Zoological Studies **58:** 0dd (2019)

Utilization of DNA Barcodes for the Identification of Larval Fishes in Tropical Estuarine Waters (Malacca Straits, Malaysia)

Cecilia Chu[1], Kar Hoe Loh[1], Ching Ching Ng[2], Ai Lin Ooi[3], Yoshinobu Konishi[4], Shih-Pin Huang[5], and Ving Ching Chong[2,*]

[1]*Institute of Ocean and Earth Sciences, University of Malaya, Kuala Lumpur, Malaysia. E-mail: cecilia@siswa.um.edu.my, khloh@um.edu.my*

[2]*Institute of Biological Science, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia. *Correspondence: E-mail: chong@um.edu.my. E-mail: ccng@um.edu.my (Ng)*

[3]*Department of Agricultural and Food Science, Faculty of Science, University Tunku Abdul Rahman, Kampar, Perak. E-mail: ooial@utar.edu.my*

[4]*Seikai National Fisheries Research Institute, Nagasaki, Japan. E-mail: ykoni21080@gmail.com*

[5]*Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan. E-mail: huangshihpin@gmail.com*

Received 3 April 2019 / Accepted 18 August 2019 / Published xx September 2019

Communicated by Benny K.K. Chan

Larval descriptions of tropical marine and coastal fishes are very few, and the taxonomic problem is further exacerbated by the high diversity of fish species in these waters. Nonetheless, accurate larval identification is crucial in ecological and early life history studies of larval fishes for the purpose of fishery management and habitat protection. The present study aimed to evaluate the usefulness of DNA barcodes to support larval fish identification since conventional dichotomous keys based on morphological traits are not efficient due to the lack of larval traits and the rapid morphological changes during ontogeny. The molecular analysis uncovered a total of 48 taxa (21 families) from the larval samples collected from Klang Strait waters encompassing both spawning and nursery grounds of marine and estuarine fishes. Thirty-two (67%) of the larval taxa were matched at the species level, whereas two taxa (4%) were identified at the genus, and 14 taxa (29%) identified at family level. The relatively low rate of species-level identification is not necessarily due to the DNA barcoding method per se, but a general lack of reference sequences for speciose and non-commercial fish families such as Gobiidae, Blenniidae, and Callionymidae. Larval morphology

remains important in species diagnosis when molecular matches are ambiguous. The lower ethanol percentage at 50% for larva preservation is also useful to keep the body of larvae intact for morphological identification, and to preserve DNA for subsequent molecular analysis. The 10% Chelex resin used for extracting DNA is also cost-effective for long term monitoring of larval fishes. Hence, future studies on larval fish ecology in the mangrove ecosystem are expected to progress further by using the DNA barcoding method since species-level identification of estuarine larval fishes is successful and made easier.

Key words: Molecular identification, Morphology, Fish larvae, Coastal fishes, Mangrove-associated.

Citation: Chu C, Loh KH, Ng CC, Ooi AL, Konishi Y, Huang SP, Chong VC. 2019. Utilization of DNA barcodes for the identification of larval fishes in tropical estuarine waters (Malacca Straits, Malaysia). Zool Stud **58:**0dd. doi:-.

BACKGROUND

One major reason why larval fish studies in the tropics are not progressing as fast as its temperate counterpart is the problem of larval fish identification. For the Indo-Pacific region, fish larvae descriptions are made more accessible only in recent publications, compared to the period prior to the first Indo-Pacific Fish Conference held in 1981 (Leis 2015). Tropical larval fish descriptions for marine and coastal fishes are very few; notable ones include Chayakul (1990) for Gulf of Thailand, Shadrin et al. (2003) for Vietnamese waters, and Leis and Carson-Ewart (2004) for the 124 families that occur in the Indo-Pacific region; the latter publication is the compilation of work done by Leis and Rennis (1983) and Leis and Trnski (1989), and the extended version of Leis and Carson-Ewart (2000). More recently, Konishi et al. (2012) provide the identification guide for 92 families of larval fishes of the important marine capture species that occur in the Southeast Asian countries. Even so, the taxonomic problem is further exacerbated by the high diversity of fish species in tropical marine waters that include several unique habitats, making it more difficult to distinguish among congenerics within family. For instance, fish larvae of more than 100 families have been reported in Thai waters (Janekarn and Kiørboe 1991), and in the Banda Sea, Indonesia,

78 families of larval fishes were collected (Soewito and Schalk 1990). In Malaysian marine and brackish waters, as many as 1481 species have been recorded, with as many as 250 being considered as marine euryhaline species (Chong et al. 2010). Liew (1992) managed to identify 61 taxa of larval fishes, mostly at family level, that was collected from the Straits of Malacca and South China Sea. Noticeably, the Malaysian estuarine coastal area has fewer families of larval fishes; 19 families in the Matang mangrove and adjacent waters (Ooi and Chong 2011), 24 families in the seagrass-mangrove area in Gelang Patah, Johor (Ara et al. 2013), and 20 families in the mangrove estuarine of Marudu Bay, Sabah (Rezagholinejad et al. 2016). From the archived ichthyoplankton samples collected in Klang Strait, 23 families of larval fishes were identified (Chu, unpublished).

The Klang Strait on the west coast of Peninsular Malaysia is a narrow, 70-km long shipping channel flanked by coastal mudflats to its right and a large sand-mud shoal on its left (Fig. 1). Several deltaic islands covered by mangrove forests enclose the strait's southern flank, breaking up its southern approach into several narrow channels. Evidently, the more offshore waters at the northern approach of Klang Strait, are spawning areas for fish and prawns while the coastal mangroves, mudflats and estuaries serve as their nursery or feeding areas (Chong and Sasekumar 1981; Chong et al. 1990). The utilization of Klang Strait as spawning and feeding ground by two engraulid species, *Stolephorus baganensis* and *Thryssa kammalensis*, was shown in the schematic diagram by Sarpedonti and Chong (2008); the former species spawn in more offshore waters before moving into the estuaries as postflexion larvae, whereas the latter species spawn closer to the shore but move further upstream until both species reach the juvenile stage and came out of the estuaries again to feed in the near inshore waters. Since fish movements performed during ontogenetic development are species-specific, the migratory routes used by other larval species may be different. So unless the larval taxonomy is resolved, the migratory routes used by many of the larval fishes in the strait during their larval development will remain unanswered.

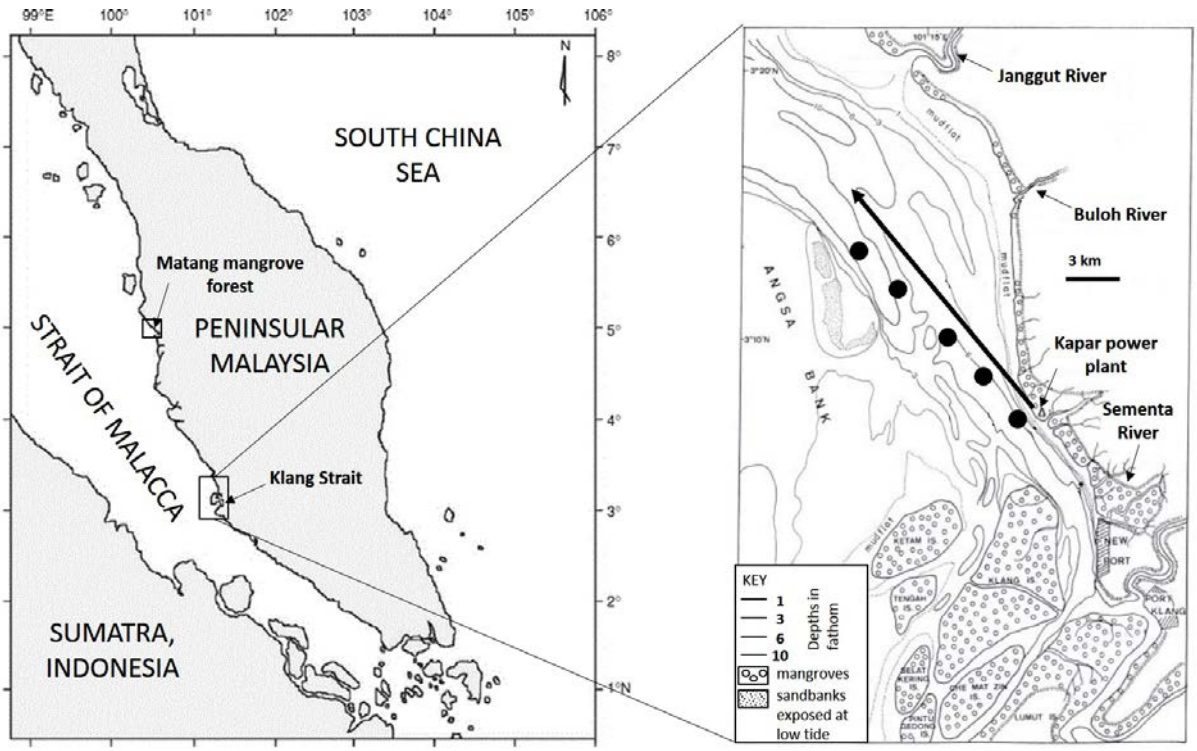

Fig. 1. Map of sampling location (left) and enlarged inset box (right) showing five sampling stations (black circles) along Klang Strait. Right arrow indicates offshore direction of transect line from Kapar power plant.

