35 PCR cycles encompass the denaturation stage at a temperature of 94^{o}C for 30 seconds, the primer annealing stage at a temperature of 58.8^{o}C for 1 minute, and the elongation stage at a temperature of 72^{o}C for 1 minute 30 seconds. Finally, the Post PCR stage at a temperature of 72^{o}C lasted for 5 minutes. The success of PCR amplification was detected by electrophoresis on 1.2% agarose gel. The PCR products are then sent to PT. Genetica Science Indonesia in Jakarta to sequence the nucleotide bases. The sequencing is done in one direction using the same forward primer as the primer in the PCR process.

Data analysis. The Control Region sequences are then homologized and aligned. The reference sequence used as a comparison from GenBank (NCBI 2020) at www.ncbi.nlm.nih.gov is *Wallago attu* (MN495550.1) and as an outgroup is *Mystus cavasius* (NC030187.1). Data analysis was performed using the MEGA program version 6.06 (Tamura et al 2013). The data analyzed included nucleotide composition, genetic distance and phylogenetic trees using the Neighbor-Joining method with 1000 times bootstrap test.

Results and Discussion. The results of multiple alignment of Control Region mitochondrial DNA sequences among *Wallago leerii* contrasted to the GenBank data (NCBI 2020) at www.ncbi.nlm.nih.gov and showed a sequence of 537 nucleotides. The nucleotide base composition of the Control Region sequence has an average of Adenine (A) bases of 37.5%, Thymine (T) bases of 29.6%, Cytosine (C) bases of 19.7% and Guanine (G) bases of 13.2% (Table 1). The results of the base composition in the Control Region sequence obtained show that Adenine (A) > Thymine (T), and Guanine (G) < Cytosine (C). Overall, the nucleotide composition of A + T is greater than G + C with an average A + T (67.1%) and G + C (32.9%), where the highest A + T composition is found in the *W. leerii* 4 (Tapung river) (68.1%), and the lowest A + T composition was found in the species *M. cavasius* (64.5%).

Table 1

The nucleotide composition of Control Region sequences of *W. leerii* from Riau Province

Species	*T(U)*	*C*	*A*	*G*	*A+T*	*G+C*
1	29.6	19.2	37.7	13.5	67.3	32.7
2	29.6	19.2	37.7	13.5	67.3	32.7
3	29.2	19.4	38.1	13.3	67.3	32.7
4	29.8	19.0	38.3	12.9	68.1	31.9
5	29.6	19.2	37.7	13.5	67.3	32.7
6	29.6	19.2	37.7	13.5	67.3	32.7
7	30.6	20.0	36.9	12.5	67.5	32.5
8	28.5	22.1	36.0	13.4	64.5	35.5
Average	29.6	19.7	37.5	13.2	67.1	32.9

Note: (1) *Wallago leerii* 1 (Kampar River), (2) *Wallago leerii* 2 (Kampar River), (3) *Wallago leerii* 3 (Tapung River), (4) *Wallago leerii* 4 (Tapung River), (5) *Wallago leerii* 5 (Indragiri River), (6) *Wallago leerii* 6 (Indragiri River), (7) *Wallago attu* (GenBank at Access Number MN495550.1) (NCBI 2020), (8) *Mystus cavasius* (GenBank at Access Number NC030187.1) (NCBI 2020).

The incompatible base composition in the Control Region sequence produces an asymmetrical pair. Reyes et al (1998) stated that the mitochondrial DNA replication process has an asymmetric nucleotide distribution, giving rise to heavy strands (lots of G, C) and light strands (lots of A, T).

Table 2

Nucleotide transition mutations of Control Region sequence of *W. leerii* from Riau Province

Species	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1								
2	0							
3	10	10						
4	12	12	6					
5	0	0	10	12				
6	0	0	10	12	0			
7	59	59	56	56	59	59		
8	65	65	64	62	65	65	58	

Note: (1) *Wallago leerii* 1 (Kampar River), (2) *Wallago leerii* 2 (Kampar River), (3) *Wallago leerii* 3 (Tapung River), (4) *Wallago leerii* 4 (Tapung River), (5) *Wallago leerii* 5 (Indragiri River), (6) *Wallago leerii* 6 (Indragiri River), (7) *Wallago attu* (GenBank at Access Number MN495550.1) (NCBI 2020), (8) *Mystus cavasius* (GenBank at Access Number NC030187.1) (NCBI 2020).

