



OPEN

DNA-based taxonomy of a mangrove-associated community of fishes in Southeast Asia

Danial Hariz Zainal Abidin^{1✉}, Siti Azizah Mohd. Nor^{2,3✉}, Sébastien Lavoué², Masazurah A. Rahim⁴, Noorul Azliana Jamaludin^{1,5} & Noor Adelyna Mohammed Akib^{1,2✉}

The Merbok Estuary comprises one of the largest remaining mangrove forests in Peninsular Malaysia. Its value is significant as it provides important services to local and global communities. It also offers a unique opportunity to study the structure and functioning of mangrove ecosystems. However, its biodiversity is still partially inventoried, limiting its research value. A recent checklist based on morphological examination, reported 138 fish species residing, frequenting or subject to entering the Merbok Estuary. In this work, we reassessed the fish diversity of the Merbok Estuary by DNA barcoding 350 specimens assignable to 134 species initially identified based on morphology. Our results consistently revealed the presence of 139 Molecular Operational Taxonomic Units (MOTUs). 123 of them are congruent with morphology-based species delimitation (one species = one MOTU). In two cases, two morphological species share the same MOTU (two species = one MOTU), while we unveiled cryptic diversity (i.e. COI-based genetic variability > 2%) within seven other species (one species = two MOTUs), calling for further taxonomic investigations. This study provides a comprehensive core-list of fish taxa in Merbok Estuary, demonstrating the advantages of combining morphological and molecular evidence to describe diverse but still poorly studied tropical fish communities. It also delivers a large DNA reference collection for brackish fishes occurring in this region which will facilitate further biodiversity-oriented research studies and management activities.

Estuaries and coastal wetlands which feature mangrove ecosystem are transition zones that link terrestrial and freshwater habitats with the sea¹. Mangrove ecosystem delivers essential ecosystem services, including shoreline protection, nutrient production and fisheries resources. In consequence, mangrove ecosystem plays a vital role in supporting local communities' socio-economic pursuits². Unfortunately, such crucial human-nature relationship is threatened by habitat pollution, destruction, and overfishing³. It is also impacted by other factors such as species invasion, and climate change¹.

The less disturbed tropical estuaries, especially their mangrove area, generally harbour rich, unique and complex faunal communities⁴, combining the presence of salinity-tolerant resident species along with regularly or occasionally frequenters. Frequenters include mainly marine species, which use this ecosystem either to feed, shelter, breed, or nurse their young⁵. Inventorying and monitoring biodiversity in these ecosystems is primordial for long term sustainability because biodiversity ensures stability and resistance towards any disturbance or potential invasion through complex species-species interactions¹. However, biodiversity is still poorly documented in many mangrove ecosystems, particularly those of Southeast Asia, which hampers further research on their functioning and management.

Malaysia is part of the Sundaland biodiversity hotspot, which is recognized for its astounding levels of diversity and endemism^{6,7}. Considering only fishes, ⁸ reported the presence of a total of 1418 marine and brackish species in Malaysian waters, occupying various coastal habitats, including the threatened mangrove ecosystems⁹.

¹Centre for Global Sustainability Studies (CGSS), Level 5, Hamzah Sendut Library, Universiti Sains Malaysia, 11800 Penang, Malaysia. ²School of Biological Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia. ³Institute of Marine Biotechnology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia. ⁴Fisheries Research Institute, 11960 Batu Maung, Penang, Malaysia. ⁵Fisheries Research Institute, Kampung Acheh, 32000 Sitiawan, Perak, Malaysia. ✉email: geneprodigy@gmail.com; s.azizah@umt.edu.my; adelyna@usm.my

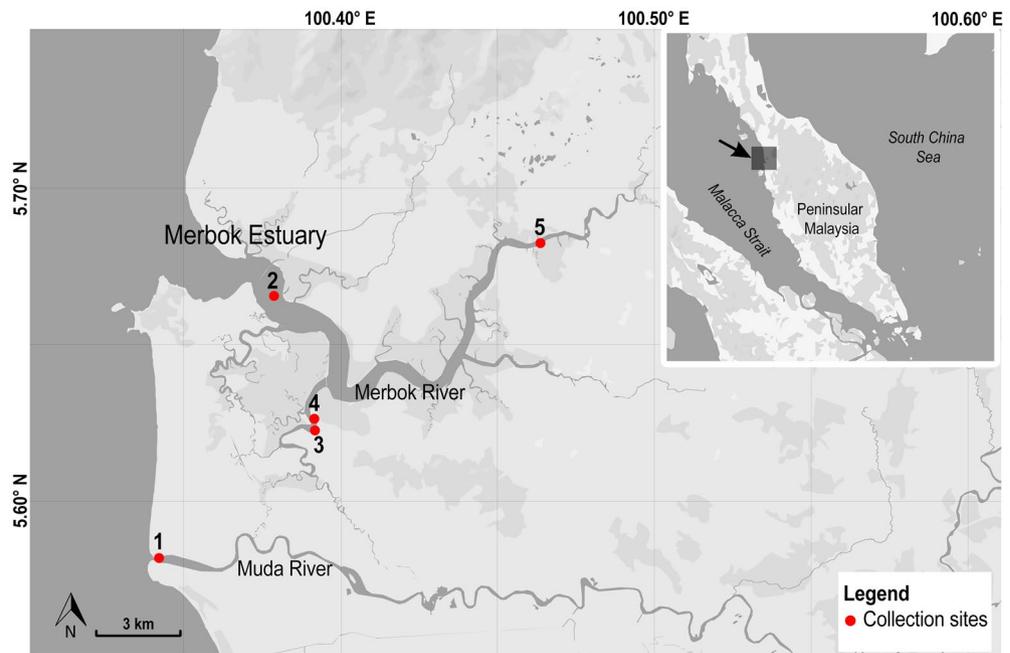


Figure 1. Sampling localities across the study area, which covers the Merbok Estuary (Merbok River) and Muda River. Sampling sites; 1: Kuala Muda Whispering Market, 2: Pompang Sungai Merbok, 3 and 4: Pompang Batu Lintang, 5: Semeling Bridge. Inset map shows the location of the study area within Peninsular Malaysia. Maps are generated using QGIS v.3.4.11 and edited in Adobe Photoshop CC 2019.

One of the largest remaining intact patches of mangrove forests is located within the Merbok Estuary, north-west Peninsular Malaysia, facing the Strait of Malacca (Fig. 1). The estuary was gazetted as a permanent forest reserve, the Sungai Merbok Mangrove Forest Reserve in 1951, and is the second largest mangrove forest in Peninsular Malaysia after the Larut Matang Forest Reserve. The Merbok Estuary and its surroundings constitute a dynamic and productive ecosystem, which supports the World's highest mangrove species diversity per unit area within a contiguous habitat, with 39 of the estimated 70 true mangroves species described globally¹⁰. This area also represents important resource grounds for local populations^{11,12}.

