BIODIVERSITAS
Volume 24, Number 4, April 2023
Pages: 2251-2260

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d240438

The importance of DNA barcode reference libraries and selection primer pair in monitoring fish diversity using environmental DNA metabarcoding

DEWI IMELDA ROESMA♥, DJONG HON TJONG, SYAIFULLAH, NOFRITA, MUHAMMAD NAZRI JANRA, FURQAN DWIKI LINTANG PRAWIRA, VIOLA MUTIARA SALIS, DYTA RABBANI AIDIL

Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Andalas. Jl. Limau Manis, Padang 25175, West Sumatra, Indonesia. Tel./fax.: +62-751-71671, ♥email: dewiroesma@sci.unand.ac.id

Manuscript received: 13 December 2022. Revision accepted: 25 April 2023.

Abstract. *Roesma DI, Tjong DH, Syaifullah, Nofrita, Janra MN, Prawira FDL, Salis VM, Aidil DR. 2023. The importance of DNA barcode reference libraries and selection primer pair in monitoring fish diversity using environmental DNA metabarcoding. Biodiversitas 24: 2251-2260.* Environmental DNA (eDNA) metabarcoding has become an alternative method used for biodiversity monitoring of an ecosystem. The eDNA metabarcoding has advantages compared to the conventional method because it is non-invasive, quick, and requires less cost. However, the effectiveness of the eDNA method is highly dependent on the coverage of the DNA barcode reference and primer pair. A study using the eDNA method was conducted for fish biodiversity monitoring in Singkarak Lake. Two-liter water samples were collected using sterile bottle samples at each sampling site (five sites). The universal primers (Fish FI and Fish R1) used for Next-generation sequencing (GRIDION, Nanopore, Oxford Technologies). The study detected 152 fish species using eDNA metabarcoding. Ten species out of the 30 originally reported in Singkarak Lake were detected using eDNA metabarcoding. The low percentage of fish detected is thought to be due to several factors; incomplete/unavailability of freshwater fish DNA barcodes in Indonesia registered in the database repository, inappropriate primer pair selection, low DNA quality, and the absence of target species DNA in collected water samples. The results demonstrated the significance of correctly registering DNA barcodes to the database and appropriate primer pair selection to identify eDNA metabarcoding. This study provides recommendations using eDNA metabarcoding for monitoring in future work.

Keywords: DNA barcode, eDNA metabarcoding, next-generation sequencing, primer pair**,** Singkarak Lake

INTRODUCTION

Indonesia has the second-highest mega biodiversity in the world (von Rintelen et al. 2017). Unfortunately, not all biodiversity data is known related to the number and types of species, especially freshwater fish groups. Data from various databases (library repositories, official websites, articles, and books) are still separate and uncollected well (Widjaja et al. 2014). These data are essential for managing and conserving fishery resources (Menning et al. 2018). Collecting data and information to assess the biodiversity of freshwater fish has been carried out using conventional methods, which are the direct collection and morphological identification with the help of taxonomists and the existing literature (Hogg et al. 2018). However, there are big challenges related to using a morphological approach, which is the lack of taxonomists; the identification can be subjective; the lack of literature; the identification doubts of cryptic species, and it takes time, effort, and great expense to estimate all communities (Bean et al. 2017; Ellingsen et al. 2017). Therefore, a more reliable approach, environmental DNA (eDNA) metabarcoding, is needed to collect biodiversity data and information (Ficetola et al. 2008).

The eDNA metabarcoding is a promising approach using DNA barcodes as identification references combined with High Throughput Sequencing (HTS) namely Next-Generation Sequencing (NGS) (Shokralla et al. 2012). The eDNA metabarcoding approach is non-invasive because it only takes DNA from living or dead organisms in the environment from water, soil, or sediment (Ficetola et al. 2008). The eDNA metabarcoding has been used in various monitoring studies to assess the community diversity of multiple taxa groups (Aylagas et al. 2014; Pawlowski et al. 2014; Thomsen and Willerslev 2015; Valentini et al. 2016; Yamamoto et al. 2017; Andriyono et al. 2019). Despite being a promising approach, there are some limitations in using eDNA metabarcoding, including incomplete DNA barcode databases and inappropriate primer pairs. Unregistered DNA barcodes in the databases will impact the limitation and occurrence of species identification errors (Hebert et al. 2016).

Freshwater ecosystems are currently facing a lot of anthropogenic pressure from various community activities (Carew et al. 2013), one of which is Singkarak Lake, the largest lake in West Sumatra. The Singkarak Lake, used for various activities, has experienced a decline in fish biodiversity, based on survey data using conventional methods from 1913 to 2011 (Weber and de Beaufort 1913, 1916; Syandri 2008; Roesma 2011). The results indicate differences in the type of species found from time to time. The lack of studies and unavailable species database

information causes the absence of factual data on freshwater fish biodiversity in Singkarak Lake. Therefore, besides surveying using the conventional method with a morphological approach (Roesma 2011), molecular studies were also conducted continuously to provide fish DNA barcodes in West Sumatra, including Singkarak Lake, which was registered to the Barcode of Life Data System (Bold System) (Roesma 2011; Roesma et al. 2018, 2019, 2020, 2022).

eDNA metabarcoding helps overcome conventional methods' limitations in assessing fish biodiversity in Singkarak Lake. Therefore, an eDNA metabarcoding study was conducted for fish biodiversity monitoring in Singkarak Lake. Some of the fish DNA barcodes from Singkarak Lake have been registered to the Bold System (Roesma et al. 2018, 2019, 2020) and are able used as references for species identification. This study is expected to give information on the connection between the availability of DNA barcodes registered in the database and appropriate primer pair, with the success of species detection using eDNA metabarcoding. Furthermore, this study also provides recommendations for future eDNA metabarcoding studies for various other taxa.

MATERIALS AND METHODS

Study area

The Singkarak Lake is the largest lake in West Sumatra, Indonesia with an area of 107.8 km² and approximately 21 km long. The natural outlet is the Batang Ombilin River which flows to the Strait of Malacca and through a tunnel to Batang Anai to drive the Singkarak hydropower generator in Lubuk Alung, Padang Pariaman district. Water samples taken for research have obtained permission from the local community and fishermen. However, no approval was needed from the local government because no living organisms were sampled, only taking the eDNA in water samples. Sample collections were carried out on May 2022 in Singkarak Lake. The sites of sample collections based on the inlet of Singkarak Lake are; i) Paninggahan, ii) Malalo, iii) Sumpur, iv) Batu Taba, and v) Sumani (Figure 1). At each site, two liters of water samples were collected.

