Using DNA Barcode to Improve the Identification of Marine Fish Larvae, Case Study Coastal Water Near Jakarta and Banda Sea, Indonesia

Arif Wibowo1*, Anthony Sisco Pangabean2, Achmad Zamroni3, Asep Priatna3 and Helman Nur Yusuf3

1Research Institute for Inland Fisheries and Fisheries Extensions, Agency for Marine and Fisheries Research, Ministry of Marine Affairs and Fisheries, Jl. H.A Bastari 08 Jakabaring, Palembang - 30763, South Sumatera – Indonesia.
2Inland Fishery Resources Development Management Department, Southeast Asia Fisheries Development Center, Jl. Gubernur H. A. Bastari No. 08 Kel. Silaberanti, Kec. Seberang Ulu I, Palembang, Sumatera Selatan, Indonesia
3Research Institute for Marine Fisheries, Ministry for Marine Affairs and Fisheries, Raya Bogor Km 47st., Nanggewer Mekar, Cibinong, Bogor, Indonesia

ABSTRACT

The sustainability of the exploitation of the Indonesian fishes depends heavily on many of fish’s basic information including both larvae distribution and dispersal. However, the identification of fish larvae and juvenile to species is very limited. In this study, we employed DNA barcoding techniques to identify marine fish larvae to a species level in Jakarta Bay and Banda Sea by comparing the queries with sequences from adult stage as reference library to contribute on biodiversity information on that particular area. The result revealed that in Jakarta Bay, the molecular marker of a 471 bp region of the mitochondrial cytochrome c oxidase I gene (COI) has been successfully found to be species-specific, genetic distance within species (0.0 - 1.30 percent). There are total of 8 families, 5 genera and 5 species from a total 15 successful PCR that could be used to calculate the accuracy of larval fish identification in three taxonomic categories. Meanwhile in Banda Sea, for the adult specimens, after some of PCR experiment, we have successfully amplified 27 individuals, only 8 sequences available. There are a total 326 eggs and larvae which been collected from 19 stations, of the 28 successfully amplified PCR samples, 11 sequences were available for DNA analysis and at least 10 species used Banda Sea and surrounding area as their spawning ground. We prove the ability of COI barcodes to identify species level resolution from query sequences and to classify species from distinct geographical origins and determine of how the data retrieved give important information for proposing plans for conserving and managing of fisheries in the sea waters.

Keywords: Jakarta Bay; Banda Sea; larvae; and DNA barcode

INTRODUCTION

Bay’s fisheries have important spots in the culture and history of local communities, eq they give countless opportunities and incomes in regards to its fishing. However many of those fisheries are experiencing eutrophication, heavy pollutions and overfishing, such as Jakarta Bay (Arifin et al., 2003; Wouthuyzen et al., 2007) and Ambon Bay (Manapa, 2014). These conditions bring negative impact to fisheries especially in the early development stage, the larval. Furthermore, improvement of the crucial resource helps spur job growth and protect the countless livelihoods that depends on the Bay’s fisheries health.

Accurate species identification is vital for scientific research and assessment purposes (Hart & Reynolds, 2002). Specifically knowledge of larval ecology provides fundamental information on the reproductive biology of a species, as well as reproduction sites and preferred times, possible migration routes and population recruitment success rates (Baumgartner et al., 2004; Bialetzki et al., 2005; Reynalte-Tataje et al., 2011; Wibowo et al., 2016). This information is extremely important for monitoring ecology, analysing environmental impacts and developing management and conservation plans (Frantine-Silva et al., 2015), establishing fishing management actions, such as the defeso period (a closed season when fishing a given species is prohibited because of reproduction) (King...
et al., 2003) and helping to protect vulnerable and threatened species (Reynalte-Tataje et al., 2011).

However, it is difficult to identify ichthyoplankton at species level. This is due to morphological keys are sometimes only effective for a particular stage, especially with a larval phase radically different in morphology where many diagnostic characters are not fully developed in the early development stages (Nakatani et al., 1997; Frantime-Silva et al., 2015; Wibowo et al., 2016) and by the enormous diversity of fish in the tropics (Hubert et al., 2015). For Indonesian aquatic organism, the misidentification of a taxon due to several large radiation of morphologically similar species (Kottelat, 2013), lack of taxonomists and available tools for species identification (Hubert et al., 2015). Therefore, the identification of species is a major bottleneck, hampering the necessary monitoring of marine biodiversity (Kochzius, 2009).