Due to its biological, ecological, and socio-economic importance, the Merbok Estuary has been the focus of research during the last two decades, including some biodiversity inventories (trees and gastropods¹³; shrimps¹⁴; fishes^{12,14}; mangrove trees¹⁰). The latest ichthyological survey has inventoried 138 fish species from 47 families in the estuary and adjacent marine environment, revealing a rich fish fauna¹⁵. However, because of taxonomical uncertainties when considering morphological characters alone, the identifications of some species were challenging, especially for some speciose families such as Mugilidae, Gobiidae or Eleotridae¹⁵. Furthermore, cryptic diversity is frequently encountered in tropical highly biodiverse regions^{16,17}, and it is possible that some morphology-based species hide more than one species. In Merbok as elsewhere, a precise account of species diversity is a necessary requirement for further researches and numerous studies have highlighted the complementarity between morphological and molecular approaches to reveal biodiversity^{18–20}. To date, there is no attempt to compare morphology-based results on fish diversity with genetics-based approach in this mangrove species-rich community.

Since its introduction in the past decades, DNA barcoding has emerged as the global molecular taxonomic method across fishes based on a standard molecular marker, a ~ 650 base pairs long fragment of the mitochondrial cytochrome oxidase I gene (COI)²¹. Several regional DNA barcoding studies have demonstrated its efficacy to delimitate marine fish species, for instance, in Australia²², South China Sea²³, Indian Ocean²⁴, and Indo-Pacific coral fishes²⁵. DNA barcoding has proven to be a reliable method in detecting cryptic and potentially new fish species^{26–29}, identifying larval fishes^{30–33}, or tracing back food origins³⁴.

In this study, we assemble a reference library of DNA barcodes of 350 fish individuals from Merbok Estuary and its adjacent waters for the purpose to describe the fish diversity in this region in providing a complementary look at previous morphology-based results. Comprehensive species lists built on integrative taxonomy have wide applications including ecosystem health management, biodiversity monitoring and conservation, aquaculture and fishery management^{35,36}. All of these uses pertain to the Merbok Estuary.

Results

Fish diversity. A total of 350 specimens (out of 441 collected) were successfully sequenced for the COI gene, representing 134 morphological species, 94 genera, 47 families, 17 orders, and two classes, Chondrichthyes and Actinopterygii (taxonomic list shown Table 1). Two of these species (i.e. *Cryptocentrus* sp. and *Johnius* sp.) were

ORDER, Family, Species	n	BOLD ID	MUSEUM CATALOGUE NO
MYLIOBATIFORMES			
Dasyatidae			
<i>Brevitrygon walga</i>	1	DBMR332-20	USMFC (1) 00003
<i>Telatrygon zugei</i>	1	DBMR237-19	USMFC (1) 00002
Gymnuridae			
<i>Gymnura poecilura</i>	1	DBMR036-19	USMFC (104) 00001
ORECTOLOBIFORMES			
Hemiscylliidae			
<i>Chiloscyllium indicum</i>	1	DBMR335-20	USMFC (114) 00001
ANGUILLIFORMES			
Ophichthidae			
<i>Pisodonophis cancrivorus</i>	1	DBMR001-19	USMFC (106) 00001
CLUPEIFORMES			
Chirocentridae			
<i>Chirocentrus nudus</i>	1	DBMR331-20	USMFC (72) 00001
Clupeidae			
<i>Anodontostoma chacunda</i>	1	DBMR341-20	USMFC (5) 00005
<i>Escualosa thoracata</i>	3	DBMR342-20—DBMR344-20	USMFC (5) 00003
<i>Sardinella albella</i>	2	DBMR224-19, DBMR315-19	USMFC (5) 00007, USMFC (5) 00009
Dussumieriidae			
<i>Dussumieria albulina</i>	3	DBMR316-19—DBMR318-19	USMFC (103) 00001
Engraulidae			
<i>Setipinna taty</i>	1	DBMR009-19	USMFC (82) 00047
<i>Stolephorus baweanensis</i>	1	DBMR011-19	USMFC (82) 00045
<i>Stolephorus mercurius</i>	1	DBMR314-19	USMFC (82) 00050
<i>Stolephorus baganensis</i>	7	DBMR220-19—DBMR223-19, DBMR311-19—DBMR313-19	USMFC (82) 00038, USMFC (82) 00049
<i>Stolephorus indicus</i>	1	DBMR340-20	USMFC (82) 00044
<i>Stolephorus tri</i>	1	DBMR010-19	USMFC (82) 00039
<i>Thryssa hamiltonii</i>	4	DBMR012-19—DBMR014-19, DBMR016-19	USMFC (82) 00041, USMFC (82) 00043
<i>Thryssa kammalensis</i>	4	DBMR018-19—DBMR021-19	USMFC (82) 00042, USMFC (82) 00046
<i>Thryssa mystax</i>	2	DBMR015-19, DBMR017-19	USMFC (82) 00051, USMFC (82) 00052
Pristigasteridae			
<i>Ilisha melastoma</i>	6	DBMR022-19—DBMR027-19	USMFC (92) 00004, USMFC (92) 00006
<i>Opisthopterus tardoore</i>	2	DBMR028-19—DBMR029-19	USMFC (92) 00007
SILURIFORMES			
Ariidae			
<i>Arius gagara</i>	3	DBMR273-19—DBMR275-19	USMFC (66) 00005
<i>Arius maculatus</i>	5	DBMR192-19 -DBMR196-19	USMFC (66) 00008
<i>Hexanematichthys sagor</i>	3	DBMR198-19—DBMR200-19	USMFC (66) 00002
<i>Ketengus typus</i>	1	DBMR201-19	USMFC (66) 00006
<i>Osteogeniosus militaris</i>	4	DBMR202-19—DBMR204-19, DBMR345-20	USMFC (66) 00009
<i>Plicofollis argyroleuron</i>	3	DBMR205-19—DBMR207-19	USMFC (66) 00003, USMFC (66) 00004
<i>Plicofollis layardi</i>	1	DBMR272-19	USMFC (66) 00007
<i>Plicofollis polystaphylodon</i>	1	DBMR197-19	USMFC (66) 00010
Plotosidae			
<i>Plotosus canius</i>	1	DBMR208-19	USMFC (93) 00002
AULOPIFORMES			
Synodontidae			
<i>Saurida micropectoralis</i>	1	DBMR329-20	USMFC (51) 00003
BATRACHOIDIFORMES			
Batrachoididae			
<i>Allenbatrachus grunniens</i>	1	DBMR336-20	USMFC (102) 00002
<i>Batrachomoeus trispinosus</i>	1	DBMR002-19	USMFC (102) 00001
GOBIIFORMES			
Eleotridae			
<i>Butis butis</i>	5	DBMR072-19, DBMR300-19—DBMR301-19, DBMR322-19, DBMR323-19	USMFC (33) 00012, USMFC (33) 00003, USMFC (33) 00014
Continued			