Samples water collection

Water samples can be collected on the water's surface from the edge site, but sampling using boats in the center of the aquatic sites will increase detection (DNA released by species can move quickly) (Goldberg and Strickler 2017). The water samples were collected using a boat on the lake's surface about 200-400 meters from the lake's edge. Two-liter water samples were taken at each sampling site (five sites) (Andruszkiewicz et al. 2017). Samples were collected using bottles soaked in the bleach solution (50%) and washed with running water for future sterile processes with an autoclave (Laramie et al. 2015; Goldberg and Strickler 2017). Water samples were taken using gloves to avoid contamination (Laramie et al. 2015; Carim et al. 2016). The bottle sample was stored in a coolbox and filtered in the laboratory. The bottles were coded with the site's name using an ethanol marker, stored in a refrigerator (24 hours), and away from light (Carim et al. 2016; Goldberg and Strickler 2017).

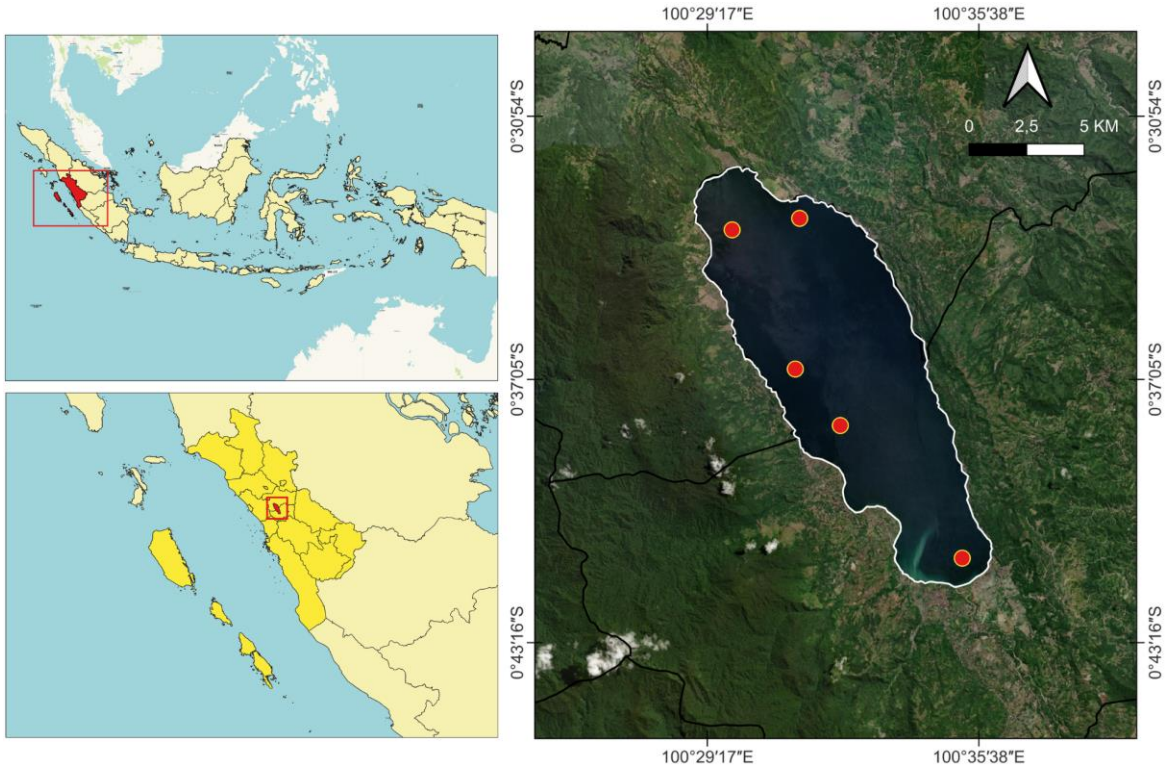

Figure 1. Map of sampling collection in Singkarak Lake, West Sumatra, Indonesia

DNA isolation, DNA amplification, and sequencing

The eDNA sample was isolated from the membrane filter using the DNA isolation kit (gSYNC™ Extraction Kit Geneaid, GS300). The filter membrane was isolated following the protocol for Solid Tissue extraction. First, the filter membrane was crushed into smaller pieces in the microtube. The membrane filter pieces were taken from the tube and air-dried to remove the ethanol. The membrane filter pieces were put into the new microtube. Then, Proteinase K and lysis buffer were added and incubated (60°C) to help the cell membrane lysis process. After incubation, RNase A added to the mix, shaken vigorously, and incubated at room temperature for 5 minutes. Next, the ethanol-absolute was put into the mix to help the DNA binding process and shaken vigorously for 10 seconds. The DNA is bound and separated from the protein and other impurities after the mix is transferred to the GS Column and centrifuged. After centrifuge, the mix was separated into pellet (bound in the GS column) and supernatant, and then discard the supernatant. The Wash buffer was added to the mix, and centrifugation was conducted to wash the DNA from impurities, thus obtaining pure DNA. Finally, the elution buffer (35µL) was added as a solvent for the DNA material (gSYNC™ Extraction Kit Geneaid, GS300). The results of DNA isolation were checked for purity and concentration using a NanoDrop spectrophotometer (IMPLEN, CA, USA) and Qubit dsDNA Assay Kit, Qubit 2.0 Fluorometer (Life Technologies, CA, USA). The DNA obtained was the total genome DNA (gDNA), and the DNA isolate was stored at -20°C (gSYNC™ Extraction Kit Geneaid, GS300).