In this study, we employed DNA barcoding techniques to identify marine fish larvae to a species level (Larson et al., 2017; Docker et al., 2016). This note provides the first description of preflexiom larvae of marine fishes and an example of our succes in applying molecular genetics. The results of this work enhance our understanding for reproductive activity of fishes in Jakarta Bay and Banda Sea for long-term monitoring of species diversity, abundance and distribution and give us the potential for real-time at sea identification.

**MATERIALS AND METHODS**

**Sample Preparations**

Larvae samples were taken during ichthyoplankton tow survey in the Jakarta Bay, (Table), (Figure 1), employing a Bongo net of 30 cm mouth diameter and 500 µm mesh, in a series of 10-min oblique tows, on 10th of April 2014. Juvenile fish at various developmental stages (from preflexion to postflexion larvae) were collected using a Bongo net of 30 cm mouth diameter and 500 µm mesh, in a series of 10-min oblique tows on water around Banda Sea (Figure 1). Larvae were kept in water, manually sorted after collection and were subsequently individually stored in 96% ethanol.

Additionally, after completing the tow missions, the particles which did not pass through the net was concentrated and trapped within the string plastic bottle in the bottom of net. The concentrated water sample was filtered by 500 lm mesh net to obtain larvae of unknown species. These larvae were collected and then preserved immediately in 96% ethanol onboard. The larvae were sorted from the ichthyoplankton samples in the laboratory under dissection microscopes and separated into different morphotypes referring to their basic morphological characters.

Adult fish collected from local fish market were identified to the species level or alternatively to the genus when systematic knowledge was inadequate for reliable identification of the species following (White et al., 1991) acted as reference species. Approximately 2 cm² piece of fin clip tissue was taken from every dead individual by using scalps and stored in 1.5 mL of 96% ethanol.

**DNA Extraction, Amplification and DNA Sequencing**

Genomic DNA from 130 morphotypes of unknown larval species was isolated using extraction kits from GENEaID, following the manufacturers’ protocols. The cytochrome c oxidase subunit-1 gene (COI) was amplified by polymerase chain reaction (PCR) using the primers Fish-COI-F (5’-ACT TCAAAC TTC CAY AAAGAY aty GG-3) and COI-Fish-R (5’-TAGACT TCT GGG TGG CCR AAR Aay CA-3’) (Ivanova et al., 2007). The primer sequence is known as one of the most commonly used to retrieval of the 52 region of the cytochrome c oxidase 1 (COI) gene.

Polymerase chain reaction (PCR) was processed as 50 µL final working volume. PCRs volume contained 5 µL DNA samples, 16 µL double distillate water, 2 µL of each primer and 25 µL of PCR ready mixture solution KAPA2G fast PCR kit. The following temperature profile was used for the PCR, an initial denaturation phase at 95°C for 10 min, followed by 35 cycles at 94°C for 1 min, 48°C for 1 min and 72°C for 1.5 min and ended with a final extension at 72°C for 7 min. All PCRs were done using a Applied Biosystems and the size of PCR products were visualized in Acrylamide 10%. The PCR products were purified using the The GenepHlow™ Gel/PCR Kit (GENEAID), following the manufacturer’s protocol. A single strand of the purified DNA was automatically sequenced at First Base, Singapore and Macrogen, South Korea. All sequences were stored in GenBank and freely accessible (Accession numbers KY091245-KY091259). Our dataset was appended by COX1 sequences aquired from GenBank database (KP856771.1, EF609414.1, JX972212.1, KT951781.1, EF607317.1, EU871690.1, KT951733.1 and EF609512.1).
DATA ANALYSIS

All sequences were edited with the program sequence navigator (version 7.0.1; Applied Biosystems) (Hall, 1999) and checked manually by in silico method. Blast data search in NCBI database and using Clustal W for multiple sequence alignment (Thompson, 1994) and sequence alignment and sequence analysis used bioedit software program V7.0.41 (Hall, 1999). Phylogenetic reconstruction was done using a distance based method, Neighbor-Joining (NJ), carried out in MEGA5 software (Tamura et al., 2007) with the K2P model of substitution. Support for nodes in NJ analyses was assessed using non-parametric bootstrapping with 100 full heuristic pseudo-replicates.