The results of the comparison of nucleotide sequences from the research conducted, showed the highest difference in the control region nucleotide transition substitution of 65 nucleotides, found among the specimens of *W. leerii* 1 (Kampar River), *W. leerii* 2 (Kampar River), *W. leerii* 5 (Indragiri River) and *W. leerii* 6 (Indragiri River), with *M. cavasius* (GenBank NC030187.1) (NCBI 2020). The lowest transition substitution value

was found in *W. leerii* fish population in Riau Province with a range of 0-12 nucleotides (Table 2). Meanwhile, the results of the analysis of the highest difference in nucleotide sequences in the Control Region sequence transversion substitution, identified 85 nucleotides among specimens of *W. leerii* 1 (Kampar River), *W. leerii* 2 (Kampar River), *W. leerii* 4 (Tapung River), *W. leerii* 5 (Indragiri River), *W. leerii* 6 (Indragiri River) and *W. attu* (GenBank MN495550.1) (NCBI 2020), with *M. cavasius* (GenBank NC030187.1) (NCBI 2020). The lowest transversion substitution value is found in the *W. leerii* fish population from Riau Province with a value range of 0-1 base (Table 3).

Table 3

Nucleotide transversion mutations of Control Region sequence of *W. leerii* from Riau Province

Species	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1								
2	0							
3	1	1						
4	0	0	1					
5	0	0	1	0				
6	0	0	1	0	0			
7	78	78	77	78	78	78		
8	85	85	84	85	85	85	85	

Note: (1) *Wallago leerii* 1 (Kampar River), (2) *Wallago leerii* 2 (Kampar River), (3) *Wallago leerii* 3 (Tapung River), (4) *Wallago leerii* 4 (Tapung River), (5) *Wallago leerii* 5 (Indragiri River), (6) *Wallago leerii* 6 (Indragiri River), (7) *Wallago attu* (GenBank at Access Number MN495550.1) (NCBI 2020), (8) *Mystus cavasius* (GenBank at Access Number NC030187.1) (NCBI 2020).

In this study, it was found that the number of transition mutations was more dominant than the transversion mutations in the populations of *W. leerii* from Riau Province. According to Kocher et al (1989), nucleotide substitution of transition occurs more often than transversion substitution in patterns of mitochondrial DNA sequence.

The results of alignment of the Control Region sequence among *W. leerii* with the other species showed that the genetic distance ranged from 0.000-0.299. The average genetic distance in the *W. leerii* population is in the range of 0.000-0.024. The closest genetic distance among the *W. leerii* population exists in Riau rivers. The difference in genetic distance is due to differences in habitat and distribution patterns of the species. According to Akbar et al (2018), the difference in genetic distance is caused by different distribution patterns so that the chance of meeting is small, resulting in no gene flow between populations. Meanwhile, the farthest genetic distance from the entire sequence was found in the species *M. cavasius* (out-group) against the *W. leerii* from Riau Province with a genetic distance value of 0.293-0.299. The results of the analysis of genetic distances between species are presented in Table 4.

Table 4

The genetic distance among *W. leerii* populations from Riau Province based on the Control Region sequence

Species	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1								
2	0.000							
3	0.022	0.022						
4	0.024	0.024	0.014					
5	0.000	0.000	0.022	0.024				
6	0.000	0.000	0.022	0.024	0.000			
7	0.273	0.273	0.265	0.267	0.273	0.273		
8	0.299	0.299	0.295	0.293	0.299	0.299	0.285	

Note: (1) *Wallago leerii* 1 (Kampar River), (2) *Wallago leerii* 2 (Kampar River), (3) *Wallago leerii* 3 (Tapung River), (4) *Wallago leerii* 4 (Tapung River), (5) *Wallago leerii* 5 (Indragiri River), (6) *Wallago leerii* 6 (Indragiri River), (7) *Wallago attu* (GenBank at Access Number MN495550.1) (NCBI 2020), (8) *Mystus cavasius* (GenBank at Access Number NC030187.1) (NCBI 2020).

The high value of genetic distance is caused by the high level of variability of the Control Region sequence. Hebert et al (2004) stated that a good genetic distance is 10 times greater than the value of the intraspecies genetic distance. The genetic distance between *W. leerii* and *W. attu* which belong to one genus ranges from 0.265 to 0.273. Meanwhile, the significant genetic distance value in the Control Region sequence is shown in the comparison of out-group species with *Wallago* representatives from Riau Province with the farthest genetic distance value of 0.299. This farthest genetic distance is not only caused by different genera, but also by different geographic locations which do not allow genetic flow to occur.