DNA amplification was carried out in the Cytochrome Oxidase I (COI) gene region because it has been used as a DNA barcode. DNA barcodes (COI gene) of fish in West Sumatra have been registered in the GenBank (BOLD System). Therefore, it can be used as a reference for eDNA metabarcoding identification. DNA amplification was conducted using the COI universal primer pair FISH F1 (5'TCAACCAACCACAAAGACATTGGA C3') forward and FISH R1 (5'TAGACT TCTGGGTGGCCAAAGAATCA3') reverse (Ward et al. 2005). The polymerase chain reaction (PCR) process used the MyTaq HS Red Mix PCR kit, 2X (Bioline, BIO-25048). The PCR reaction was carried out using 11 µL MyTaq HS Red Mix (Bioline), 3 µL DNA isolates (35ng/µL), 9 µL nucleus-free water (Invitrogen), and 1 µL forward and reverse primers (10 µM) for 25 µL total volumes. The PCR products were visualized using electrophoresis on a 1% agarose gel to confirm the presence of the target DNA band.

The PCR products were prepared for library preparation using Oxford Nanopore Technologies Prep Kit. Moreover, the sequencing was performed with the GRIDION machine, Oxford Nanopore Technologies, UK. Oxford Nanopore Technologies has developed a new generation of DNA/RNA sequencing. Oxford Nanopore has an advantage in sequencing reads compared to the previous Next-generation Sequencing because it offers real-time analysis with the principle of the pore-based flow cell and membrane technology. Oxford Nanopore technologies can sequence any fragment length from the short to the long fragment (Oxford Nanopore Technologies 2008). The GRIDION sequencing can run five flow cells and generate upwards of 10 Gb of data on each flow cell at a time. Sequencing starts with the attachment of rapid 1D sequencing adapters and loading them into the GRIDION machine (Oxford Nanopore Technologies 2008). The sequencing machine contains a series of tiny holes with nanopores embedded in an electro-resistant membrane. Each nanopore is connected to channels, and a sensor chip corresponds with its electrode, which can measure the electric current that flows in the nanopore. The molecule through the nanopore will disrupt the current and result in a characteristic' 'squiggle'. It was translated using basecalling algorithms to determine the DNA/RNA sequence reads in real-time (Oxford Nanopore Technologies 2008). The process of DNA isolation, DNA amplification, and DNA sequencing was carried out at the Genetika Science Indonesia Laboratory.

Data analysis

Sequencing data were obtained using MinKNOW software version 20.0.9 (https://nanoporetech.com/about-us/news/introducing-new-minknow-app) Oxford Nanopore technologies. MinKNOW is software that drives nanopore sequencing. The software was used for several core tasks; data acquisition, real-time analysis, basecalling, and data streaming, while providing device control for selecting run parameters, sample identification, and ensuring the platform chemistry is worked (Oxford Nanopore Technologies 2008). Basecalling from sequencing data was carried out using Guppy version 4.0.11 with high accuracy mode (Wick et al. 2019). The FASTQ data was generated after basecalling, which contains sequence quality score information. The quality score in Phred Q is related to the probability of basecalling error. FASTQ files were filtered using Filtlong software (https://github.com/rrwick/Filtlong) to obtain good-quality sequences and visualized using NanoPlot (de Coster et al. 2018). The filtered data are classified using a Centrifuge classifier based on sequence similarities (Kim et al. 2016). The consensus sequences and variant calls were assembled using Medaka v1.5.0 software (https://github.com/nanoporetech/medaka). The consensus was performed by assembling the fragment sequence reads. Assembly was carried out using Flye v 2.8.1 software (Kolmogorov et al. 2019). The result of the assembly was aligned with the genome references in the NCBI database using BLAST (https://blast.ncbi.nlm.ni.gov) and the BOLD System. After sequences aligned with the reference sequences, their annotation was assessed to determine the Operational Taxonomic Unit (OTU) placement on each sequence read.

RESULTS AND DISCUSSION

Raw data sequencing

The NGS sequencing results are data in FASTQ format, a text-based format for storing nucleotide sequences and assessing sequence quality (Frampton and Houlston 2012). The FASTQ has become the standard for storing the output

of HTS. The FASTQ contains sequence quality score information in Phred Q, which is logarithmically related to the basecalling error probabilities P (Meacham et al. 2011). P is the probability of incorrect basecalling (Meacham et al. 2011). The higher the score, the better the base call (Meacham et al. 2011). In FASTQ files, raw sequence quality control was checked by filtering and trimming sequences based on Q quality to obtain the sequence's good quality. Sequences with low Q scores will be trimmed and not included in the analysis (Frampton and Houlston 2012). The filter sequences were sorted and aligned, then grouped into OTU and species annotation.

The quality of the FASTQ visualized using the Nanoplot on the sequencing results of eDNA from Singkarak Lake is shown in Table 1. The total bases are the total bases number for all sequences read. The number of reads is the total number of sequences read. The mean read length is the mean value of the total base length of all sequence reads. Mean read quality is the mean quality value for all sequence reads (Q score). The median read length is the median value of the total read length of all sequence reads. Median read quality is the median value of all sequence quality reads (Q score). Meanwhile, OTU is defined as an operational unit to classify an individual into species groups based on sequences similarities (Blaxter et al. 2005; Porter and Hajibabaei 2018). The results obtained as many as 137,013-166,199 total reads at the five sampling locations.

All sequence reads are grouped in 3,152 OTUs, including OTU from the bacteria, fungi, and plantae groups. As many as 500 OTU from 3,152 OTUs cannot be grouped into clear taxa. Bacteria, plantae, and fungi groups are the most sequence reads. Bacteria groups have the highest number of OTU (800 OTU). However, bacteria, fungi, and plantae groups were not included in the further analysis. The total number of OTU detected at each sampling point is shown in Table 1. Among all identified OTU, most of the same OTU (more than 60%) were found in three of the five sampling locations.

Analysis of the barcode species in the reference libraries

The group included for further analysis was the phyla in the Animalia kingdom. Results of BLAST with the DNA barcode database at GenBank obtained 20 phyla detected is shown in Table 2. Table 2 shows the list of phyla detected, the mean number of species detected in five sites, the mean number of species detected without barcodes in five sites, and the mean percentage of barcode gaps for the five sampling sites. The barcode gap is present when organism groups in phyla are detected but cannot be assigned to the species level. So, they considered the barcode gaps (the group doesn't yet have its barcode DNA among the other species detected). The mean percentage of the barcode gap resulted from the divided mean number of species detected without barcode (only assigned to genus/family level) with the mean number of species detected. Overall, in the 20 phyla detected, 82.12% of species had DNA barcodes registered at GenBank, and an average of 17.88% of species without DNA barcodes. Phyla with the highest mean number of species and having DNA barcodes registered in the database are chordata (375) and arthropods (74), with low barcode gaps percentage values of 5.6% and 28.4%, respectively. Mollusca, Nematoda, and Platyhelminthes were detected with a mean of 19, 19, and 15 species, respectively. In comparison, the other phyla had a similar distribution of the mean number of species detected at each location ranging from one to eight species.