Species validation of unknown larval morphotypes followed 97% sequence similarity for species identification with the references sequences. This threshold has been often used for specimen identification in different taxonomic groups (e.g. Hebert et al., 2003) and the technique is similar with eq. Hubert et al. (2015). However, in some cases this may lead to clumping of closely related species, as coalescent depths among species will vary due to differences in population size, rate of mutation and time since speciation (Monaghan et al., 2009; Fujita et al., 2012). The identified larvae are used to study spatial patterns of species occurrence.

RESULTS AND DISCUSSIONS

Results

Jakarta Bay

There are twenty-seven samples were unsuccessful PCR for COI even after repeated attempts and could not be classified to species level from a total of 42 individuals, hence it was not included in the COI neighbour-joining (NJ) tree, of the remaining 15 had successful PCR. This left a total of 8 families, 5 genera and 5 species that could be used to calculate the accuracy of larval fish identification in three taxonomic categories. After using a bongo net for sampling larval fish, preservation at dissolved ethanol from water contained larvae are not usually adequate. Certain one-vial fixatives must use certain larval of fish, the impractical happen due to mechanism of sorting process. The process of selecting larvae for its counterpart easier had been done in the lab instead of field.

The Kimura Two Parameters (K2P) divergences of COI sequences among the fish larvae were evaluated and the genetic variability among the species is shown (Fig.1). The molecular marker a 471 bp region of the mitochondrial cytochrome c oxidase I gene (COI) has been successfully found to be species-specific, within species (0.0-1.30%). Sequences of the 6 identifiable larval morphotypes were clustered into monophyletic groups in the NJ tree. Ambassis marianus and Platyccephalus indicus were found to be exhibited higher interspecific variations (1.30%) than the other three species groups. An increase in genetic variation was observed with increase in taxonomic level. COI barcodes are essential to verify the species identification of fish larvae, especially of morphologically similar species, within our results, six specimens can be recognized to the species level (Fig 2). All together, 8 fish species were found in the Jakarta Bay habitat at the larval stage, within this study the largest number of fish species was detected in the sampling point number three on the Jakarta Bay (S3; n = 5), whereas only one species was observed in the sampling point number one. A species, such as the broad-mouthed mullet (P. parmatus) was observed in two sampling points and a few specimens in the family of Gerreidae and Sparidae were found in three sampling points. In contrast, 5 species were detected only in a single site, despite the small sample sizes, the occurrence of fish larvae varied.

Banda Sea

A barcode reference sequence library for 56 species with robust identification of adult specimen from morphology technique. The dataset was used as diagnostic tool to screen queries DNA sequences from fish larva specimens collected in Banda Sea, Indonesia. For the adult specimens, after some of PCR experiment, we have successfully amplified 27 individuals, only 8 sequences available (6 species). The species identification using those sequences are corresponding with the morphology counterpart, Figure 3.
Figure 1. Neighbour-joining (NJ) tree of COI sequences, showing the placement of 15 larvae in a total of 8 families, 5 genera and 5 species, plus 8 taxa retrieved from GenBank, with accession numbers in brackets.

- *Platycephalus indicus* (3.6 mm L)
- *Ambassis marianus* (3.84 mm L)
- *Apogon quadrafasciatus* (4.08 mm L)
- *Eleutheronema tetradactylum* (3.5 mm L)
- *Paramugil parmatus* (3.2 mm L)

Figure 2. *Early life stages* of known *fish* species from marine system of Jakarta Bay, Indonesia. The two specimens (lowest part) were identified to be identical species through barcoding.
A total of 326 eggs and larvae samples were taken from 19 station (with 28 successful PCR reaction, 11 sequences were available for DNA analysis) during ichthyoplankton tow survey in the Banda Sea, Indonesia. The phylogeny tree comprises from adult specimens, larvae and reference sequences from public domain is presented in Figure 4. More than 90% (10 sequences) of the available larvae’s sequences were identified in species level, based on categories that the sequences can be identified in species level if the similarity are above 97%. Even though the generated sequence from 1 specimen considered a good chromatogram result however it cannot classified to the species level due to the incomplete of reference data base. We can only classified the larvae in the genus level, Apogon sp. The effort to identify larvae using morphological performa, Figure 5 was not succesful, although repeated exercised.