ORDER, Family, Species	n	BOLD ID	MUSEUM CATALOGUE NO
<i>Butis humeralis</i>	8	DBMR065-19—DBMR071-19, DBMR306-19	USMFC (33) 00001, USMFC (33) 00002, USMFC (33) 00004
<i>Butis koilomatodon</i>	1	DBMR302-19	USMFC (33) 00013
Gobiidae			
<i>Acentrogobius caninus</i>	1	DBMR082-19	USMFC (34) 00015
<i>Boleophthalmus boddarti</i>	1	DBMR083-19	USMFC (34) 00016
<i>Brachygobius aggregatus</i>	3	DBMR285-19—DBMR287-19	USMFC (34) 00020
<i>Exyrias puntang</i>	3	DBMR084-19, DBMR304-19, DBMR339-20	USMFC (34) 00018, USMFC (34) 00025
<i>Favonigobius gymnauchen</i>	2	DBMR085-19—DBMR086-19	USMFC (34) 00011, USMFC (34) 00014
<i>Glossogobius aureus</i>	6	DBMR087-19—DBMR092-19	USMFC (34) 00010, USMFC (34) 00012
<i>Hemigobius hoevenii</i>	1	DBMR296-19	USMFC (34) 00023
<i>Psammogobius biocellatus</i>	1	DBMR305-19	USMFC (34) 00026
<i>Pseudapocryptes elongatus</i>	2	DBMR093-19—DBMR094-19	USMFC (34) 00017
<i>Pseudogobius fulvicaudus</i>	3	DBMR297-19—DBMR299-19	USMFC (34) 00024
<i>Pseudogobius avicennia</i>	2	DBMR294-19—DBMR295-19	USMFC (34) 00022
<i>Stigmatogobius sadanundio</i>	5	DBMR095-19—DBMR096-19, DBMR291-19 -DBMR293-19	USMFC (34) 00019, USMFC (34) 00021
<i>Trypauchen vagina</i>	1	DBMR321-19	USMFC (34) 00027
<i>Trypauchen pelaeos</i>	1	DBMR097-19	USMFC (34) 00013
<i>Cryptocentrus</i> sp.	1	DBMR338-20	N/A
ATHERINIFORMES			
Phallostethidae			
<i>Neostethus lankesteri</i>	3	DBMR282-19—DBMR284-19	USMFC (108) 00001
BELONIFORMES			
Adrianichthyidae			
<i>Oryzias javanicus</i>	3	DBMR279-19—DBMR281-19	USMFC (101) 00002
Belonidae			
<i>Strongylura strongylura</i>	3	DBMR003-19—DBMR005-19	USMFC (89) 00002
Hemiramphidae			
<i>Hyporhamphus dussumieri</i>	1	DBMR219-19	USMFC (105) 00002
<i>Hyporhamphus quoyi</i>	4	DBMR006-19—DBMR008-19, DBMR324-19	USMFC (105) 00001, USMFC (105) 00005
Zenarchopteridae			
<i>Dermogenys sumatrana</i>	3	DBMR288-19—DBMR290-19	USMFC (105) 00004
CARANGIFORMES			
Carangidae			
<i>Alepes melanoptera</i>	2	DBMR043-19—DBMR044-19	USMFC (69) 00013, USMFC (69) 00014
<i>Atule mate</i>	1	DBMR045-19	USMFC (69) 00015
<i>Carangoides coeruleopinnatus</i>	3	DBMR046-19—DBMR047-19, DBMR247-19	USMFC (69) 00012
<i>Caranx ignobilis</i>	4	DBMR048-19—DBMR051-19	USMFC (69) 00019, USMFC (69) 00020
<i>Caranx sexfasciatus</i>	1	DBMR346-20	USMFC (69) 00022
<i>Megalaspis cordyla</i>	5	DBMR052-19—DBMR056-19	USMFC (69) 00009
<i>Scomberoides commersonianus</i>	1	DBMR060-19	USMFC (69) 00016
<i>Scomberoides tala</i>	1	DBMR059-19	USMFC (69) 00018
<i>Scomberoides tol</i>	2	DBMR057-19—DBMR058-19	USMFC (69) 00017
<i>Selaroides leptolepis</i>	2	DBMR328-20, DBMR330-20	USMFC (69) 00010
<i>Trachinotus blochii</i>	2	DBMR061-19—DBMR062-19	USMFC (69) 00021
<i>Ulua mentalis</i>	1	DBMR246-19	USMFC (69) 00011
Cynoglossidae			
<i>Cynoglossus arel</i>	1	DBMR268-19	USMFC (84) 00007
<i>Cynoglossus cf. cynoglossus</i>	3	DBMR187-19, DBMR325-19, DBMR334-20	USMFC (84) 00003, USMFC (84) 00006
<i>Cynoglossus monopus</i>	4	DBMR269-19—DBMR271-19, DBMR319-19	USMFC (84) 00002, USMFC (84) 00005
<i>Cynoglossus bilineatus</i>	3	DBMR188-19—DBMR190-19	USMFC (84) 00008
<i>Cynoglossus oligolepis</i>	1	DBMR320-19	USMFC (84) 00004
Paralichthyidae			
<i>Pseudorhombus arsius</i>	1	DBMR327-20	USMFC (107) 00001
Latidae			
<i>Lates calcarifer</i>	5	DBMR102-19—DBMR106-19	USMFC (76) 00001
Polynemidae			
Continued			