The number of OTU detected in the order of the Chordata at each sampling site is shown in Figure 2. A long-distance separate each sampling location from one other. The figure compared the number of OTU detected from the order at each sampling site. The highest OTU in each order was found at Sumpur and Batu Taba sites. The lowest number of OTU in each order was found at the Malalo site. Among the order in the Chordata, Actinopterygii has the highest number of OTU, with a mean number of 185 OTUs from five sampling sites, followed by Aves and Mammalia. The Ascidiacea, Leptocardii, and Reptilia were the lowest orders detected at five sampling sites.

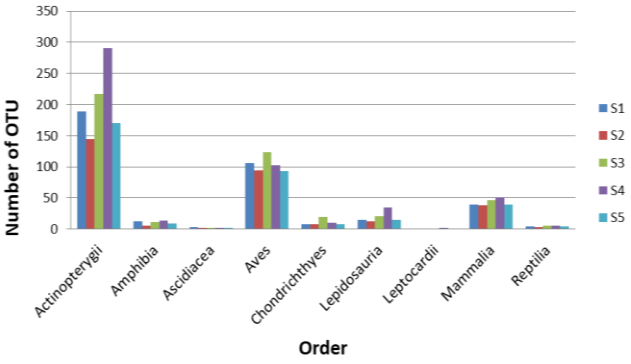

Figure 2. The total number of OTU was detected in the order of the Chordata using the eDNA method in five sites. S1: Paninggahan, S2: Malalo, S3: Sumpur, S4: Batu Taba, S5: Suman

Table 1. The quality of FASTQ files was assessed by NanoPlot

FASTQ Data	S1	S2	S3	S4	S5
Mean read length	815.2	754.6	803.4	783.2	756.6
Mean read quality	11.3	12.6	11.5	11.7	12.2
Number of reads	160.004	137.013	165.234	166.199	167.830
Median read length	764	754	759	756	755
Median read quality	11.2	12.8	11.4	11.8	12.4
Total bases	130,438,811	103,393,939	132,750,479	130,166,384	126,979,530
Total OTUs	1,131	1,341	1,465	1,759	1,579

Note: S1: Paninggahan site, S2: Malalo site, S3: Sumpur site, S4: Batu Taba site, S5: Sumani site

Table 2. List of phyla, the mean number of species detected, the mean number of species without barcode, and the mean percentage of the Barcode Gap in five sites

Phyla	The mean number of species detected in five sites	The mean number of species detected without a barcode in five sites	Mean percentage of the Barcode Gap (%)
Annelida	7	4	57.1
Arthropoda	74	21	28.4
Bryozoa	1	0	0
Chaetognatha	1	0	0
Chordata	375	21	5.6
Cnidaria	3	1	33.3
Echinodermata	8	1	12.5
Euglenozoa	3	0	0
Evosea	1	0	0
Gastrotricha	2	1	50
Hemichordata	1	0	0
Heterolobosea	1	0	0
Mollusca	19	2	10.5
Nematoda	19	1	5.2
Nemertea	2	1	50
Oomycota	4	1	25
Platyhelminthes	15	2	13.3
Porifera	3	1	33.3
Priapulida	1	0	0
Rotifera	3	1	33.3

Comparison of Actinopterygii identified using eDNA metabarcoding and those known to occur in Singkarak Lake

Actinopterygii, as the highest most detected order, was the target of this study. Previously, the freshwater fish diversity data in Singkarak Lake has been reported using conventional methods (morphological approach). Fish diversity data in the previous studies are shown in Table 3. A total of 30 fish species were found in the three surveys, among them 28 species in the first survey in 1913, 19 species in the second survey, and 16 species fish in the last survey in 2011 (Weber and de Beaufort 1913, 1916; Syandri 2008; Roesma 2011). However, the types of species found in each survey were sometimes different. So then, this study (eDNA) detected as many as 343 species, consisting of 152 freshwater fishes and 191 marine fishes. All marine fish were not included in the further analysis. All detected marine fishes are impossible to find living in freshwaters such as Singkarak Lake. Therefore, marine fish detected using the eDNA method is considered a false positive to avoid misinterpretation. False positive is the detection of DNA from non-living organisms in the system. Among 152 freshwater fish detected, only ten species (33.3%) were previously found in Singkarak Lake. Meanwhile, 140 other fishes detected have never been reported in Singkarak Lake, which is native species, and living on other islands in Indonesia and other countries. So, the fish detected in Singkarak Lake using the eDNA method can be considered a false positive. The results from the first to the end survey (eDNA) showed a decrease in the species number found over time. When compared with the types of species found in the last survey (2011), only six species were detected using eDNA metabarcoding.

Among the ten species detected by the eDNA method, not all species were found at all sampling sites. Based on the number of species found per site, Paninggahan has the highest number of species detected (eight species), and Sumani is the site with the least number of species detected (two species). *Osteochilus vittatus* and *Cyprinus carpio* were the species found at all sampling sites. *Rasbora jacobsoni* was found at four sites, and *Clarias batrachus, Oreochromis niloticus,* and *Mystacoleucus padangensis* were found at the same two sites. In contrast, other species are only found in one different site. *Rasbora argyrotaenia* and *R. jacobsoni*, known locally as Bada, are among the species reported only last time they were found during the first survey. The discovery of these species using the eDNA method proves that this method can detect the presence of species that are difficult to find using conventional methods. *Barbonymus schwanefeldii* and *M. padangensis*, native species of Singkarak Lake, which have continued to experience a decline in population numbers, are only found at the Paninggahan and Sumpur sites.