Previously, Sarpédonti et al. (2000) identified two engraulid species, *S. baganensis* and *T. kammalensis*, based on the morphology of their digestive system. But for most other larval taxa, identification is mostly done at the family, subfamily or genus level (Leis and Carson-Ewart 2004). This is because some features such as fins and scales are not fully developed, so there is a lack of characters that can be diagnosed for species level identification. Therefore, conventional dichotomous keys used to identify adult fish are not so useful for fish larvae due to too few characters and often, the rapid morphological changes during ontogeny (Leis 2015). With the advent of molecular techniques, DNA barcodes and other gene markers have been used to differentiate or confirm the identity of larval fish species (Hubert et al. 2010; Wibowo et al. 2015; Azmir et al. 2017; Aoyama et al. 2018) and crustacean larvae (*e.g.*, Chen et al. 2013; Wong et al. 2014) after morphological identification. The DNA technique transcends the limitation posed by morphological diagnoses, in that, the molecular identities of adults are enough to identify all other ontogenetic stages using readily available DNA barcodes (Hubert et al. 2015). As DNA barcodes are increasingly being deposited into accessible databases such as GenBank and BOLD (Barcode of Life Data system), more species of larval fishes can now be identified through matching of the

DNA sequences. Because of this utility, DNA barcodes have been used to link larvae to their adult species so as to answer questions concerning their early life stages (Valdez-Moreno et al. 2010). In other cases, DNA barcoding is also able to reveal cryptic species within the family (Mat Jaafar et al. 2012; Dahruddin et al. 2016). However, while the goal of barcoding is to accumulate more DNA barcodes, these barcodes must come from specimens that are accurately identified (Hubert et al. 2008). The availability of accurate DNA barcodes of adults is thus crucial to the accurate identification of their larval counterparts.

In light of the new direction in larval fish identification, further collections of fish larvae are necessary for molecular identification using DNA barcoding. The present study thus aimed to evaluate the usefulness of DNA barcodes to support the use of existing morphological traits for larval fish identification. Among the common tropical fish taxa, subsets of the diverse families of Gobiidae and Engraulidae that are typical of mangroves and tropical estuaries were tested. Problems and limitations that arise from utilization of the method are also discussed. Additionally, molecular identification using larval specimens were further refined by comparing methods for larval preservation and DNA extraction that were used by past researchers. The best methods for preserving fish larvae, extracting high DNA output and cost effectiveness in processing large quantity of larvae were evaluated.

MATERIALS AND METHODS

Collection and preservation of larval samples

Plankton samples were collected at five sampling stations, along an 18-km transect line from Kapar power plant to offshore direction in the Klang Strait (Fig. 1). The samples were collected by twin bongo nets with mouth openings of 45-cm diameter, and mesh sizes of 180 µm and 363 µm. Day-samplings were carried out in June and July of 2015, and continued in January, March, and June of 2016. At each station, the net was hauled obliquely by a moving trawl boat for 10 minutes at an approximate speed of 1.5 knots. In initial trials, collected plankton samples were immediately preserved in 99%, 80%, and 50% HPLC grade ethanol to determine the best concentration of ethanol that will keep the larval body intact, whilst preserving its DNA for molecular analysis. Since 50% ethanol showed the best results for all larval stages, subsequent samples were preserved

in this ethanol concentration. At the laboratory, all preserved samples were kept in a -20°C freezer until sorted out.

Sorting and morphological identification of fish larvae

Plankton samples were thawed at room temperature. Then, fish larvae and early juveniles were sorted out from the samples under a stereomicroscope (Leica M125) connected to an imaging system (Leica Application Suite v4.10). For subsequent molecular analysis, all sorted specimens were kept in individual vials containing 50% ethanol, and were individually photographed and measured using the imaging system. Pre- and flexion larvae were measured for notochord length (NL, mm), whereas the postflexion larvae and early juveniles were measured for their total length (TL, mm). The vials were labelled and stored at -20°C. Before DNA extraction, the specimens were identified to the lowest possible taxon by their morphological characters using key descriptions found in Leis and Carson-Ewart (2004) and/or Okiyama (2014).

Extraction and amplification of DNA

One eye of each specimen was taken out for DNA extraction. For preflexion larva, the whole body was used for extracting DNA because the eye tissue was very little. DNA were extracted using 10% Chelex resin following the procedure found in Hyde et al. (2005), with some modification. Either an eye or a small amount of tissue was inserted into a tube containing 150 µl of 10% Chelex resin (Bio-Rad) in distilled water, and the sample was first heated at 60°C for 2 minutes then at 103°C for 25 minutes. The heated tube containing the sample was then left at room temperature to cool down. PCR-amplification of the Chelex-isolated DNA ensued after cooling down, or stored at -20°C pending PCR.

Amplification of the partial cytochrome *c* oxidase subunit I (*COI*) was done using the barcoding primers and thermal program described in Ward et al. (2005). A 20 µl PCR reaction was prepared using dry Maxime PCR PreMix (iNtRON Biotechnology) by adding 1 µl of template DNA, 0.5 µl of each primer (10 µM) and 18 µl of UV-distilled water. Successful PCR products were outsourced to First Base Laboratory Private Limited (Malaysia) for purification and sequencing. Initially, both forward and reverse sequencing direction was performed per sample, and

subsequently, only one (forward) sequencing direction was done because of the robustness of the sequences obtained.

Analysis of DNA sequence

Raw DNA sequences were checked and edited using Sequence Scanner v1.0 (Applied Biosystem) to confirm the correct base calling. Then, the trimmed sequences were searched against GenBank database (https://blast.ncbi.nlm.nih.gov/) and BOLD system (http://www.boldsystems.org/) to find molecular matches. Molecular matches to larval sequences are listed in table S1. Reference sequences nearest to the subject were downloaded from the GenBank and were aligned with the targeted sequence using MEGA v.6 (Tamura et al. 2013). The genetic distances were then calculated using the Kimura 2-parameter model, and a 3% threshold for species delineation as suggested by Hebert et al. (2003) was followed in the present study. Species name was only assigned to the target sequence if it corroborated with the morphological identification *i.e.*, the genus/family of the specimens. The steps for acquiring the final larval identification are shown in figure S1.

Adult collection

Adult fish specimens were collected from local landing sites and fish markets, and directly from bag net catches from the fishing villages of Janggut River, Buloh River and Sementa River (Fig. 1). The first collection was made in September 2017, and a second collection in October 2018. Additional tissue samples of 22 species of Gobiidae and three species of Eleotridae were borrowed from the Biodiversity Research Center, Academia Sinica, Taiwan; the fishes were previously collected by Huang et al. (2013) mostly from the Matang mangrove forest located approximately 250 km to the north of the main study area. The adults were identified to species and photographed, and tissue were collected and preserved in 95% ethanol. DNA extraction, amplification, sequencing, and sequence analysis were done according to the same methods described above. *COI* sequences of adults were used to match the sequences of larvae through phylogenetic analysis using MEGA v.6 (Tamura et al. 2013).

RESULTS

Effects of larval preservation in ethanol

The fish larvae preserved directly in 99% ethanol just after collection by bongo net were badly distorted in form or damaged, so that their morphological characters were almost impossible to recognize. However, larvae kept in 50% ethanol were intact. Larger larvae at postflexion or the juvenile stage could be kept in 80% ethanol without body distortion. Essentially, samples preserved in 50% ethanol yielded sufficient DNA concentration i.e. between 7–68 ng/µl for subsequent analysis.