ORDER, Family, Species	n	BOLD ID	MUSEUM CATALOGUE NO
<i>Eleutheronema tetradactylum</i>	7	DBMR137-19—DBMR143-19	USMFC (68) 00005, USMFC (68) 00006
<i>Leptomelanosoma indicum</i>	1	DBMR144-19	USMFC (68) 00004
MUGILIFORMES			
Mugilidae			
<i>Crenimugil buehanani</i>	2	DBMR226-19, DBMR236-19	USMFC (81) 00006
<i>Crenimugil crenilabis</i>	3	DBMR225-19—DBMR228-19	USMFC (81) 00007
<i>Osteomugil perusii</i>	2	DBMR229-19—DBMR230-19	USMFC (81) 00003, USMFC (81) 00004
<i>Planiliza subviridis</i>	12	DBMR030-19—DBMR035-19, DBMR231-19—DBMR235-19, DBMR333-20	USMFC (81) 00001, USMFC (81) 00002, USMFC (81) 00005, USMFC (81) 00008
PERCIFORMES			
Gerreidae			
<i>Gerres filamentosus</i>	4	DBMR073-19—DBMR074-19, DBMR081-19, DBMR248-19	USMFC (91) 00005, USMFC (91) 00006
<i>Gerres limbatus</i>	4	DBMR077-19—DBMR080-19	USMFC (91) 00004
<i>Gerres oyena</i>	2	DBMR075-19—DBMR076-19	USMFC (91) 00003
Ambassidae			
<i>Ambassis vachellii</i>	8	DBMR238-19—DBMR245-19	USMFC (30) 00002, USMFC (30) 00003
<i>Ambassis interrupta</i>	3	DBMR040-19—DBMR042-19	USMFC (30) 00006
<i>Ambassis macracanthus</i>	5	DBMR037-19—DBMR039-19, DBMR347-20—DBMR348-20	USMFC (30) 00004, USMFC (30) 00005, USMFC (30) 00007
Haemulidae			
<i>Pomadasy kaakan</i>	4	DBMR098-19—DBMR101-19	USMFC (71) 00002, USMFC (71) 00003, USMFC (71) 00004, USMFC (71) 00005
Lethrinidae			
<i>Lethrinus lentjan</i>	3	DBMR128-19—DBMR130-19	USMFC (79) 00001
Lutjanidae			
<i>Lutjanus argentimaculatus</i>	1	DBMR136-19	USMFC (49) 00012
<i>Lutjanus johnii</i>	2	DBMR131-19—DBMR132-19	USMFC (49) 00007, USMFC (49) 00011
<i>Lutjanus russellii</i>	7	DBMR134-19—DBMR135-19, DBMR250-19 -DBMR253-19, DBMR133-19	USMFC (49) 00010, USMFC (49) 00006, USMFC (49) 00008, USMFC (49) 00009
Sciaenidae			
<i>Dendrophysa russelii</i>	1	DBMR150-19	USMFC (48) 00009
<i>Johnius</i> sp.	5	DBMR254-19—DBMR257-19, DBMR259-19	USMFC (48) 00008
<i>Nibea soldado</i>	3	DBMR151-19—DBMR153-19	USMFC (48) 00005
<i>Otolithes ruber</i>	5	DBMR154-19—DBMR158-19	USMFC (48) 00004
<i>Panna microdon</i>	1	DBMR260-19	USMFC (48) 00007
<i>Pennahia anea</i>	1	DBMR266-19	USMFC (48) 00006
<i>Pennahia ovata</i>	5	DBMR261-19—DBMR265-19	USMFC (48) 00003
Serranidae			
<i>Cephalopholis formosa</i>	1	DBMR278-19	USMFC (74) 00008
<i>Epinephelus bleekeri</i>	1	DBMR277-19	USMFC (74) 00009
<i>Epinephelus coioides</i>	5	DBMR159-19—DBMR163-19	USMFC (74) 00006
<i>Epinephelus heniochus</i>	1	DBMR276-19	USMFC (74) 00005
<i>Epinephelus sexfasciatus</i>	1	DBMR352-20	USMFC (74) 00007
Sillaginidae			
<i>Sillago sihama</i>	4	DBMR170-19—DBMR173-19	USMFC (53) 00003
Sphyraenidae			
<i>Sphyraena barracuda</i>	1	DBMR309-19	USMFC (62) 00006
<i>Sphyraena jello</i>	2	DBMR174-19, DBMR310-19	USMFC (62) 00004, USMFC (62) 00007
<i>Sphyraena zenie</i>	1	DBMR267-19	USMFC (62) 00005
Platycephalidae			
<i>Grammoplites scaber</i>	1	DBMR326-20	USMFC (56) 00006
<i>Platycephalus indicus</i>	1	DBMR191-19	USMFC (56) 00007
Tetrarogidae			
<i>Trichosomus trachinoides</i>	2	DBMR308-19, DBMR337-20	USMFC (115) 00001, USMFC (115) 00002
Cichlidae			
<i>Oreochromis mossambicus</i>	1	DBMR063-19	USMFC (47) 00005
ACANTHURIFORMES			
Drepaneidae			
Continued			

ORDER, Family, Species	n	BOLD ID	MUSEUM CATALOGUE NO
<i>Drepane punctata</i>	4	DBMR064-19, DBMR349-20—DBMR351-20	USMFC (60) 00003
Leiognathidae			
<i>Deveximentum ruconius</i>	2	DBMR124-19—DBMR125-19	USMFC (50) 00013
<i>Deveximentum indicium</i>	2	DBMR123-19, DBMR249-19	USMFC (50) 00012, USMFC (50) 00015
<i>Deveximentum hanedai</i>	1	DBMR126-19	USMFC (50) 00014
<i>Eubleekeria jonesi</i>	1	DBMR107-19	USMFC (50) 00016
<i>Leiognathus brevisrostris</i>	3	DBMR108-19—DBMR110-19	USMFC (50) 00011
<i>Leiognathus equula</i>	3	DBMR111-19—DBMR113-19	USMFC (50) 00017
<i>Nuchequula gerreoides</i>	9	DBMR114-19—DBMR122-19	USMFC (50) 00010, USMFC (50) 00018
Scatophagidae			
<i>Scatophagus argus</i>	5	DBMR145-19—DBMR149-19	USMFC (109) 00001
Siganidae			
<i>Siganus fuscescens</i>	1	DBMR164-19	USMFC (67) 00001
<i>Siganus javus</i>	5	DBMR165-19—DBMR169-19	USMFC (67) 00002
SCOMBRIFORMES			
Stromatidae			
<i>Pampus argenteus</i>	6	DBMR175-19—DBMR180-19	USMFC (110) 00001, USMFC (110) 00002
Trichiuridae			
<i>Lepturacanthus savaia</i>	3	DBMR184-19—DBMR186-19	USMFC (73) 00001
CENTRARCHIFORMES			
Terapontidae			
<i>Terapon jarbua</i>	2	DBMR182-19, DBMR307-19	USMFC (88) 00004, USMFC (88) 00006
<i>Terapon theraps</i>	2	DBMR181-19, DBMR183-19	USMFC (88) 00003, USMFC (88) 00005
TETRAODONTIFORMES			
Tetraodontidae			
<i>Arothron reticularis</i>	1	DBMR209-19	USMFC (40) 000010
<i>Dichotomyctere cf. fluviatilis</i>	1	DBMR210-19	USMFC (40) 00006
<i>Dichotomyctere nigroviridis</i>	1	DBMR303-19	USMFC (40) 00011
<i>Lagocephalus lunaris</i>	6	DBMR211-19—DBMR216-19	USMFC (40) 00004, USMFC (40) 00005, USMFC (40) 00007, USMFC (40) 00008
<i>Takifugu oblongus</i>	1	DBMR217-19	USMFC (40) 00009
Triacanthidae			
<i>Triacanthus nieuhofii</i>	1	DBMR218-19	USMFC (61) 00002, USMFC (61) 00003

Table 1. List of morphology-based species from Merbok Estuary region studied through DNA barcoding with the number of specimens examined (n), the BOLD IDs of their respective COI sequences, and the museum catalogue numbers of each species.

only identified to the generic level using morphology whereas for three other species, *Dichotomyctere cf. fluviatilis*, *Brachygobius cf. kabiliensis*, and *Cynoglossus cf. cynoglossus*, we used open nomenclature.