As many as 20 other species reported in Singkarak Lake using conventional methods were not detected by the eDNA method. As many as 17 undetected species have registered DNA barcodes. As many as three species (*Barbonymus belinka*, *Nemacheilus olivaceus,* and *Gobiopterus* cf. *brachypterus*) do not have registered DNA barcodes in the GenBank. Species *B. belinka* and *Gobiopterus* cf. *brachypterus* are native to Singkarak Lake (Sumatra) that have substantial economic value and continue to experience a decline in population numbers. However, the DNA barcode data of this species has yet to be registered in the database reference. In contrast, the availability of the target species' barcode data is crucial for the success of species detection using the eDNA method. Among the species that already have DNA barcodes, three species (*Hampala bimaculata, Hemibagrus planiceps,* and *Hemibagrus velox*) only have one DNA barcode sequence each, registered in the GenBank database. Species

undetected that inhabit or are present in an aquatic system are said to be a false negative. A false negative is the organisms or target DNA present and living in the system that was undetected in the survey.

A total of 142 other freshwater fish species detected were species that had never been previously reported in Singkarak Lake (data not shown). However, two species among them have been reported present in other lakes in West Sumatra, namely *Barbonymus gonionotus* and *Rasbora sumatrana*. Both species were found in Diatas Lake, geographically related to Singkarak Lake as sympatric populations (Roesma et al. 2018). Therefore, this species can be a new detection of species found in Singkarak Lake. On the other hand, other species detected were only found on other islands in Indonesia and other countries that have never been reported in Singkarak Lake. Therefore, detecting a species that has never inhabited a system was said to be a false positive.

Discussion

In this study, we collected water samples from five sampling sites in Singkarak Lake to quickly monitor fish biodiversity using the eDNA method with the NGS technique. NGS sequencing detected 3,152 OTUs, of which 800 OTUs were non-target groups, namely bacteria, plantae, and fungi. Non-target groups were not included in the analysis. Among the analyzed OTUs, 20 phyla were detected in the kingdom Animalia. These results indicate that detection using eDNA does not only detect fishes but other vertebrate and invertebrate groups, especially chordates as shown in Table 2 and Figure 2. Valdez-Moreno et al. (2019) reported monitoring fish groups using eDNA-detected OTU groups from other vertebrates, such as amphibians, reptiles, birds, and mammals. Monitoring fish and metazoan with eDNA metabarcoding (Schwentner et al. 2021) also detected 942 OTU (72.8%) non-metazoan groups.

Table 3. Species list in Singkarak Lake was reported in the previous studies using conventional methods and eDNA metabarcoding

Order	Family	Genus	Species	Sequence recovered from eDNA per sites					Barcode species in GenBank
				S1	S2	S3	S4	S5	
Cypriniformes	Balitoridae	Homaloptera	*Homaloptera gymnogaster***	0	0	0	0	0	✓
	Cyprinidae	Barbonymus	*Barbonymus belinka**^*	0	0	0	0	0	-
			*Barbonymus schwanefeldii**^ᵛ*	1	0	0	0	0	✓
		Cyclocheilichthys	*Cyclocheilichthys apogon**^ᵛ*	0	0	0	0	0	✓
			*Cyclocheilichthys armatus**^ᵛ*	0	0	0	0	0	✓
		Cyprinus	*Cyprinus carpio**	1	1	1	1	1	✓
		Hampala	*Hampala bimaculata**ᵛ*	0	0	0	0	0	✓
			*Hampala macrolepidota**^ᵛ*	0	0	0	0	0	✓
		Mystacoleucus	*Mystacoleucus padangensis**^ᵛ*	1	0	1	0	0	✓
		Osteochilus	*Osteochilus vittatus/ Osteochilus hasseltii**^ᵛ*	1	1	1	1	1	✓
		Tor	*Tor tambroides**^*	0	0	0	0	0	✓
	Danioninae	Rasbora	*Rasbora argyrotaenia**	0	0	1	0	0	✓
			*Rasbora jacobsoni**	1	1	1	1	0	✓
			*Rasbora spilotaenia**	0	0	0	0	0	✓
	Nemacheilidae	Nemacheilus	*Nemacheilus olivaceus**	0	0	0	0	0	-
	Gobiidae	Gobiopterus	*Gobiopterus* cf.*brachypterus**^*	0	0	0	0	0	-
Perciformes	Anabantidae	Anabas	*Anabas testudineus**^ᵛ*	0	0	0	0	0	✓
	Channidae	Channa	*Channa lucius**^ᵛ*	0	0	0	0	0	✓
			Channa striata^ᵛ*	0	0	0	1	0	✓
	Cichlidae	Oreochromis	*Oreochromis mossambicus**	0	0	0	0	0	✓
			*Oreochromis niloticus**^*	1	0	1	0	0	✓
	Osphronemidae	Osphronemus	*Osphronemus goramy**^*	0	0	0	0	0	✓
		Trichopodus	*Trichopodus trichopterus**^*	0	0	0	0	0	✓
Siluriformes	Bagridae	Hemibagrus	*Hemibagrus planiceps**^*	0	0	0	0	0	✓
			*Hemibagrus velox*ᵛ*	0	0	0	0	0	✓
	Claridae	Clarias	*Clarias batrachus**^ᵛ*	1	0	1	0	0	✓
	Sisoridae	Glyptothorax	*Glyptothorax platypogon**ᵛ*	0	0	0	0	0	✓
Synbranchiformes	Mastacembelidae	Mastacembelus	*Mastacembelus erythrotaenia**ᵛ*	1	0	0	0	0	✓
			*Mastacembelus unicolor**^ᵛ*	0	0	0	0	0	✓
Tetraodontiformes	Tetraodontidae	Tetraodon	*Tetraodon leiurus**^ᵛ*	0	0	0	0	0	✓

Note: *****First Survey (Weber and de Beaufort 1913; 1916), ^Second Survey (Syandri 2008), *ᵛ*Third Survey (Roesma 2011). S1: Paninggahan site, S2: Malalo site, S3: Sumpur site, S4: Batu Taba site, S5: Sumani site

Overall, eDNA metabarcoding detected 152 freshwater fish from the five sampling sites in Singkarak Lake. Among 152 freshwater fish, only ten species (6.57%) were reported from previous surveys using conventional methods (Weber and de Beaufort 1913, 1916; Syandri 2008; Roesma 2011). Based on the species number (30 species) found in the previous surveys, only ten species (33.3%) were re-detected using the eDNA metabarcoding (Weber and de Beaufort 1913, 1916; Syandri 2008; Roesma 2011). Andruszkiewicz et al. (2017) reported that among 72 fish species detected in Monterey Bay using eDNA metabarcoding, only 52 species were known to have been reported in Monterey Bay. The detection of other species that have never been reported inhabit the waters is said to be a false positive, which can occur in studies using the eDNA (Darling and Mahon 2011; Goldberg et al. 2015; Andruszkiewicz et al. 2017; Valdez-Moreno et al. 2019; Schwentner et al. 2021).