Although there are sub-groups between individuals originating from Kampar and Indragiri rivers, and the sub-group in the Tapung river, the reconstruction of phylogenetic trees of the Control Region sequence using the Neighbor Joining method (1000 replications) (Tamura et al 2013) resulted in a large group of *W. leerii* from Riau Province (Figure 1). The *W. leerii* from Riau Province is separated from *W. attu* (GenBank MN495550.1) and *M. cavasius* (GenBank NC030187.1) (NCBI 2020).

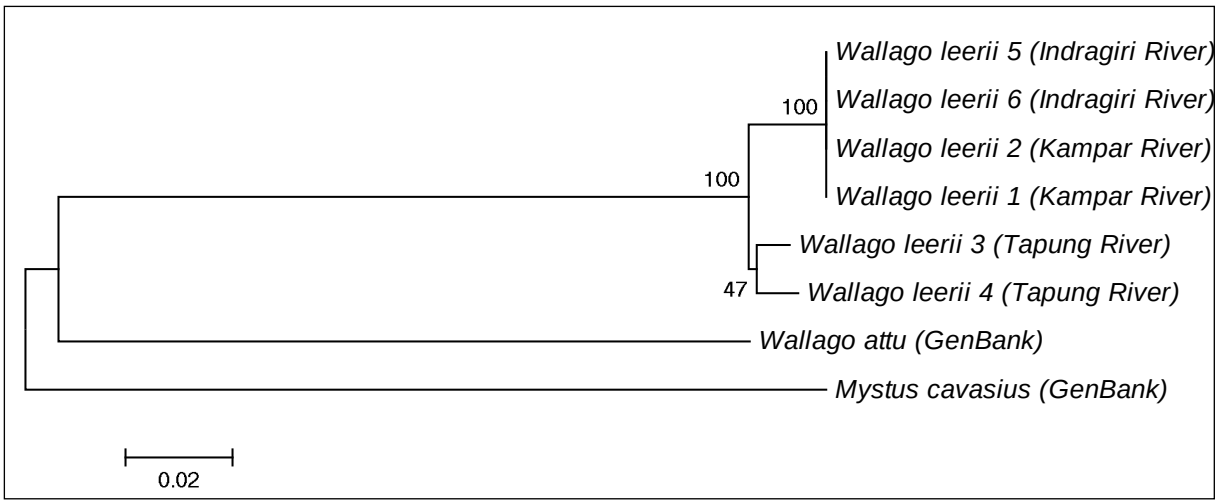

Figure 1. Phylogenetic tree of *W. leerii* from Riau Province based on the Control Region sequence.

The reconstruction of the phylogenetic tree using the mitochondrial DNA Control Region sequence shows that, this sequence can differentiate *W. leerii* specimens at interspecies level. This accuracy can be seen by the formation of a separate main cluster between the species *W. leerii*, *W. attu* and *M. cavasius* (out-group). Hillis and Bull (1993) state that the bootstraps value ≥70% is the minimum value to obtain a precise phylogram accuracy, while the bootstrap value ≥95% is an indication that the resulting phylogram has a high degree of accuracy or is the shape of the actual cluster. Meanwhile at the intraspecies level, the Control Region sequence resulted in two sub groups, namely the *W. leerii* populations from the Indragiri and Kampar Rivers, and the *W. leerii* from the Tapung River.

Conclusions. Based on the mitochondrial DNA Control Region sequence of *W. leerii* specimens, a 537 nucleotides sequence was obtained. The composition of the Control Region sequence has an average of Adenine (A) bases of 37.5%, Thymine (T) bases of 29.6%, Cytosine (C) bases of 19.7% and Guanine (G) bases of 13.2%. The mutations in the Control Region sequence indicate that the mutation in transition > transversion. The reconstruction of the phylogenetic tree using the Control Region sequence of the mitochondrial DNA shows that, this sequence can differentiate *W. leerii* specimens at interspecies level. Meanwhile, at intraspecies level, *Wallago leerii* formed two sub groups between river locations.

Acknowledgements. This research was funded by a grant from the Ministry of Research and Technology/National Research and Innovation Agency, Republic of Indonesia (2020) under "The University Excellence Basic Research Grant" on the behalf of Dr. Roza Elvyra.

Conflict of Interest. The authors declare no conflict of interest.

References

Akbar N., Aris M., Irfan M., Tahir I., Baksir A., Surahman, Madduppa H. H., Kotta R., 2018 [Phylogenetic of tuna fish (*Thunnus* spp.) in North Mollucas Sea, Indonesia]. Jurnal Iktiologi Indonesia 18(1):1-11 [in Indonesian].