Figure 4. Phylogeny tree from NJ Using Kimura 2 Parameter highlight the relatedness among examined samples.

Figure 6. The morpological appearence of the collected fish larvae in banda Sea.
Discussion

The larval were divided into distinct clades according to species using the COI, which implied that COI genes are suitable for species identification of marine larval from Jakarta Bay and Banda Sea. The positioning of organisms on a tree is generally based on their genetic similarity to one another. The tree reflects how much genetic change has occurred and therefore roughly how much time has passed (a solid bar in the tree indicating the scale, since lineages split from one other). Because branch lengths reflect the evolutionary distance between two points on tree (Freeland, 2005). Congeneric species that diverged from a common ancestor relatively new and will be close to each other on the tree. Confamilial genera are further apart on the tree because their common ancestor was more remote, and member of different families are even more widely spaced.

This picture depicts the incomplete of COI database for Indonesian fishes, thus less larval fishes can be identified to the species level. DNA barcoding is a system designed to provide accurate, fast and automatable species identification by using short and standardized gene regions as internal species tag. This DNA sequence, targeted as the 5’ end of the gene coding for the subunit I of the cytochrome c oxidase subunit I (COI), is sufficiently diverse so as to allow the specific identification of a great majority of animal species. Numerous studies have proven the success of this approach in the animal kingdom, and using various sources of tissue samples (Hebert et al., 2003). DNA barcoding relies on the development of DNA barcode reference libraries for known species in order to foster automated assignment of unknown specimens to known species. For automation purposes, both stability of the reference libraries and reproducibility of the molecular identifications should be guaranteed (Hubert et al., 2008).

Success in this study, however, may be largely as a result of the limited species diversity of marine species collected in Jakarta Bay and Banda Sea, where each species represents a different genus. Uncertainties in identification using COI divergence is, therefore, less likely to be encountered, as the mean interspecific divergences deviate far from these cut-off values. Consequently, species identification of marine larvae in Jakarta Bay can be confirmed with a relatively high degree of confidence.

A broader scale of sampling may well challenge the effectiveness of the COI marker because of the potential increase in intraspecific divergence over the entire geographical range of the species concerned, as well as the increase in number of species and the difficulties in achieving complete sampling. Such factors are considered more prominent in tropical regions (e.g. Barber & Boyce, 2006) because of the higher species diversity, which may impair the resolution power of the markers compared with their effectiveness in temperate regions (e.g. Hebert et al., 2003).

The studies, however, have expressed the lack of resolution power, which resulted in ambiguous taxonomic identities. The importance of having a more complete and accurate barcode database for larval fish molecular identification can also be demonstrated in this paper. It should also be noted that a full phylogenetic study of marine fishes in this region is yet to be undertaken; therefore, the genetic diversity of marine species should be evaluated with the inclusion of a complete species inventory, the accuracy of which would be dependent on the completeness of taxonomic sampling (Barber & Bellwood, 2005). Having more complete COI database for all fish species can make fish egg and fish larva identification more successful. The DNA barcoding method was developed in 2003 (Hebert et al., 2003) and began to be applied to fishes in the earliest stage. The first country that established a COI databank for fishes was Australia (Ward et al., 2005). The Fish-BOL campaign was conceived in 2004 (Hebert et al., 2003), and as of October 2012 there have been 136,758 COI sequences belonging to 12,909 fish species that have been deposited in BOLD (Ko et al., 2013).

CONCLUSION

This study establishes the baseline information on marine larval species composition in Jakarta Bay and Banda Sea and demonstrates the validity of DNA barcoding for marine larvae identification on a regional scale. The results of this work enhance our understanding for reproductive activity of fishes for long-term monitoring of species diversity, abundance and distribution and give us the potential for real-time at sea identification. This study not only describes the use of DNA barcoding in the identification of marine larvae, but also provides some insights into distribution of marine larvae in the region. Marine larvae obtained in Jakarta Bay and Banda Sea waters could now be identified to species with a high degree of confidence although there is a lack of potential reference species.

ACKNOWLEDGEMENT

This study was a part of the collaboration research between Research Institute for Marine fisheries and
Research Institute for Inland Fisheries and fisheries extension. The authors thank to Head of Research Institute for Marine Fisheries - Agency for Marine Affairs and Fisheries Research and Human Resources (AMAFRHR) for facilitating and supporting this work.

REFERENCES