Among the species re-detected using the eDNA, most species were found at the Paninggahan site. Based on the water quality around the sampling site, Paninggahan is a site that has good water quality because no floating net cages, or bombs were found to catch fish. Fishing may only be conducted using nets to maintain water quality and fish populations. The water quality in Paninggahan is still good, as evidenced by the detection of *M. padangensis*, *R. jacobsoni*, and *B. schwanefeldii*. *B. schwanefeldii* is native fish in Singkarak Lake only detected at the Paninggahan site. The site with the lowest number of species found at Sumani. Sumani is a site that has poor water quality than other sites because this site still uses Floating Net Cages to catch fish. The damage to the water lake due to fishing gear is proven by detecting only two fish species at the Sumani site: *C. carpio* and *O. vittatus.* This species is a fish cultivated in Floating Net Cages and can live in poor water quality.

Compared to all the species reported in Singkarak Lake, 20 species were not detected. While compared to the last survey in 2011, ten species were not found using eDNA. These results indicate a lower percentage of fish species detected again using eDNA than surveys using conventional methods. Schwentner et al. (2021) reported a low percentage of fish species known to inhabit the Elbe estuary, which can be detected using eDNA metabarcoding. The low percentage of species detection is due to several factors (Darling and Mahon 2011; Goldberg et al. 2015; Roesma et al. 2021a; 2021b): the unavailability of target species DNA barcodes, inappropriate primers pair selection, the low quality of the target species' DNA, and the absence of target species DNA in the collected water samples.

Among the undetected species using eDNA metabarcoding, three native species in Singkarak Lake (*B. belinka*, *N. olivaceus*, and *Gobiopterus* cf. *brachypterus*) did not have DNA barcodes registered in GenBank. *Gobiopterus* cf. *brachypterus* (Rinuak) is one of the main fish caught by fishermen and continues to decline in population. While taxonomic status of *Gobiopterus* cf. *brachypterus* (Rinuak) in Gobiidae is still unclear. Roesma et al. (2020) reported Rinuak fish as different species with genus members of the only transparent fish in Gobiidae (*Gobiopterus*). Even though it has a similar morphology to *G. brachypterus*, it has a high genetic distance (>20%) between them. Thus, Rinuak fish is temporarily designated as *Gobiopterus* cf. *brachypterus* (Roesma et al. 2020). Further study is still being conducted to determine the correct scientific name, so that the Rinuak DNA barcode cannot be published to NCBI and the Bold System as public data. Based on interviews with local fishermen, Rinuak is still present at all sites in Singkarak Lake. Therefore, the undetected Rinuak fish in this study can occur due to GenBank's unavailability of DNA barcodes. *B. belinka* and *N. olivaceus* are fish with limited distribution in Sumatra, especially in Singkarak Lake, which has limited biological information included did not have DNA barcodes registered in GenBank.

Overall, based on the results obtained and references from the previous studies, the false positives and false negatives detection are caused by several main factors: incomplete/unavailability of freshwater fish DNA barcodes in Indonesia registered in the database repository, inappropriate primer pair selection, low DNA quality, and the absence of target species DNA in collected water samples. Andruszkiewicz et al. (2017) reported that false detection occurred because the OTUs detected did not have representative sequences matches with the GenBank. More than half of the total OTUs are not matching with the GenBank database can be explained as part of the presence of barcode gaps in the repository database. The eDNA metabarcoding studies cannot provide information about species composition without a reliable database reference (Valdez-Moreno et al. 2019), such as the available DNA barcode databases for Mexican freshwater fishes (Schmitter-Soto 1999), which cover 93% of species known. Specchia et al. (2020) stated that the success of species identification in DNA metabarcoding studies depends on the completeness of the reference libraries. The availability of complete DNA sequence references in the DNA barcode database is critical because able to be used as a reference in determining the species taxa obtained. The use of eDNA will result in the identification of species accurately and quickly compared to using conventional methods.

The use of Fish FI and RI universal primers is thought to cause the detection of non-target other taxa groups. Primer is short nucleotide sequences needed as attachment points for DNA polymerase enzymes which are helpful in the initiation process of DNA sequence elongation in the PCR process. Meanwhile, the universal primer is a general primer capable of binding various DNA templates. Therefore, the universal primer can detect various species in certain taxa groups in one sequencing process. The Fish FI and RI universal primers were chosen because the primer has been widely used in general to identify fish groups (Ward et al. 2005). In addition, previous studies have used primer Fish FI and RI to create DNA barcodes of fish in Singkarak Lake (Roesma et al. 2018, 2019, 2020, 2022) and have been registered in Bold System. Furthermore, fish and metazoan monitoring (Schwentner et al. 2021) with eDNA metabarcoding using the COI universal primer also detects non-fish and non-metazoan

groups. Meanwhile, marine vertebrate monitoring (Andruszkiewicz et al. 2017) using the specific primer sets resulted in the detection of more fish groups compared to the non-target taxa.

The success of eDNA metabarcoding for detecting the target groups was closely related to the appropriate primer pair selection. Although the other taxa were also detected, the Fish FI and RI primers were adequate for fish monitoring in Singkarak Lake. Schwentner et al. (2021) stated that the main advantage of using universal primers able to detect all taxa in a primer at one reaction. However, Specchia et al. (2020) stated that the design combination of primers to cover specific groups is needed to overcome the limitations of species detection using eDNA metabarcoding. Therefore, spesific primers for freshwater fish groups have to be designed for future DNA metabarcoding work to obtain the optimal result.

In addition to DNA barcodes and the selection of primers, the presence of eDNA and eDNA quality in the collection of water samples also affects the success of species detection. The absence of DNA of target species in the water samples led to false negatives. Likewise, the low DNA quality of the target species, so cannot be continued to the PCR and sequencing process. Bohmann et al. (2014), Goldberg et al. (2015), and Shelton et al. (2016) reported that false detection might result from eDNA from an organism not being captured in the water sample and low DNA quality. The results study showed the eDNA method could be applied to monitor species quickly, lower cost, and without killing/disturbing organisms. Even so, the eDNA method has some disadvantages compared to conventional methods, but these can still be overcome by considering various factors that influence the success of eDNA detection.