Larval composition

A total of 671 larval fish were collected, which consisted of preflexion, flexion and post-flexion larvae, and early juveniles. Initial identification based on morphology recorded the families of Clupeidae (*n* = 225), Blenniidae (*n* = 105), and Ambassidae (*n* = 98) as the most abundant; together, they accounted for at least 63.8% of the total larval fish collection. Families of Engraulidae (*n* = 51), Sciaenidae (*n* = 40), and Gobiidae (*n* = 34) were moderately abundant and contributed about 18.7% of the total larval collection. The other 8.0% of the larvae consisted of families having less than 10 individuals: Sillaginidae (*n* = 9), Mugilidae (*n* = 8), Callionymidae (*n* = 6), Apogonidae (*n* = 5), Carangidae (*n* = 5), Tetrarogidae (*n* = 5), Polynemidae (*n* = 3), Platycephalidae (*n* = 3), Scatophagidae (*n* = 2), Soleidae (*n* = 2), Stromateidae (*n* = 2), Hemiramphidae (*n* = 1), Gerreidae (*n* = 1), Cynoglossidae (*n* = 1), and Triacanthidae (*n* = 1). Damaged specimens made up the last 9.5% of the total larval collection.

Species assignment through DNA barcodes

Of the total larval collection, only 250 individuals that ranged from 2.0 mm to 46.0 mm TL and representing the various families and ontogenetic stages were selected for molecular identification. About 77% (193 samples) were successfully amplified for *COI* gene using the barcoding primers. The 57 samples that failed to amplify were from the families Sciaenidae (*n* = 14),

Blenniidae (*n* = 13), Gobiidae (*n* = 9), Engraulidae (*n* = 7), Ambassidae (*n* = 5), Callionymidae (*n* = 3), Tetrarogidae (*n* = 2), Apogonidae (*n* = 1), and Scatophagidae (*n* = 1).

The *COI* sequences of larvae ranged from 546–710 bp, and consisted of 48 taxa belonging to 21 families in 7 orders (Table 1). About 67%, or 32 out of the 48 taxa were identified to species level, whereas 4% (two taxa) were identified to the genus level, and 29% (14 taxa) were identified to the family level. Despite being the most abundant, the families Clupeidae and Blenniidae contained only two species each. These two families, together with Ambassidae, each consisted of one abundant species. In contrast, the families Engraulidae and Gobiidae contained the highest number of taxa, six and 13, respectively, in spite of being only moderately abundant. Family Sciaenidae, also from the moderately-abundant group, had only two successful amplifications out of the 16 extracted samples; the two successfully-amplified samples came from two species.

Table 1. List of the identified larval fish specimens and their GenBank accession number

Order	Family	Species	n	Accession No.
Beloniformes	Hemiramphidae	*Hyporhamphus quoyi*	1	MH673896
Clupeiformes	Clupeidae	*Anodontostoma chacunda*	3	MH673897 –MH673899
		Escualosa thoracata	42	MH673906 –MH673947
	Engraulidae	*Coilia dussumieri*	2	MH673900 –MH673901
		Coilia sp. 1	1	MH673902
		Stolephorus commersonii	1	MH673903
		Stolephorus dubiosus	2	MH673904 –MH673905
		Stolephorus insularis	1	MH673948
		Stolephorus tri	6	MH673949 –MH673954
Mugiliformes	Mugilidae	*Paramugil parmatus*	7	MH673955 –MH673961
		Osteomugil cunnesius	1	MH673962
Perciformes	Ambassidae	*Ambassis gymnocephalus*	34	MH673963 –MH673996
	Apogonidae	*Ostorhinchus fasciatus*	4	MH673997 –MH674000
	Blenniidae	Blenniidae sp. 1	1	MH674001
		Blenniidae sp. 2	28	MH674002 –MH674029
	Callionymidae	Callionymidae sp. 1	3	MH674030 –MH674032
	Carangidae	*Alepes djedaba*	1	MH674033
		Alepes kleinii	4	MH674034 –MH674037
	Gerreidae	*Gerres limbatus*	1	MH674038
	Gobiidae	*Acentrogobius cyanomos*	6	MH674039 –MH674044
		Hemigobius hoevenii	1	MH674045
		Parapocryptes serperaster	1	MH674046
		Tridentiger barbatus	1	MH674047
		Trypauchen sp. 1	1	MH674048
		Gobiidae sp. 1	1	MH674049
		Gobiidae sp. 2	2	MH674050 –MH674051
		Gobiidae sp. 3	1	MH674052
		Gobiidae sp. 4	1	MH674053
		Gobiidae sp. 5	1	MH674054
		Gobiidae sp. 6	1	MH674055
		Gobiidae sp. 7	6	MH674056

Order	Family	Species	n	Accession No.
				–MH674061
		Gobiidae sp. 8	2	MH674062
				–MH674063
	Polynemidae	*Eleutheronema tetradactylum*	3	MH674064
				–MH674066
	Scatophagidae	*Scatophagus argus*	1	MH674067
	Sciaenidae	*Pennahia anea*	1	MH674068
		Johnius carouna	1	MH674069
	Sillaginidae	*Sillago asiatica*	1	MH674072
		Sillago sihama	4	MH674073
				–MH674076
		Sillaginidae sp. 1	2	MH674070
				–MH674071
	Stromateidae	*Pampus argenteus*	1	MH674077
		Pampus minor	1	MH674078
Pleuronectiformes	Cynoglossidae	*Cynoglossus lingua*	1	MH674079
	Soleidae	*Zebrias zebra*	1	MH674080
		Soleidae sp. 1	1	MH674081
Scorpaeniformes	Platycephalidae	*Kumococius rodericensis*	2	MH674082
				–MH674083
		Platycephalidae sp. 1	1	MH674084
	Tetrarogidae	*Tetraroge barbata*	3	MH674085
				–MH674087
Tetraodontiformes	Triacanthidae	*Trixiphichthys weberi*	1	MH674088
TOTAL: 7 Orders, 21 Families, 48 taxa (n = 193)				

The intra-species divergence among larval sequences ranged from 0.0% to 2.2%, whereas the inter-species divergence started from 4.8% to 36.8%. For families with two or more species, the intra-family divergence ranged slightly higher at 1.7% to 32.7%. Interestingly, the families Sillaginidae, Engraulidae, and Gobiidae with respectively three, six and 13 taxa each, had smaller intra-family divergence of 14.6%, 15.7% and 19.8%, whereas the families Soleidae and Sciaenidae each with two taxa had higher intra-family divergence at 29.3% and 32.7% respectively. The smallest inter-family divergence was between Ambassidae and Apogonidae with 19.3%, and the largest was between Clupeidae and Cynoglossidae at 34.6% (Phylogenetic trees can be found in Figs. S2–S6).

Clupeidae

The larval specimens of Clupeidae could be differentiated into two species based on their monophyletic groupings. The first monophyletic group consisted of three larval specimens that matched the adult of *Anodontostoma chacunda* with an intra-species divergence of 0.4%. Sequences of the other 42 larval specimens of Clupeidae matched the reference sequences of the engraulid species *Stolephorus indicus* by 99% (KX223955) and 98% (FJ238040 and EU595310), and also to a clupeid species *Escualosa thoracata* by 98% (AP011601 and MH429324). Since they are clupeid larvae by having a long gut of which the anus is located posterior to the dorsal-fin base (Fig. 2),

matches to *S. indicus* were dismissed as wrong identifications; these 42 specimens were identified as *E. thoracata*.

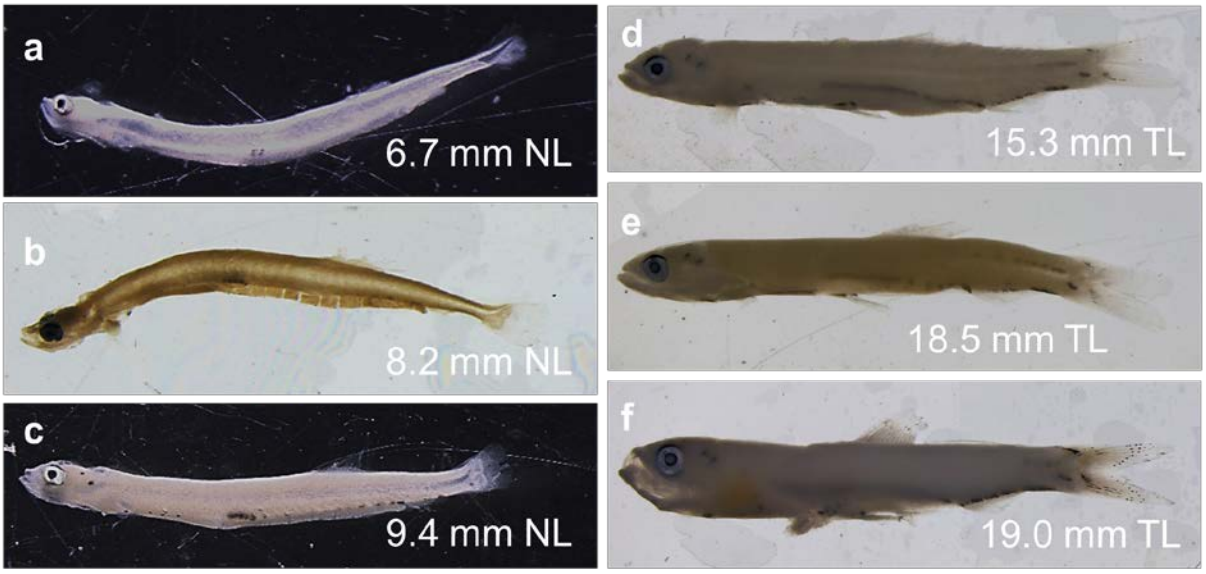

Fig. 2. Ontogenetic series of *E. thoracata* at preflexion (a); flexion (b, c); postflexion (d, e); early juvenile (f).