The most diverse orders were Perciformes (42 species representing 31.1% of the total number of species), followed by Carangiformes (21 species, 15.5%), Gobiiformes (18 species, 13.3%), and Clupeiformes (16 species, 11.9%) (Fig. 2a). At the family level, Gobiidae has the highest species richness with 15 species (11.1%), followed by Carangidae (12 species, 8.9%), Engraulidae (9 species, 6.7%), and Ariidae (8 species, 5.9%) (Fig. 2b). The three most diverse genera were the anchovy genus *Stolephorus*, the flatfish genus *Cynoglossus* with five species each, followed by the grouper genus *Epinephelus* with four species. According to the International Union for Conservation of Nature (IUCN) Red List, five species are “Near Threatened” (four Chondrichthyes: *Telatrygon zugei*, *Brevitrygon walga*, *Gymnura poecilura*, *Chiloscyllium indicum* and one actinopterygian: *Arius gadora*) whereas others are listed as “least concern” or “data deficient”. One recorded species is an alien invasive species (AIS), *Oreochromis mossambicus* (the Mozambique tilapia) from the African region³⁷.

DNA-based delimitation. Sequence length for all 350 generated barcodes was longer than 600 bp with no indels or stop codon detected. The nucleotide composition showed a mean percentage of 18.32% (G), 27.97% (C), 24.07% (A), and 30.7% (T). More than half of the species (56%, 76 species) were represented by multiple specimens while 59 species were represented by a single specimen (Table 1). Mean number of specimens per species was 2.59. Increment in the K2P genetic divergence was directly related to the hierarchical taxonomic relationship: within species mean divergence = 0.85% (SE = 0.01), within congeners mean divergence = 16.7%, (SE = 0.01) and within families mean divergence = 18.17% (SE = 0) (Table 2).

Deep intraspecific K2P divergences, exceeded the standard threshold distance of 2%^{21,38}, were observed in seven species: *Eleutheronema tetradactylum* (16.66%), *Osteomugil perusii* (14.24%), *Planiliza subviridis* (13.44%), *Deveximentum indicium* (9.05%), *Lagocephalus lunaris* (5.62%), *Gerres oyena* (4.29%) and *Lutjanus russellii*

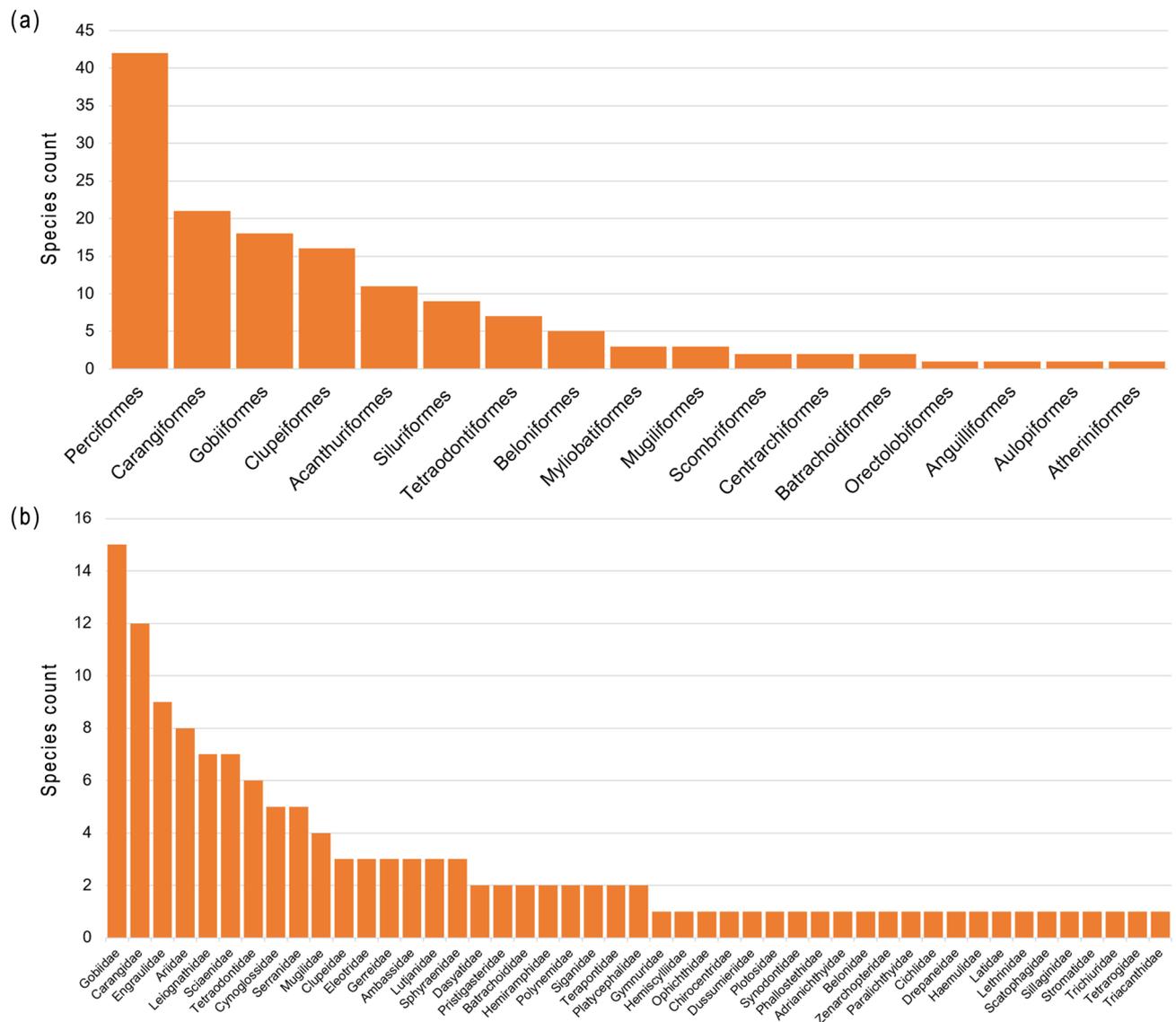


Figure 2. Species count rankings according to (a) orders and (b) families recorded in this study.