In conclusion, fish biodiversity monitoring using eDNA metabarcoding detected 152 fish species. The eDNA metabarcoding detected ten species that have been previously reported in Singkarak Lake using conventional methods. A total of 20 other species known in Singkarak Lake are undetected using eDNA metabarcoding. The low percentage of species detection was due to several factors; incomplete/unavailability of freshwater fish DNA barcodes in Indonesia registered in the database repository, inappropriate primer pair selection, the low quality of the target species' DNA, and the absence of target species DNA in the collected water samples. This study shows that freshwater fish biodiversity monitoring could be conducted using eDNA metabarcoding, which considers the availability of DNA barcodes and the use of spesific primers for freshwater fish groups. Thus, miss identification can be avoided in determining taxa. Therefore, the provision of DNA barcodes registered to the reference database is the main work that needs attention for the successful future application of eDNA metabarcoding in biodiversity monitoring.

ACKNOWLEDGEMENTS

The research was supported by Andalas University under the PDU-KRP1GB Unand scheme (No.62/UN16.17/ PT.06/2022). We extend our thanks to the management for the funding support. We also thank the Biology Department of Andalas University for a field and Laboratory work permit. We also thank our students who helped us with sample collection and laboratory work in the Genetic and Biomolecular Laboratory, Biology Department, Faculty of Mathematics and Natural Sciences, Andalas University.

REFERENCES

Andriyono S, Alam MDJ, Kim H. 2019. Environmental DNA (eDNA) metabarcoding: Diversity study around the Pondok Dadap fish landing station, Malang, Indonesia. Biodiversitas 20 (12): 3772-3781. DOI: 10.13057/biodiv/d201241.

Andruszkiewicz EA, Starks HA, Chavez FP, Sassoubre LM, Block BA, Boehm AB. 2017. Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. Plos One 12 (4): e0176343. DOI: 10.1371/journal.pone.0176343.

Aylagas E, Borja A, Rodriguez-Ezpeleta N. 2014. Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (GAMBI). PLoS One 9: e90529. DOI: 10.1371/journal.pone.0090529.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005. Defining operational taxomic units using DNA barcode data. Philos Trans R Soc Lond B Biol Sci 360: 1935-1943. DOI: 10.1098/rstb.2005.1725.

Bean TP, Greenwood N, Beckett R, Biermann L, Bignell JP, Brant JL, Copp GH, Devlin MJ, Dye S, Feist SW, Fernand L, Foden D, Hyder K, Jenkins CM, van der Kooij J, et al. 2017. A review of the tools used for marine monitoring in the UK: Combining historic and contemporary methods with modeling and socioeconomics to fulfill legislative needs and scientific ambitions. Front Mar Sci 4: 263. DOI: 10.3389/fmars. 2017.00263.

Bohmann K, Evans A, Gilbert MTP, Carvalho R, Creer S, Knapp M, Yu DW, Bruyn M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol Evol 29: 358-367. DOI: 10.1016/j.tree.2014.04.003.

Carew ME, Pettigrove, VJ, Metzeling L, Hoffmann AA. 2013. Environmental monitoring using Next-generation sequencing: Rapid identification of macroinvertebrate bioindicator species. Front Zool 10: 45. DOI: 10.1186/1742-9994-10-45.

Carim KJ, McKelvey KS, Young MK, Wilcox TM, Schwartz MK. 2016. A Protocol for Collecting Environmental DNA Samples from Streams. Gen Tech Rep RMRS-GTR-355. Department of Agriculture, Forest Service, Rocky Mountain Research Station. Fort Collins, Colorado, United states of America. DOI: 10.2737/RMRS-GTR-355.

Darling JA, Mahon AR. 2011. From molecules to management: Adopting DNA based methods for monitoring biological invasions in aquatic environments. Environ Res 111: 978-988. DOI: 10.1016/j.envres.2011.02.001.

de Coster W, D'Hert S, Schultz DT, Cruts M, van Broeckhoven C. 2018. NanoPack: Visualizing and processing long-read sequencing data. Bioinformatics 34: 2666-2669. DOI: 10.1093/bioinformatics/bty149.

Ellingsen KE, Yoccoz NG, Tveraa T, Hewitt JE, Thrush SF. 2017. Longterm environmental monitoring for assessment of change: Measurement inconsistencies over time and potential solutions. Environ Monit Assess 189: 595. DOI: 10.1007/s10661-017-6317-4.

Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species detection using environmental DNA from water samples. Biol Lett 4 (4): 423-425. DOI :10.1098/Rsbl. 2008.0118.

Frampton M, Houlston R 2012. Generation of artificial FASTQ files to evaluate the performance of Next-Generation Sequencing pipelines. Plos one 7 (11): e49110. DOI: 10.1371/journal.pone.0049110.

Goldberg CS, Strickler KM, Pilliod DS. 2015. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. Biol Conserv 183: 1-3. DOI: 10.1016/j.biocon.2014.11.040.

Goldberg C, Strickler K. 2017. Environmental DNA Protocol for Freshwater Aquatic Ecosystems version 2.2 Washington State University, Washington.

Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, Pedersen S, Jannetta P, deWaard JR. 2016. Counting animal species with DNA barcodes: Canadian insects. Philos Trans R Soc B 371: 20150333. DOI: 10.1098/rstb.2015.0333.

Hogg ID, Banks SM, Wood SM. 2018. Biodiversity Genomics: Monitoring Restoration Efforts Using DNA Barcoding and Environmental DNA. In: Hamilton DP et al. (eds.) Lake Restoration Handbook. University of Waikato, Hamilton, New Zealand. DOI: 10.1007/978-3-319-93043-5_12.

Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res 26 (12): 1721-1729. DOI: 10.1101/gr.210641.116.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, errorprone reads using repeat graphs. Nat Biotechnol 37: 540-546. DOI: 10.1038/s41587-019-0072-8.

Laramie MB, Pilliod DS, Goldberg CS, Strickler KM. 2015. Environmental DNA Sampling Protocol-Filtering Water to Capture DNA from Aquatic Organisms: U.S. Geological Survey Techniques and Methods. U.S. Geological Survey, Reston, Virginia. DOI: 10.3133/tm2A13.