Engraulidae

Among the 13 larval sequences of the Engraulidae, 10 of the sequences matched the adult sequences of *Coilia dussumieri* (two matches), *Stolephorus dubiosus* (two matches), and *Stolephorus tri* (six matches) with an intra-species divergence of 1.1%, 1.6% and 0.2%, respectively. Two other specimens each belonged to *Stolephorus commersonii* and *Stolephorus insularis* based on matches with reference sequences. The last specimen has a long, distinctly tapering tail (Fig. 4c) which is a specific character for the genus *Coilia*. Since no molecular matches were found on both databases, the genus *Coilia* was applied based on its morphology and was named as *Coilia* sp. 1. Inter-species divergence between all six species ranged from 13.5% (between *Coilia* sp. 1 and *C. dussumieri*) to 24.3% (between *S. dubiosus* and *S. insularis*).

Ambassidae

All 34 sequences of the ambassid larvae matched the adult sequences of *Ambassis gymnocephalus* collected from Janggut River, with an intra-species divergence of only 0.2%. But morphologically, the ambassid larval specimens in this study had at least three morphotypes based on pigmentation on the top of the head (Fig. 3).

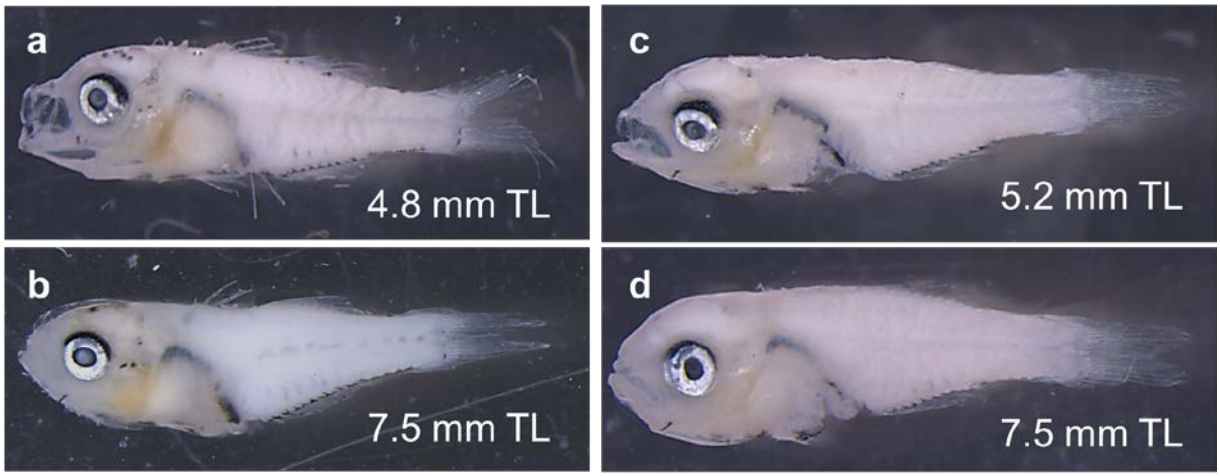

Fig. 3. Three degrees of pigmentation on the top of head of *A. gymnocephalus* larvae; heavy pigment (a); moderate pigment (b); sparse pigment (c, d).

Blenniidae

Specimens recognized as belonging to the Blenniidae could be distinguished into Blenniidae sp. 1 and Blenniidae sp. 2 based on their monophyletic groupings as well as the pigmentation along the anal-fin base; Blenniidae sp. 2 (Fig. 4b) was less pigmented at all ontogenetic stages compared to Blenniidae sp. 1 (Fig. 4a). Both types of specimens showed similar characteristics to the Tribe Omobranchini by having pigmentation on the head, pectoral fin, and along the anal-fin base, and a pair of long preopercular spine that decreases with growth (Leis and Carson-Ewart 2004). Intra-species divergence of Blenniidae sp. 2 was 0.3%, and inter-species divergence between Blenniidae sp. 1 and Blenniidae sp. 2 was large at 28.1%. The initial match to Blenniidae sp. 1 was to *Cirripectes stigmaticus* (KX223895), but the identification was doubtful since the latter did not cluster with other reference sequences of *C. stigmaticus* (Fig. S3); the sequence divergence between KX223895 and the *Cirripectes* group was 18.4%. Similarly, the initial matches to Blenniidae sp. 2 with 99% similarity were to *Lutjanus apodus* (KX223917, KX223918) and *Acentrogobius* sp. (KX144848) but both were dismissed as wrong identifications. Thus, both taxa of the Blenniidae

were not identified, and named Blenniidae sp. 1 and Blenniidae sp. 2 following the morphological identification of their family.

Gobiidae

Out of the 13 taxa recognized for the family Gobiidae, four taxa were identified as *Acentrogobius cyanomos*, *Hemigobius hoevenii*, *Parapocryptes serperaster* and *Tridentiger barbatus* based on sequence matches with an intra-species divergence of 0.4%, 0.7%, 0.8% and 0.8%, respectively. The larval sequence that was embedded between *Trypauchen pelaeos* and *Trypauchen vagina* with an inter-species divergence of 7.0% and 10.0%, respectively, could be another species of *Trypauchen*, thus named as *Trypauchen* sp. 1. Morphologically, the larva specimen showed the characteristics of the subfamily Amblyopinae as described by Leis and Carson-Ewart (2004) which contains the genus *Trypauchen*. The eight other taxa of gobiid larvae were unidentified because no close matches (98–99% similarity) were found, but they formed monophyletic groupings with less than 3% intra-species divergence. Gobiidae sp. 1, Gobiidae sp. 2, and Gobiidae sp. 3 seemed to be embedded within the species complex of *Mahidolia mystacina*; the nearest inter-species divergence of 4.8% was between Gobiidae sp. 2 and Gobiidae sp. 3, and the largest was between Gobiidae sp. 1 and Gobiidae sp. 3 with 21.0% divergence. Gobiidae sp. 4 was nearest to another gobiid species *Odontamblyopus rubicundus* with an inter-species divergence of 17.1%. Gobiidae sp. 5 was separated from Gobiidae sp. 6 by 13.9% divergence, whereas the latter was very close to *Parachaeturichthys polynema* by 4.2% divergence. Gobiidae sp. 7 had the largest intra-species divergence of 2.2% among all recognized taxa although they seemed to consist of two very closely related taxa; the distance between these two branches was 3.4%, but since they had similar morphology and pigmentation (Fig. 4o, 4p), they were similarly classified as Gobiidae sp. 7. Lastly, Gobiidae sp. 8 was only distantly related to *Scartelaos gigas* by 12.4% divergence.