Category	n	Taxa	Comparisons	Minimum (%)	Mean (%)	Maximum (%)	SE (%)
Within species	285	75	532	0	0.85	16.66	0.01
Within genus	169	24	427	1.11	16.7	23.59	0.01
Within family	207	16	1197	0.16	18.17	26.02	0

Table 2. K2P divergence values from 350 analysed specimens with increasing taxonomic levels. *SE* standard error.

(4.12%) (Table 3). Barcoding gap analysis demonstrated that almost all species represented by multiple sequences are supported by a barcode gap (Fig. 3). Notably, only one species, *D. indicium*, had its maximum intraspecific distance (9.05%) similar to its nearest neighbour distance (9.04%).

Both Bayesian Inference (BI) (Fig. 4) and Maximum Likelihood (ML) (Figure S1) trees were fully resolved exhibiting minimal differences in topologies. Node-supports in the BI tree were overall higher than in ML tree leading us to use the BI tree to visualize our Molecular Operational Taxonomic Unit (MOTU) delimitation results (Fig. 4). The three MOTU delimitation analyses (using RESL, ABGD and GMYC methods) yielded moderately variable numbers of MOTUs, although always higher than our initial 134 morphology-based species. The RESL analysis revealed 139 MOTUs assigned to dedicated BINs. The ABGD analysis identified the same 139 MOTUs ($P=0.0010-0.0599$) within the initial partition for all substitution models (Table S2). The single-threshold GMYC analysis recognised 140 MOTUs that were taxonomically concordant with those obtained with the other two

Species/MOTUs	Max. intraspecific distance (%)	Nearest neighbour distance (%)
Species comprising two MOTUs		
<i>Deveximentum indicium</i>	9.05	9.04
BOLD:ADZ6313	0	9.06
BOLD:AAF1238	0	8.41
<i>Eleutheronema tetradactylum</i>	16.66	19.40
BOLD:AAB8457	0	14.44
BOLD:AAB8458	0	14.44
<i>Lagocephalus lunaris</i>	5.62	17.76
BOLD:AAF8798	0	5.07
BOLD:ADL4007	0.31	5.07
<i>Gerres oyena</i>	4.29	18.49
BOLD:AAC1288	0	4.15
BOLD:AAC1290	0	4.15
<i>Lutjanus russellii</i>	4.12	14.24
BOLD:AAB2905	0.31	3.69
BOLD:AAB2904	0	3.69
<i>Osteomugil perusii</i>	14.24	15.90
BOLD:AAG3686	0.00	12.75
BOLD:AAW7354	0.00	12.75
<i>Planiliza subviridis</i>	13.44	15.30
BOLD:ABU7210	0.92	11.98
BOLD:ACC0823	0	11.98
Pairs of species sharing one MOTU		
<i>Alepes melanoptera</i>	0	0.16
<i>Caranx sexfasciatus</i>	0	0.16
BOLD:AAB5775	0.16	10.45
<i>Dichotomyctere cf. fluviatilis</i>	0	1.11
<i>Dichotomyctere nigroviridis</i>	0	1.11
BOLD:AAF2344	1.02	11.67

Table 3. List of morphological species comprising two MOTUs (= BINs) or sharing one MOTU. The summary statistics include the BIN of each MOTU, their maximum intraspecific distance and distance to the nearest neighbour (i.e. minimum interspecific distance).

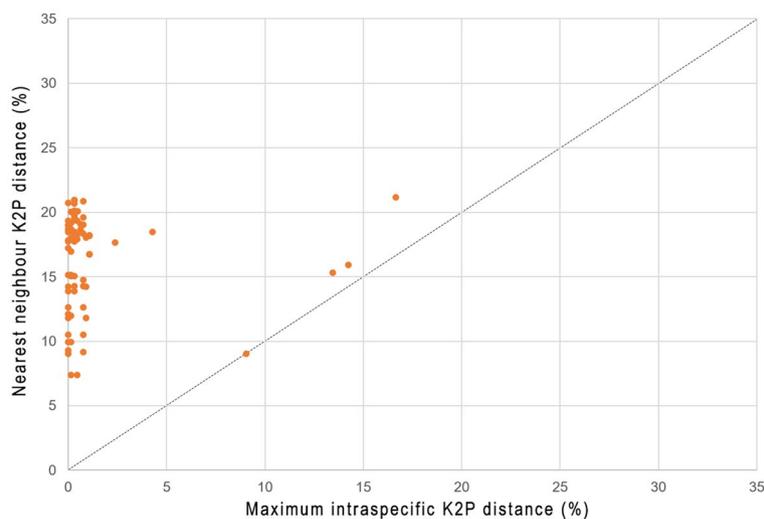


Figure 3. Scatterplot of maximum intraspecific K2P distances vs. the nearest neighbour K2P distances.