Meacham F, Boffelli D, Dhahbi J, Martin DI, Singer M, Pachter L. 2011. Identification and correction of systematic error in high-throughput sequence data. BMC Bioinformatics 12: 451. DOI: 10.1186/1471-2105-12-451.

Menning D, Simmons T, Talbot S. 2018. Using redundant primer sets to detect multiple native Alaskan fish species from environmental DNA. Conserv Gen Resour 12: 109-123. DOI: 10.1007/s12686-018-1071-7.

Oxford Nanopore Technologies. 2008. Nanopore Sequencing. https://nanoporetech.com/how-it-works.

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA. 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. Mol Ecol Resour 14 (6): 1129-1140. DOI: 10.1111/1755-0998.12261.

Porter TM, Hajibabaei M. 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Mol Ecol 27 (2): 313-38. DOI: 10.1111/mec.14478.

Roesma DI. 2011. Species Diversity and Genetic Relationship of Cyprinidae in Lakes and the Adjoining Rivers in West Sumatra. Dissertation. Andalas University, Padang. [Indonesian]

Roesma DI, Djong HT, Munir W, Aidil DR. 2018. New record species of Puntius (Pisces: Cyprinidae) from West Sumatra based on Cytochrome Oxidase I gene. Intl J Adv Sci Eng Inf Technol 8 (1): 250-256. DOI: 10.18517/ijaseit.8.1.4170.

Roesma DI, Tjong DH, Karlina W, Aidil DR. 2019. Taxonomy confirmation of *Puntius* cf. *binotatus* from Gunung Tujuh lake, Jambi, Indonesia based on cytochrome oxidase I (COI) gene. Biodiversitas 20 (1): 54-60. DOI: 10.13057/biodiv/d200107.

Roesma DI, Tjong DH, Aidil DR. 2020. Phylogenetic analysis of transparent gobies in three Sumatran lakes, inferred from mitochondrial Cytochrome Oxidase I (COI) gene. Biodiversitas 21 (1): 43-48. DOI: 10.13057/biodiv/d210107.

Roesma DI, Tjong DH, Janra MN, Aidil RA. 2021a. Fish diversity monitoring in Maninjau Lake, West Sumatra using the eDNA with the next generation sequencing (NGS) technique. IOP Conf Ser Earth Environ Sci 819: 012045. DOI: 10.1088/1755-1315/819/1/012045.

Roesma DI, Tjong DH, Janra MN, Aidil RA. 2021b. Freshwater vertebrates monitoring in Maninjau Lake, West Sumatra, Indonesia using environmental DNA. Biodiversitas 22: 2794-2802. DOI: 10.13057/biodiv/d220543.

Roesma DI, Tjong DH, Janra MN, Aidil RA. 2022. DNA barcoding of freshwater fish in Siberut Island, Mentawai Archipelago, Indonesia. Biodiversitas 23: 1795-1806. DOI: 10.13057/biodiv/d230411.

Schmitter-Soto JJ. 1999. Distribution of continental fishes in Northern Quintana Roo, Mexico. Southwest Nat 44 (2): 166-172.

Schwentner M, Zahiri R, Yamamoto S, Husemann M, Kullmann B, Thiel R. 2021. eDNA as a tool for non-invasive monitoring of the fauna of a turbid, well-mixed system, the Elbe estuary in Germany. Plos One 16 (4): e0250452. DOI: 10.1371/journal. pone.0250452.

Shelton AO, O'Donnell JL, Samhouri JF, Lowell N, Williams GD, Kelly RP. 2016. A framework for inferring biological communities from environmental DNA. Ecol Appl 26 (6): 1645-1659. DOI: 10.1890/15-1733.1.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next generation sequencing technologies for environmental DNA Research. Mol Ecol 21: 1794-1805. DOI: 10.1111/J.1365-294x.2012.05538.X.

Specchia V, Tzafesta E, Marini G, Scarcella S, D'Attis S, Pinna M. 2020. Gap analysis for DNA barcode reference libraries for aquatic macroinvertebrate species in the Apulia Region (Southeast of Italy). J Mar Sci Eng 8: 538. DOI: 10.3390/jmse8070538.

Syandri H. 2008. Threats to The Germplasm of Bilih Fish (*Mystacoleucus padangensis* Blkr) and Conservation Efforts in Singkarak Lake. Scientific Oration. University of Bung Hatta, Padang, 08 Juli 2008. [Indonesian]

Thomsen PF, Willerslev E. 2015. Environmental DNA an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183: 4-18. DOI: 10.1016/j.biocon.2014.11.019.

Valdez-Moreno M, Ivanova NV, Elias-Gutie rrez M, Pedersen SL, Bessonov K, Hebert PDN. 2019. Using eDNA to biomonitor the fish community in a tropical oligotrophic lake. Plos one 14: e0215505. DOI: 10.1371/journal.pone.0215505.

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin JM, Peroux, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Moller PR, Willerslev E, Dejean T. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol 25: 929-942. DOI: 10.1111/mec.13428.

von Rintelen K, Arida E, Hauser C. 2017. A review of biodiversity-related issues and challenges in megadiverse Indonesia and other Southeast Asian countries. RIO 3: e20860. DOI: 10.3897/rio.3.e20860.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia's fish species. Phil Trans R Soc B 360: 1847-1857. DOI: 10.1098/rstb.2005.1716.

Weber M, de Beaufort LF. 1913. The Fishes of the Indo-Australian Archipelago Vol. I. EJ Brill, Leiden, Netherland.

Weber M, de Beaufort LF. 1916. The Fishes of the Indo-Australian Archipelago Vol. II. EJ Brill, Leiden, Netherland.

Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore. Genome Biol 20: 129. DOI: 10.1186/s13059-019-1727-y.

Widjaja EA, Rahayuningsih Y, Rahajoe JS, Ubaidillah R, Maryanto I, Walujo EB, Semiadi G. 2014. Kekinian Keanekaragaman Hayati Indonesia 2014. LIPI Press, Jakarta. [Indonesian]

Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, Minamoto T, Miya M. 2017. Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. Sci Rep 7: 40368. DOI: 10.1038/srep40368.