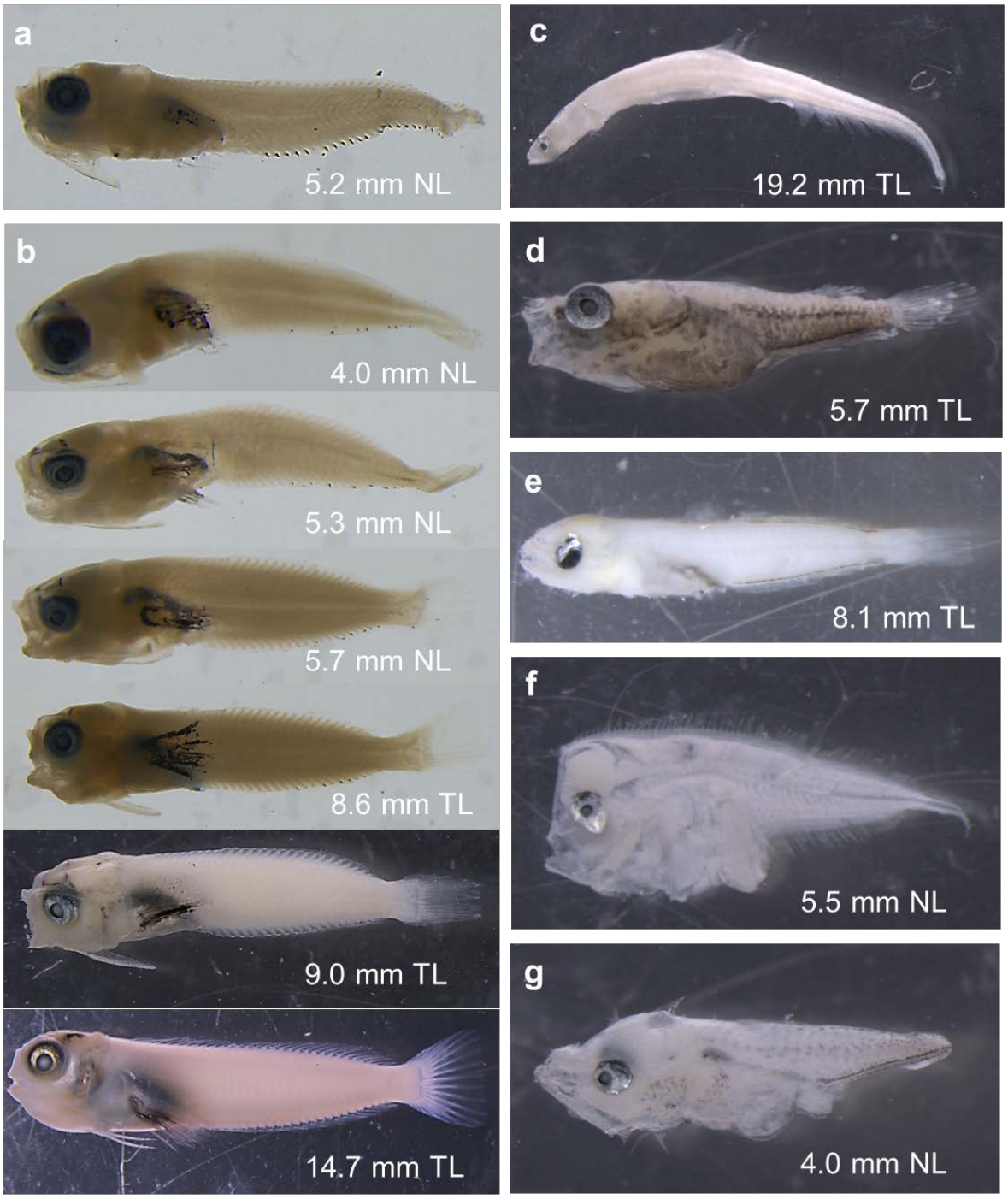

Fig. 4. Unidentified taxa. Blenniidae sp. 1 (a); Blenniidae sp. 2 (b); *Coilia* sp. 1 (c); Callionymidae sp. 1 (d); Sillaginidae sp. 1 (e); Soleidae sp. 1 (f); Platycephalidae sp. 1 (g).

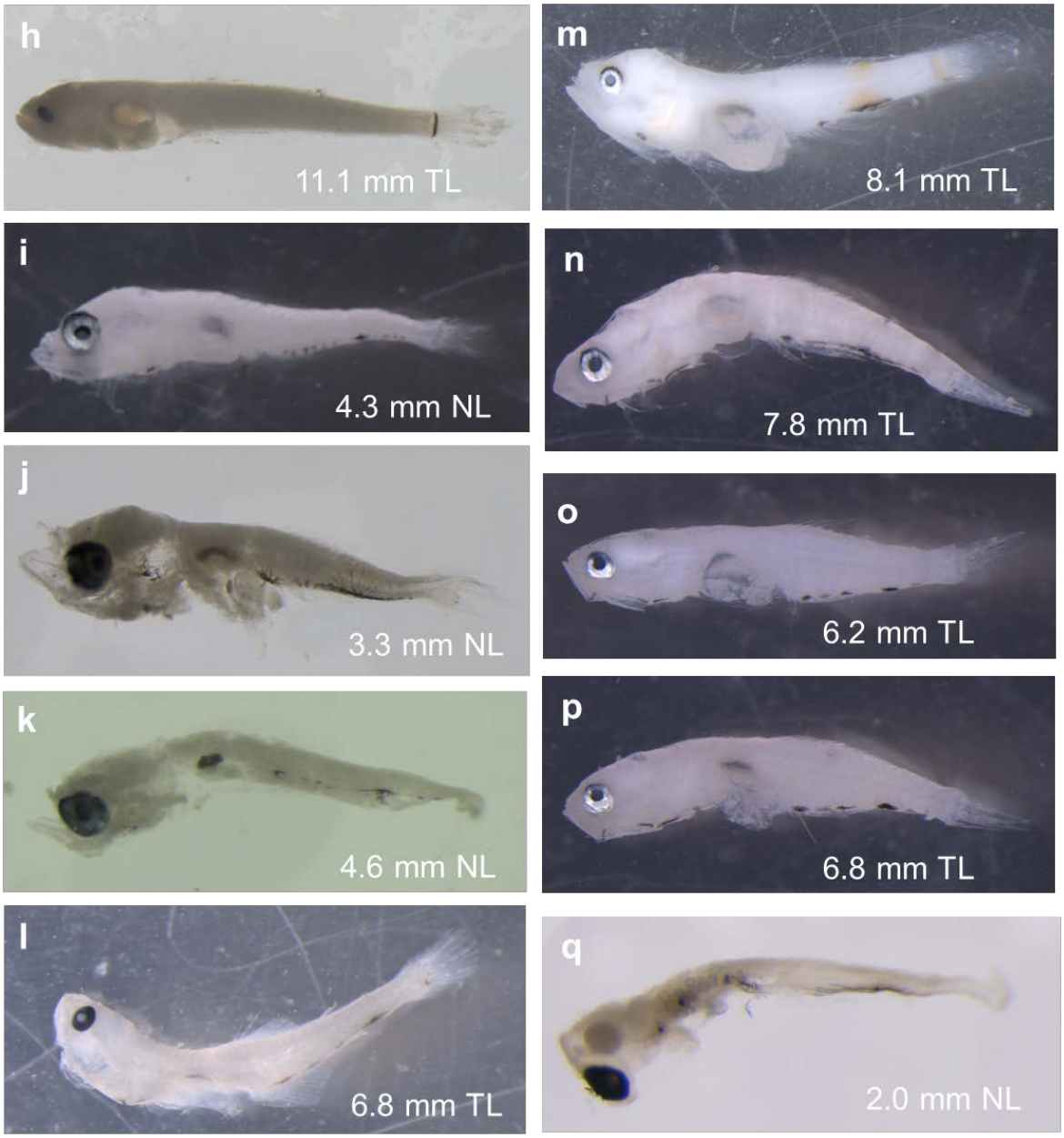

Fig. 4. Unidentified taxa. *Trypauchen* sp. 1 (h); Gobiidae sp. 1 (i); Gobiidae sp. 2 (j); Gobiidae sp. 3 (k); Gobiidae sp. 4 (l); Gobiidae sp. 5 (m); Gobiidae sp. 6 (n); Gobiidae sp. 7-1 (o); Gobiidae sp. 7-2 (p); Gobiidae sp. 8 (q).

Other families

One larval sequence of the Sillaginidae matched the adult sequence of *Sillago asiatica* with an intra-species divergence of 0.3%, whereas four other larval sequences were identified as *Sillago sihama* with 0.4% divergence. The remaining two sillaginid sequences were only distantly related to *Sillago ingenuua* with 13.4% divergence. All specimens of the families Mugilidae, Apogonidae,

Sciaenidae and Stromateidae were identified based on matches with their adult sequences i.e. the mugilid *Paramugil parmatus* and *Osteomugil cunnesius*, the apogonid *Ostorhinchus fasciatus*, the sciaenid *Pennahia anea* and *Johnius carouna*, and the stromateid *Pampus argenteus* and *Pampus minor* with an intra-species divergence of 0.4%, 0.2%, 0.2%, 0.4%, 1.0%, 0.2% and 0.5%, respectively. Specimens of the families Hemiramphidae (*Hyporhamphus quoyi*), Carangidae (*Alepes djedaba* and *A. kleinii*), Gerreidae (*Gerres limbatus*), Polynemidae (*Eleutheronema tetradactylum*), Scatophagidae (*Scatophagus argus*), Cynoglossidae (*Cynoglossus lingua*), Soleidae (*Zebrias zebra*), Platycephalidae (*Kumococius rodericensis*), Tetrarogidae (*Tetraroge barbata*) and Triacanthidae (*Trixiphichthys weberi*) were identified based on matches with their reference sequences. The last three unidentified taxa which were identified solely on their morphology were Callionymidae sp. 1 (three specimens), Soleidae sp. 1 (one specimen) and Platycephalidae sp. 1 (one specimen).