Avise J. C., 2000 Phylogeography: The history and formation of species. Harvard University Press, Cambridge MA.

Elvyra R., Solihin D. D., Affandi R., Junior M. Z., Suhendra M., 2020 Molecular characteristics and phylogenetic relationships of silurid catfishes (*Kryptopterus*, *Ompok* and *Phalacronotus*) from the Kampar River, Indonesia, based on the cytochrome b gene. Biodiversitas 21(8):3539-3546.

Guo X., Liu S., Liu Y., 2003 Comparative analysis of the mitochondrial DNA control region in Cyprinids with different ploidy level. Aquaculture 224:25-38.

Hebert P. D. N., Stoeckle M. Y., Zemlak T. S., Francis C. M., 2004 Identification of birds through DNA barcodes. PLoS Biology 2(10):1657–1663.

Hillis D. M., Bull J. J., 1993 An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42(2):182-192.

Kocher T. D., Thomas W. K., Meyer A., Edwards S. V., Paabo S., Villablanca F. X., Wilson A. C., 1989 Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86:6196-6200.

Kottelat M., Whitten A. J., Kartikasari S. N., Wirjoatmodjo S., 1993 Freshwater fishes of western Indonesia and Sulawesi. Periplus Editions Ltd., Indonesia.

Kuehn R., Schroeder W., Pirchner F., Rottmann O., 2003 Genetic diversity, gene flow and drift in Bavarian red deer populations (*Cervus elaphus*). Conservation Genetics 4:157–166.

Kumar G., Kocour M., Kunal S. P., 2014 Mitochondrial DNA variation and phylogenetic relationships among five tuna species based on sequencing of D-loop region. Mitochondrial DNA, Early Online:1–5.

Kumar A., Ghazi M. G. U., Hussain S. A., Bhatt D., Gupta S. K., 2017 Mitochondrial and nuclear DNA based genetic assessment indicated distinct variation and low genetic

exchange among the three subspecies of swamp deer (*Rucervus duvaucelii*). Evol Biol 44:31–42.
Kyle C. J., Wilson C. C., 2007 Mitochondrial DNA identification of game and harvested freshwater fish species. Forensic Science International 166:68–76.
Ng P. K. L., 1992 The giant Malayan catfish, *Wallago leerii* Bleeker, 1851, and the identities of *Wallagonia tweediei* Hora & Misra, 1941, and *Wallago maculatus* Inger & Chin, 1959 (Teleostei: Siluridae). Raffles Bulletin of Zoology 40(2):245-263.
Pereira S. L., 2000 Mitochondrial genome organization and vertebrate phylogenetics. Gen Mol Biol 23:745-752.
Purnama A. A, Mubarak J., Daruwati I., Roslim D. I., Elvyra R., 2019 First report of morphological and molecular identification of greater scissortail *Rasbora caudimaculata* from Rokan Hulu District, Riau Province, Indonesia. AACL Bioflux 12(1):34-41.
Reyes A., Gissi C., Pesole G., Saccone C., 1998 Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol Biol Evol 15(8):957–966.
Suarez J., Bautista J. M., Almodovar A., Machordom A., 2001 Evolution of the mitochondrial control region in Palaearctic brown trout (*Salmo trutta*) populations: the biogeographical role of the Iberian Peninsula. Heredity 87:198–206.
Tamura K., Stecher G., Peterson D., Filipski A., Kumar S., 2013 MEGA 6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725-2729.
Yu J. Z., Kong X., Yu Z., Zhou L., 2005 Comparative analysis of mitochondrial control region sequence from three flatfish species (Pleuronectidae). Journal of Ocean University of China (Oceanic and Coastal Sea Research) 4(1):80-84.
*** National Center for Biotechnology Information (NCBI), 2020 The National Center for Biotechnology Information (NCBI). Available at: http://www.ncbi.nlm.nih.gov/. Accessed: March 2020.

Received: 8 December 2020. Accepted: 24 February 2021. Published online: 30 December 2021.
Authors:
Roza Elvyra, Department of Biology, Faculty of Mathematics and Natural Science, Universitas Riau, 28293 Pekanbaru, Riau, Indonesia, e-mail: roza.elvyra@gmail.com
Afdizan, Department of Biology, Faculty of Mathematics and Natural Science, Universitas Riau, 28293 Pekanbaru, Riau, Indonesia, e-mail: afdizanj@gmail.com

How to cite this article:
Elvyra R., Afdizan, 2021 Analysis of mitochondrial control region sequence of tapah fish (*Wallago leerii*) from Riau Province, Indonesia. AACL Bioflux 14(6):3799-3805.