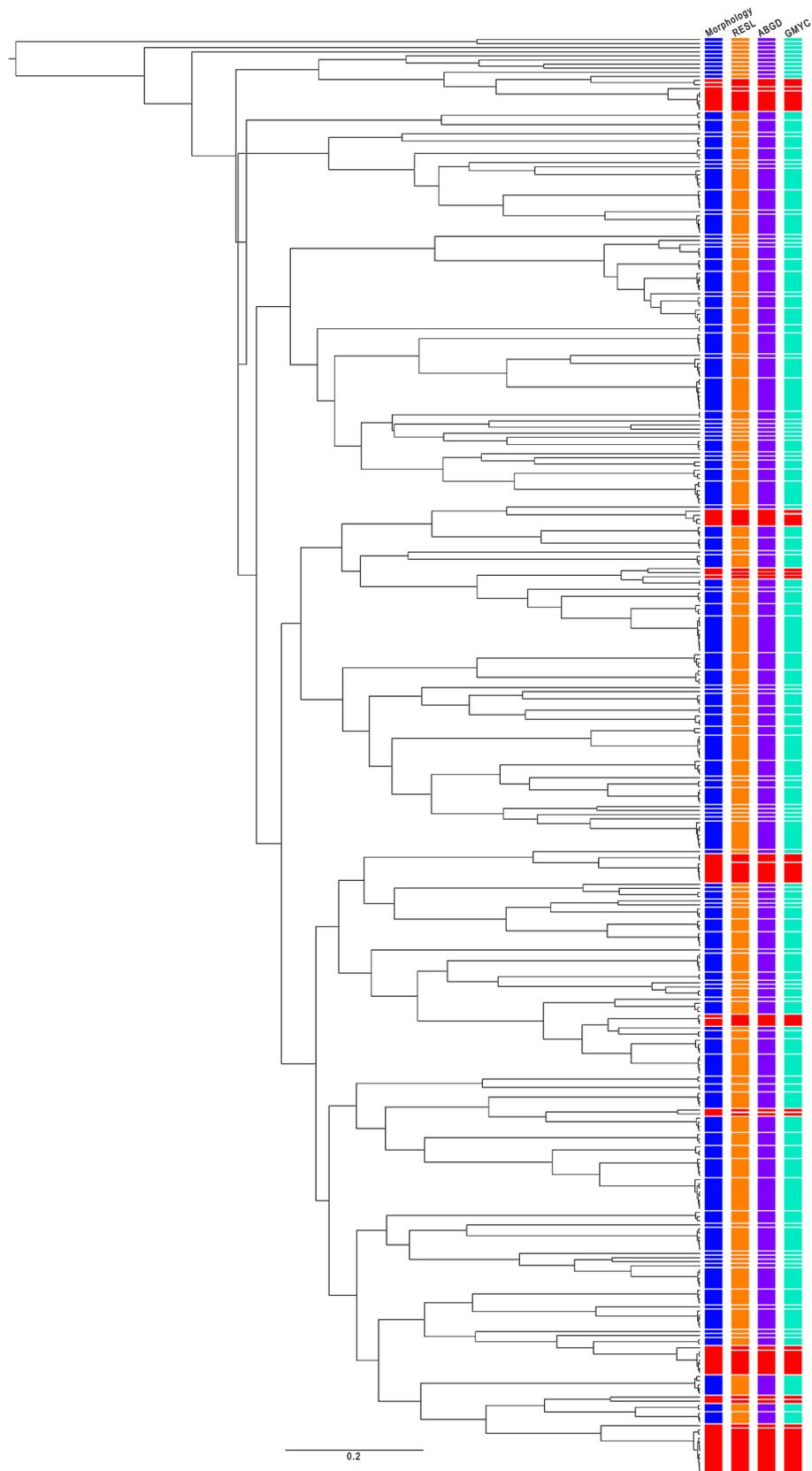


Figure 4. Bayesian Inference gene tree based on the 350 DNA barcodes with delineated MOTUs. Colour bars indicate (from left to right): morphological species (blue), MOTUs delineated by RESL (orange), ABGD (purple), and GMYC (green). Red bars indicate discrepancies among the different schemes (either morphology-genetics discrepancies or genetics-genetics discrepancies).

analyses except for one species, *Hyporhamphus quoyi*, that is partitioned into two MOTUs. All incongruences between MOTUs and morphology-based species delimitation are highlighted in Fig. 4 (red bars) and detailed in Table 3. In seven (eight with GMYC) cases, two MOTUs were delimited within one morphology-based species (see above the case of *Hyporhamphus quoyi* with GMYC). In two occasions, we found two of our morphology-based species sharing the same MOTU: *Alepes melanoptera* and *Caranx sexfasciatus* (BIN “BOLD:AAB5775”) and *Dichotomyctere nigroviridis* and *Dichotomyctere cf. fluviatilis* (BIN “BOLD:AAF2344”) (Table 3). Within each of these two species-pairs, interspecific genetic divergence was <2% resulting in the recognition of only one MOTU.

Discussion

Species delimitation. One of the premises of DNA barcoding is the detection of the so-called “barcode gap”, which can be estimated in comparing the maximum intraspecific distance with the minimum interspecific distance (also known as the nearest neighbour genetic distance)³⁹. The presence of a gap within a morphological species is good evidence for species-level cryptic diversity⁴⁰. However, the absence of gap between two morphological species is indicative either that they are different forms within one species or of shared ancestral polymorphism and/or hybridization followed by introgression between these two species. In this case, a multi-gene (i.e. genomic) approach will help to determine the reciprocal taxonomic status of the two morphological species.

Employing multiple “automatic species delimitation” methods and schemes in clustering the generated DNA barcodes provide an efficient approach in identifying putative species (= MOTUs). Even though these methods may have individual pitfalls, especially in analysing singletons, they can yield a robust outcome when combined⁴¹. Despite different analytical assumptions supporting each method, all three methods yielded similar results: RESL and ABGD analyses delimited each 139 MOTUs in our dataset whereas the GMYC analysis identified 140 MOTUs. These results demonstrate a robust pattern of MOTUs in our dataset; even the GMYC method which is known to overestimate MOTUs counts compared to other methods⁴², delimited only one additional MOTU. Because both RESL and ABGD analyses had closer correspondence to the number of species defined by morphological identification, we based our discussion on species account on these two methods.

Our results show that DNA barcoding (using COI gene) and morphology-based approach converge on the delimitation of 123 species (about 90% of the examined species) in Merbok Estuary region. DNA barcoding approach further revealed possible cryptic diversity within six species whereas it did not detect significant difference between two pairs of morphological species. Such results call for further taxonomic studies.

The mean conspecific K2P divergence (0.85%) was 20-fold lower than the mean congeneric divergence (16.7%). This increase in genetic divergence with increment in taxonomic levels is logical³⁵. However, both mean genetic estimates are higher than those previously recorded in other regions. Most molecular assessment of marine fishes displayed conspecific divergence within the range of 0.25–0.39% whereas congeneric divergence were within the range of 4.56–9.93%^{22–24,36,43}, but²⁵ found similar pattern of high average conspecific and congeneric divergence within the Indo-Pacific coral reef fishes (1.06% and 15.34%, respectively).

Taxonomic conundrum. We found that seven of our morphological species comprised two MOTUs: *Eleutheronema tetradactylum* (inter-MOTU COI-based genetic distance = 16.66%), *Osteomugil perusii* (14.24%), *Planiliza subviridis* (13.44%), *Deveximentum indicium* (9.05%), *Lagocephalus lunaris* (5.62%), *Gerres oyena* (4.29%) and *Lutjanus russellii* (4.12%). Such high intraspecific genetic divergence suggests either misidentification or the presence of morphologically cryptic species^{25,44}. The first possibility is unlikely because the morphological examination of incriminated specimens, based on existing keys, seems consistent. Therefore, such genetic variability may more likely be the signal of hidden diversity. Large genetic differentiation has been reported in *E. tetradactylum* (family Polynemidae) among allopatric populations within the Indian Ocean⁴⁵. Our results are consistent with⁴⁵, further indicating that differentiation in this lineage is not only allopatrically but, also, sympatrically distributed. Recent molecular taxonomic studies on the family Mugilidae in which are included *O. perusii* and *P. subviridis*, evidenced a very high level of cryptic diversity in the Indo-West Pacific region^{46,47}. Several mullet species (*P. subviridis* and *O. perusii* are among them) are, actually, each, a complex of several morphologically similar species for which extensive taxonomic revisions are needed. The taxonomy of *D. indicium* (family Leiognathidae) is still in flux with continual descriptions of new species in several genera, including *Deveximentum*⁴⁸. The taxonomy of the genus *Lagocephalus* is difficult and the current identification key is likely incomplete making the delimitating between species challenging. Our results indicate the presence of two sympatric species under *D. indicium* in Merbok Estuary. *Gerres oyena* (family Gerreidae) and *L. russellii* (family Lutjanidae) exhibit intraspecific differentiation of lower magnitude than those observed for the first five species discussed above, although still well above the threshold of 2%. *Lutjanus russellii* natively occurs in this region⁴⁹ but it is also farmed in Merbok estuary. Aquaculture activities regularly import non-native seeds from various sources, with no or poor records of origins. The divergence observed within this species (4.12%) could be the consequence of the presence of both native and alien (escaped from aquaculture farms) individuals in Merbok estuary¹⁵.