DISCUSSION

Sample preservation

Normally, plankton samples including fish larvae are preserved in buffered 4% formalin in seawater immediately after collection. But for molecular analysis, larval fish samples like their adults are kept directly in high concentration of ethanol *i.e.*, ≥ 95% to preserve DNA (Hubert et al. 2015; Azmir et al. 2017). In this study, the bodies of larval specimens were found to be highly distorted and almost impossible to identify morphologically if specimens were preserved in 95–99% ethanol. Previously, two sets of samples were taken to enable both molecular and morphological diagnoses, by preserving one in ethanol and the other in formalin (Hubert et al. 2010). However, the problem is that the two sets may not contained similar species especially if larval collections are rare or few in numbers. To circumvent this problem, Wibowo et al. (2015) preserved all larval specimens in alcohol only; initially in 50% ethanol while in the field, and later transferred to 95% ethanol after taking photos of the specimens in the laboratory. In the present study, larvae preserved in 50% ethanol not only had intact bodies suitable for morphological identification, but were also fine for molecular analysis even without gradual transfer through an alcohol series. Nevertheless, 80% ethanol preserves better the body of larger larvae i.e. postflexion or early juvenile stage.

However, since DNA were extracted from larval samples within a year from collection, the effect of preservation time on DNA quality is unknown. Following the extraction method of Hyde et al. (2005) by using only one eye of the larva, it is possible to keep the whole body intact for future examination. Exception is only for the preflexion larva where the eyes may not be sufficient; here, the whole body is used to ensure that sufficient DNA can be obtained for PCR amplification.

DNA extraction method

In the present study, five other extraction methods were tried on the larval samples, *i.e.*, NucleoSpin Tissue XS (Macherey-Nagel), DNeasy® Blood & Tissue (Qiagen), G-Spin™ Total DNA extraction kit (iNtRON Biotechnology), DNAzol (MRC, Inc.) and TRI reagent (MRC, Inc.) with an average concentration of 100.2 ng/µl, 74.2 ng/µl, 60.3 ng/µl, 24.9 ng/µl, 14.0 ng/µl, respectively. Although the 10% Chelex extraction method described in the methods section yielded much lower DNA concentration *i.e.*, 7–68 ng/µl, the handling steps are very easy (involving single-tube extraction), fast (≤ 30 min per extraction), and much less expensive (USD0.07/sample) in comparison with the extraction kits (USD1.28–3.26/sample). As in other extraction methods, the Chelex-isolation method requires fresh- and well-preserved specimens (Hajibabaei et al. 2005), but the presence of inhibitors in the Chelex-isolated DNA is commonly reported, and often leads to low amplification rate (Vigilant, 1999; Casquet et al. 2012). In our case, 77% of the larval samples extracted using Chelex managed to get positive amplification, which is slightly higher than those reported by Vigilant (1999) at 70% success rate when using naturally shed hairs as sample.

Species identification

Except for a small percentage of fish larvae with unique morphological characters, the accuracy of species-level identification by using morphological traits is very low (Ko et al. 2013). However, through the matching of the *COI* sequences of larva and adult, or with reference sequences from public databases, two thirds of the larval specimens in the present study were identified at the species level. The limited identification at species level is attributed to the lack of reference sequences, especially for speciose and non-commercial families such as Gobiidae, Blenniidae, and Callionymidae, and not due to the barcoding method *per se*. In fact, other possible and related species of the unidentified taxa of the families of Gobiidae, Blenniidae, Callionymidae,

Engraulidae, Soleidae, Platycephalidae, and Sillaginidae were also included in the phylogenetic analyses (Figs. S2–S4), but none of them matched the larval sequences. For example, the other *Coilia* species that was listed in the Klang Strait was *C. macrognathos* (Chong et al. 2012), but the larval sequence of *Coilia* sp. 1 was only distantly related to the reference sequences of *C. macrognathos* by 24.1% divergence. Besides, only two of the 22 species of known adult gobiids used in this study matched with the larval sequences. As in other recent studies, these findings implied that there is still a lot of under-studied taxa of fish larvae (Wibowo et al. 2015; Isari et al. 2017) and a wide gap in public databases to be filled (Hubert et al. 2015; Dahruddin et al. 2016; Kimmerling et al. 2017). Additionally, our larval collections covered only a few months of sampling and may not represent larvae with seasonal occurrence. There is also the possibility that not all known species of adults are found in the area as larvae or early juveniles, and *vice versa*. For example, several adult species of the Pristigasteridae have been reported in the Klang Strait by Chong et al. (2012) as well as in other Malaysian and Thai waters (Lavoué et al. 2018), but their larvae were not encountered in the present study. On the hand, larvae from the family Tetrarogidae or better known as wasp fish is a first record in the Klang Strait, although adults have never been previously reported (Chong et al. 2012). Interestingly, tetrarogid larvae were absent in the 1985-1986 and 2013-2014 archival collections of larval fishes in the Klang area (Chu, unpublished). However, a member of the family, *T. barbata*, generally resides in mangrove and coastal waters of the tropics (Myers 1991). A related species to the wasp fishes, *Vespicula trachinoides* from the family Scorpaenidae, or locally called “depu” has however been listed as present in the Klang Strait (Chong et al. 2012). Likewise, the gobiid species *T. barbatus* has never been reported in Malaysia; this species is only known to inhabit the marine and brackish waters of the Northwest Pacific and eastern Asia, although a non-indigenous occurrence in Californian waters has been widely believed to be introduced via ballast waters (Wonham et al. 2000). Since Klang Strait is the main shipping lane to Malaysia’s largest port (Port Klang), it is possible that *T. barbatus* is also a recent introduction.

Nonetheless, despite the utility of DNA barcoding as a molecular tool for classifying and identifying unknown taxa, species assignment is not a straightforward process. This is because some larval sequences may have ambiguous matches, or their molecular matches does not corroborate with the morphological identification of the larvae. For example, the *COI* sequences of the larval specimens of *E. thoracata* were identified as either an engraulid or a clupeid even with a 98–99% identity match based on the reference sequences. Morphologically, larvae of both clupeids

and engraulids that belonged to the same order (Clupeiformes) have very elongated bodies with a cross-hatching pattern in the lateral body, but the position of the anal fin relative to the dorsal fin differentiates between these two families (Leis and Carson-Ewart 2004). Larvae of the Engraulidae have an overlapped anal- and dorsal fin, or the location of the beginning of anal fin is just vertically under the end of the dorsal fin. On the other hand, the origin of the anal fin is situated posterior to the end of the dorsal-fin base for the clupeid larvae (Fig. 2). In the case for the ambassid larvae, there are at least three degrees of pigmentation on the top of the head. Interestingly, these specimens were clustered well with the adult of *A. gymnocephalus*, with such low intra-species divergence. Despite pigmentation variability within the same species, possibly due to ontogenetic development or adaptation to the environment (Gray et al. 2006), other gene markers such as the control region may help to elucidate such morphotypes or separates the subspecies (Chu et al. 2013). As for the ambiguous identity of Gobiidae sp. 1, Gobiidae sp. 2, and Gobiidae sp. 3 which were embedded among the species of *Mahidolia mystacina*, they could represent many lineages or cryptic species hidden within *Mahidolia* (Thacker et al. 2011). However, Zemlak et al. (2009) showed that sequence divergences of more than 3.5% indicate congeneric species and not a subpopulation. Therefore, Gobiidae sp. 1, Gobiidae sp. 2 and Gobiidae sp. 3 which had larger than 3.5% inter-species divergence are regarded as separate species in this study. In contrast, Gobiidae sp. 7-1 and Gobiidae sp. 7-2 clearly reflect the two lineages of similar pigmentation and morphology, rather than two separate species.

Regardless of the successful identification and naming of species, DNA barcode is also particularly useful for recognizing larvae of various ontogenetic stages as belonging to a particular species. So that when the larvae of various ontogenetic stages were lined up, developmental series of that particular species, for example, *E. thoracata* (Fig. 2) can be described. Most importantly, larval distribution and migratory route used by larvae at different ontogenetic stages can be done by matching larval sequences with those of adult species found in the area. As the collection of adult sequences further develops, the identification of more larval species becomes increasingly feasible.

CONCLUSIONS

Identification of the larval fishes has been made more effective in the present study by combining and refining the methods developed by past workers. A good compromise between the

special needs for morphological and molecular diagnoses has been determined, in that fish larvae can be directly preserved in either 50% ethanol or 80% ethanol (for postflexion larvae or larger) to keep the body intact while adequately preserving the DNA. The Chelex resin as a cost-effective extraction method could be considered for future monitoring of larval fishes. As this study is a first attempt to identify the species of larval fishes that occur in the Klang Strait using DNA barcode, limited identification at the species level was observed for speciose and non-commercial families such as Gobiidae, Blenniidae, and Callionymidae, as well as for some fish families with economic importance such as Engraulidae, Sillaginidae, Soleidae, and Platycephalidae. This is because many adult species found in the area have not been sequenced, or even documented. Nonetheless, the usefulness of DNA barcodes as a technique for larval species identification has been demonstrated, as most of the recognized taxa have smaller intra-species divergence compared to the inter-species divergence. In some cases, larval morphology remains important in species diagnosis especially if the molecular matches are ambiguous. However, there may be exceptions to the use of DNA barcodes as in the case of *Ambassis* where intra-species variation in morphology (pigmentation levels) warrants the use of other gene markers to elucidate the incongruency between morphological and molecular identity. It is anticipated that larval ecology and fishery studies will greatly benefit with the added use of DNA barcodes to help resolve species-level identification of the diverse taxa of larval fishes found near or in crucial nursery areas, such as mangroves and estuaries.

List of abbreviations

DNA, deoxyribonucleic acid.
COI, cytochrome *c* oxidase subunit I.
BOLD, Barcode of Life Data System.
HPLC, high pressure liquid chromatography.
PCR, polymerase chain reaction.
TL, total length.
NL, notochord length.
n, number of sample.

Acknowledgments: We are grateful to the Department of Fisheries Malaysia for granting a research permit to conduct fish samplings and to the Marine Police Division for a permit to work

within the harbor area. We gratefully thank Lim Yew Chai, our skipper, for skillfully using his boat for plankton sampling, Li Lee Chew for conducting field samplings, and various student members in Laboratory B201 for field assistance. This work was supported by the Ministry of Higher Education of Malaysia (MoHE) via the High Impact Research grant (UM.C/625/1/HIR/MOHE/SC/19; 2015–2017), and University of Malaya through PPP grant (PG193-2015A; 2015–2018).

Authors' contributions: CC acquired funding, performed the field work, optimised the protocol, performed the molecular and phylogenetic analyses, and wrote the manuscript. VCC, KHL, CCN, and ALO supervised CC during fieldwork and laboratory work. YK confirmed the morphological identification of larval specimens. SPH identified adult specimens of Gobiidae. VCC designed the research, acquired funding and helped in the manuscript writing. All authors participated in revising the manuscript, and approved the final manuscript.

Competing interests: CC, KHL, CCN, ALO, YK, SPH, and VCC declare that they have no conflict of interest.

Availability of data and materials: DNA sequences of larvae were deposited into GenBank database under the accession numbers MH673896–MH674088. For specimens with one eye taken out, their remaining bodies were kept as voucher specimens, and were deposited in the Zoological Museum of University of Malaya; and the list of samples with their voucher codes are given in table S2. Photos and morphological details of all specimens will be used to set up a database for Malaysian fishes; data can be shared upon request.

Consent for publication: Not applicable.

Ethics approval consent to participate: Not applicable.

REFERENCES

Aoyama J, Wouthuyzen S, Miller MJ, Sugeha HY, Kuroki M, Watanabe S, Syahailatua A, Tantu FY, Hagihara S, Triyanto, Otake T, Tsukamoto K. 2018. Reproductive ecology and biodiversity of freshwater eels around Sulawesi Island Indonesia. Zool Stud **57:**30. doi:10.6620/ZS.2018.57-30.

Ara R, Arshad A, Amin SMN, Mazlan AG. 2013. Temporal and spatial distribution of larval fish assemblage in different ecological habitat in Johor Strait, Peninsular Malaysia. Asian J Anim Vet Adv **8:**53–62. doi:10.3923/ajava.2013.53.62.

Azmir I, Esa Y, Amin S, Md Yasin I, Md Yusof F. 2017. Identification of larval fish in mangrove areas of Peninsular Malaysia using morphology and DNA barcoding methods. J Appl Ichthyol **33:**998–1006. doi:10.1111/jai.13425.

Casquet J, Thebaud C, Gillespie RG. 2012. Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. Mol Ecol Resour **12:**136–141. doi:10.1111/j.1755-0998.2011.03073.x.

Chayakul R. 1990. Larvae and juvenile fishes occurring in the Gulf of Thailand. Technical Paper of Marine Resources Survey Group, Marine Fisheries Division, Thailand Dep Fisher **4/190:**1–98.

Chen H-N, Høeg JT, Chan BKK. 2013. Morphometric and molecular identification of individual barnacle cyprids from wild plankton: an approach to detecting fouling and invasive barnacle species. Biofouling **29:**133–145. doi:10.1080/08927014.2012.753061.

Chong VC, Sasekumar A. 1981. Food and feeding habits of the white prawn *Penaeus merguiensis*. Mar Ecol Prog Ser **5:**185–191. doi:10.3354/meps005185.

Chong VC, Lee PKY, Lau CM. 2010. Diversity, extinction risk and conservation of Malaysian fishes. J Fish Biol **76:**2009–2066. doi:10.1111/j.1095-8649.2010.02685.x.

Chong VC, Sasekumar A, Zgozi S. 2012. Ecology of fish and shrimp communities. *In*: Sasekumar A, Chong VC (eds) Mangrove and coastal environment of Selangor, Malaysia Monograph series 14. University of Malaya, Kuala Lumpur, pp. 215–241.

Chong VC, Sasekumar A, Leh MUC, D'Cruz R. 1990. The fish and prawn communities of a Malaysian coastal mangrove system, with comparisons to adjacent mud flats and inshore waters. Estuar Coast Shelf Sci **31:**703–722. doi:10.1016/0272-7714(90)90021-I.

Chu C, Rizman-Idid M, Chong VC. 2013. Phylogenetic relationships of selected genera of Lutjanidae inferred from mitochondrial regions, with a note on the taxonomic status of *Pinjalo pinjalo*. Cienc Mar **39:**349–361.

Dahruddin H, Hutama A, Busson F, Sauri S, Hanner R, Keith P, Hadiaty R, Hubert N. 2016. Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species. Mol Ecol Resour **17:**288–299. doi:10.1111/1755-0998.12528.

Gray AK, Kendall Jr AW, Wing BL, Carls MG, Heifetz J, Li Z, Gharrett AJ. 2006. Identification and first documentation of larval rockfishes in Southeast Alaskan waters was possible using mitochondrial markers but not pigmentation patterns. Trans Am Fish Soc **135:**1–11.

Hajibabaei M, deWaard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL, Mackie PM, Hebert PDN. 2005. Critical factors for assembling a high volume of DNA barcodes. Phil Trans R Soc B **360:**1959–1967. doi:10.1098/rstb.2005.1727.

Hebert PDN, Ratnasingham S, deWaard JR. 2003. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B Biol Sci **270:**S96–S99. doi:10.1098/rsbl.2003.0025.

Huang S-P, Shao K-T, Huang H-M, Chong VC, Chen I-S. 2013. An annotated checklist of gobioid fishes from the mangrove estuary of Matang, Malay peninsula, with comments on a new *Pseudogobius* (Teleostei: Gobiidae) species. J Mar Sci Tech **21:**106–116. doi:10.6119/JMST-013-1219-8.

Hubert N, Delrieu-Trottin E, Irisson J-O, Meyer C, Planes S. 2010. Identifying coral reef fish larvae through DNA barcoding: a test case with the families Acanthuridae and Holocentridae. Mol Phylogenet Evol **55:**119–1203. doi:10.1016/j.ympev.2010.02.023.

Hubert N, Espiau B, Meyer C, Planes S. 2015. Identifying the ichthyoplankton of a coral reef using DNA barcodes. Mol Ecol Resour **15:**57–67. doi:10.1111/1755-0998.12293.

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, Watkinson D, Dumont P, Curry A, Bentzen P, Zhang J, April J, Bernatchez L. 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE **3:**e2490. doi:10.1371/journal.pone.0002490.

Hyde J, Lynn E, Humphreys Jr R, Musyl M, West A, Vetter R. 2005. Shipboard identification of fish eggs and larvae by multiplex PCR, and description of fertilized eggs of blue marlin, shortbill spearfish, and wahoo. Mar Ecol Prog Ser **286:**269–277.

Isari S, Pearman JK, Casas L, Michell CT, Curdia J, Berumen ML, Irigoien X. 2017. Exploring the larval fish community of the central Red Sea with an integrated morphological and molecular approach. PLoS ONE **12:**e0182503. doi:10.1371/journal.pone.0182503.

Janekarn V, Kiørboe T. 1991. Temporal and spatial distribution of fish larvae and their environmental biology in Phang-Nga Bay, Thailand. Phuket Mar Biol Cent Res Bull **56:**23–40.

Kimmerling N, Zuqert O, Amitai G, Gurevich T, Armoza-Zvuloni R, Kolesnikov I,Berenshtein I, Melamed S, Gilad S, Benjamin S, Rivlin A, Ohavia M, Paris CB, Holzman R, Kiflawi M, Sorek R. 2017. Quantitative species-level ecology of reef fish larvae via metabarcoding. Nat Ecol Evol **2:**306–316. doi:10.1038/s41559-017-0413-2.

Ko H-L, Wang Y-T, Chiu T-S, Lee M-A, Leu M-Y, Chang K-Z, Chen W-Y, Shao K-T. 2013. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS ONE **8:**e53451. doi:10.1371/journal.pone.0053451.

Konishi Y, Chayakul R, Chamchang C, Duangdee T. 2012. Early stages of marine fishes in the Southeast Asian region. Southeast Asian Fisheries Development Center, Samutprakan, Thailand.

Lavoué S, Ratmuangkhwang S, Ho HC, Chen WJ, Siti Azizah MN. 2018. Swimbladder evolution of longfin herrings (Pristigasteridae, Teleostei). Zool Stud **57:**39. doi:10.6620/ZS.2018.57-39.

Leis JM. 2015. Taxonomy and systematics of larval Indo-Pacific fishes: a review of progress since 1981. Ichthyol Res **62:**9–28. doi:10.1007/s10228-014-0426-7.

Leis JM, Carson-Ewart BM (eds). 2000. The larvae of Indo-Pacific coastal fishes: an identification guide to marine fish larvae (Fauna Malesiana Handbooks, vol. 2). Brill, Leiden.

Leis JM, Carson-Ewart BM (eds). 2004. The larvae of Indo-Pacific coastal fishes: an identification guide to marine fish larvae (Fauna Malesiana Handbooks, vol. 2). Brill, Leiden.

Leis JM, Rennis DS. 1983. The larvae of Indo-Pacific coral reef fishes. New South Wales University Press and University Press of Hawaii, Sydney and Honolulu.

Leis JM, Trnski T. 1989. The larvae of Indo-Pacific shorefishes. New South Wales University Press and University Press of Hawaii, Sydney and Honolulu.

Liew HC. 1992. Larval fish occurrence and recruitment in the seas around Malaysia. In: Sugimoto T, Romimohtarto K, Soemodihardjo S, Nakata H (eds) Proceedings of Third ORI-LIPI Seminar on Southeast Asia Marine Science, Oceanography for Fisheries, Tokyo, 1992.

Mat Jaafar TNA, Taylor MI, Mohd Nor SA, de Bruyn M, Carvalho GR. 2012. DNA barcoding reveals cryptic diversity within commercially exploited Indo-Malay Carangidae (Teleosteii: Perciformes). PLoS ONE **7:**e49623. doi:10.1371/journal.pone.0049623.

Myers RF. 1991. Micronesian reef fishes: a practical guide to the identification of the inshore marine fishes of the tropical central and western Pacific. Coral Graphics, Guam.

Okiyama M (ed). 2014. An atlas of early stage fishes in Japan second edition. Tokai University, Hadano.

Ooi AL, Chong VC. 2011. Larval fish assemblages in a tropical mangrove estuary and adjacent coastal waters: offshore–inshore flux of marine and estuarine species. Cont Shelf Res **31:**1599–1610. doi:10.1016/j.csr.2011.06.016.

Rezagholinejad S, Arshad A, Amin SMN, Ara R. 2016. Composition and diversity of larval fish in the mangrove estuarine area of Marudu Bay, Sabah, Malaysia. J Environ Biol **37:**697–703.

Sarpedonti V, Chong VC. 2008. Abundance and distribution of *Stolephorus baganensis* Hardenberg 1933 and *Thryssa kammalensis* (Bleeker 1849) larvae in relation to ontogeny and environmental factors in a Malaysian estuary. Trop Zool **21:**195–208.

Sarpédonti V, Ponton D, Chong VC. 2000. Description and ontogeny of young *Stolephorus baganensis* and *Thryssa kammalensis*, two Engraulidae from Peninsular Malaysia. J Fish Biol **56:**1460–1476. doi:10.1111/j.1095-8649.2000.tb02157.x.

Shadrin AM, Pavlov DS, Astakhov DA, Novikov GG. 2003. Atlas of eggs and larvae of coastal fishes of southern Vietman. Geos, Moscow.

Soewito, Schalk PH. 1990. Spatial and seasonal patterns on fish larvae distribution in the Banda Sea, Indonesia. Neth J Sea Res **25:**591–600. doi:10.1016/0077-7579(90)90081-Q.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725–2729.

Thacker CE, Thompson AR, Roje DM. 2011. Phylogeny and evolution of Indo-Pacific shrimp-associated gobies (Gobiiformes: Gobiidae). Mol Phylogenet Evol **59:**168–176. doi:10.1016/j.ympev.2011.02.007.

Valdez-Moreno M, Vásquez-Yeomans L, Elías-Gutiérrez M, Ivanova NV, Hebert PDN. 2010. Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. Mar Freshwater Res **61:**655–671. doi:10.1071/MF09222.

Vigilant L. 1999. An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs. Biol Chem **380:**1329–1331. doi:10.1515/BC.1999.169.

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.

Wibowo A, Sloterdijk H, Ulrich SP. 2015. Identifying Sumatran peat swamp fish larvae through DNA barcoding, evidence of complete life history pattern. Procedia Chem **14:**76–84. doi:10.1016/j.proche.2015.03.012.

Wong JY, Chen H-N, Chan BKK, Tan IKP, Chong VC. 2014. A combined morphological and molecular approach in identifying barnacle cyprids from the Matang Mangrove Forest Reserve in Malaysia: essentials for larval ecology studies. Raffles B Zool **62:**317–329.

Wonham MJ, Carlton JT, Ruiz GM, Smith LD. 2000. Fish and ships: relating dispersal frequency to success in biological invasions. Mar Biol **136:**1111–1121. doi:10.1007/s002270000303.

Zemlak TS, Ward RD, Connell AD, Holmes BH, Hebert PDN. 2009. DNA barcoding reveals overlooked marine fishes. Mol Ecol Resour **9:**237–242. doi:10.1111/j.1755-0998.2009.02649.x.

Supplementary Materials

Table S1. List of larval specimens, their morphological identification, and molecular matches from GenBank and BOLD Systems databases. (download)

Table S2. List of larval sample with the code, GenBank accession number, code of voucher specimens, and its species name. (download)

Fig. S1. Steps for acquiring the final identification of larvae. (download)

Fig. S2. Neighbour-Joining phylogenetic tree of Gobiidae based on 598 bp of the *COI* gene. Bootstrap values more than 80% are shown at the branch node. Larval sequences from this study is shown in blue font, whereas adult sequences are in red. The yellow highlights indicate the lineages of *Mahidolia mystacina* from reference sequences. (download)

Fig. S3. Neighbour-Joining phylogenetic tree of Blenniidae based on 604 bp of the *COI* gene. Bootstrap values more than 80% are shown at the branch node. Larval sequences from this study is shown in blue font. The yellow highlights indicate reference sequences that may be misidentified. (download)

Fig. S4. Neighbour-Joining phylogenetic tree of unidentified taxa based on 530 bp of the *COI* gene. Bootstrap values more than 80% are shown at the branch node. Larval sequences from this study is shown in blue font, whereas adult sequences are in red. (download)

Fig. S5. Neighbour-Joining phylogenetic tree of identified taxa based on 489 bp of the *COI* gene. Bootstrap values more than 80% are shown at the branch node. Larval sequences from this study is shown in blue font, whereas adult sequences are in red. The yellow highlights indicate the reference sequences that might be misidentified. (download)

Fig. S6. Neighbour-Joining phylogenetic tree of identified taxa based on 596 bp of the *COI* gene. Bootstrap values more than 80% are shown at the branch node. Larval sequences from this study is shown in blue font, whereas adult sequences are in red. (download)