Two cases of shared MOTUs between species were detected involving the pairs *Alepes melanoptera* and *Caranx sexfasciatus* (BOLD:AAB5775), and *Dichotomyctere nigroviridis* and *Dichotomyctere cf. fluviatilis* (BOLD:AAF2344). The first case is striking because *A. melanoptera* and *C. sexfasciatus* are morphologically easily distinguishable (specimens are housed in the USMFC collections and available for morphological verification) and the two COI sequences (one from each of these two species) are only slightly different, which seems to exclude the possibility of a contamination. This observation warrants future investigation based on more specimens.

The second case is interesting because the marking patterns of the specimens of *D. nigroviridis* and *D. cf. fluviatilis* are distinctly different¹⁵. However, the genetic distance between these two species is only 1.1%. We hypothesize that, in this case, the COI-based genetic differentiation (<2%) between *D. nigroviridis* and *D. cf.*

fluviatilis does not reflect their actual taxonomical status. Recent hybridisation among these two closely related species and incomplete lineage sorting of a recent, on-going speciation event could account for this observation⁵⁰. Guimarães-Costa et al.⁵¹ who studied the fish diversity in the Parnaíba Delta, also suggested that the rate of molecular variation does not necessarily accompany recent (sympatric) speciation event that lead to morphological differentiation.

Towards the establishment of a comprehensive DNA barcoding library of the fish community of Merbok Estuary.

Precise identification of organisms is a prerequisite for assessing the biological and ecological status of an ecosystem. The current study illustrates yet another example of the complementarity of the morphological and molecular techniques to achieve this goal. DNA barcoding offers a quick and easy approach in aquatic diversity assessment and requires minimal expertise in conventional taxonomy^{52,53}. Comprehensive DNA barcode reference library is crucial in any biodiversity assessment for providing selective autecological and biogeographic information for comparative analysis with previous assessment. Even though DNA databases like BOLD⁵⁴ and GenBank⁵⁵ are publicly available, a localised taxon-specific reference library is synoptically important as it is easier to curate and is a more practical reference for a focused site.

Our DNA barcodes reference library associated with voucher collections previously established¹⁵ can be used for further biological evaluation and biomonitoring effort in Merbok Estuary and nearby regions. Future research endeavours to assess ecosystem health status in which a reference DNA barcoding library is needed, such as COI-based environmental DNA (eDNA) surveys or metabarcoding assays, can use this database. The barcode data generated in this study will contribute to the local as well as regional conservation efforts of fish diversity. Notwithstanding, to improve the resolution of the taxonomic coverage of the mangrove-associated of the fish community of Merbok Estuary, the number of DNA barcodes for the singleton specimens and also the not-yet examined species should be increased through more sampling and increased number of sites within the estuary and around.

Of the 134 species examined in this study, 61 species (~46%) were identified with high commercial value⁵⁶. Protection planning and proper fishery management of these species are vital. Furthermore, we manage to barcode an invasive species—the Mozambique tilapia, *Oreochromis mossambicus*; its monitoring should be conducted either using traditional methods or eDNA methods.

We DNA barcoded a rich and diverse mangrove-associated fish community. Of the 134 species initially identified based on morphology, barcodes of 123 species support their validity. We found hidden diversity within seven species whereas the divergences between two pairs of valid species are below the interspecific threshold standard calling for further taxonomic studies. The comparison with previous species lists in and around this region⁴⁹ shows that our taxonomic coverage in Merbok Estuary is certainly not complete, although the degree of incompleteness is unknown. Further researches are needed to expand the results of this study, especially towards small, elusive, transient and non-commercial fish species. The establishment of a local DNA barcoding reference library is an essential step for future studies of fisheries, conservation and ecological management of this important site.

Methods

Ethics statement. This project was conducted according to the relevant national and international guidelines and did not involve any endangered or protected fish species. All fish specimens were either collected from the local fishermen, caught using non-invasive fishing gear by the authors, or bought from the local market. This study was carried out following the recommendations and approval by the Universiti Sains Malaysia Animal Ethics Committee.

Sample collection. A total of 441 specimens were sampled between December 2018 to October 2019 at multiple locations along the Merbok Estuary and its vicinity (Fig. 1). Specimens were collected either from local fishermen (who use the barrier-net method locally called 'pompong'), direct sampling by dip-net or bought from the major fish landing site (Kuala Muda Whispering Market). All specimens were caught within Merbok River and its adjacent waters. Samples collected from the fish landing site were retrieved from fishing vessels that operate within Zone A (from the shoreline up to 5 nautical miles) and Zone B (from 5 to 12 nautical miles)⁵⁷. Information on the sampling localities (geographical coordinates) is shown in Table S1. Other collection data—dates, taxonomy and details of voucher specimens can be retrieved from the online project datasheet implemented in BOLD with project code—DBMR.

Sample processing and morphological identification. A fin clip from each fresh specimen was taken and stored in 90% ethanol. Voucher specimens were fixed in 10% formalin for at least one week and then transferred into 70% ethanol for long term storage. All specimens were catalogued and deposited at the Museum of Biodiversity, Universiti Sains Malaysia.

Morphology-based species identifications and nomenclature follow¹⁵ with few reidentifications: *Pseudogobius avicennia* (museum number: USMFC (34) 00022; identified as *Pseudogobius olorum* in¹⁵), *Trypauchen vagina* (USMFC (34) 00027; *Trypauchen pelaeos* in¹⁵), *Trypauchen pelaeos* (USMFC (34) 00013; *Trypauchen vagina* in¹⁵), *Cynoglossus bilineatus* (USMFC (84) 00008; *Cynoglossus lingua* in¹⁵), *Cynoglossus monopus* (USMFC (84) 00002, 00005; *Cynoglossus cynoglossus* in¹⁵), *Cynoglossus cf. cynoglossus* (USMFC (84) 00003, 00006; *Cynoglossus puncticeps* in¹⁵), *Pseudorhombus arsius* (USMFC (107) 00001; *Pseudorhombus elevatus* in¹⁵), *Stolephorus baganensis* (USMFC (82) 00038, 00049; *Stolephorus dubiosus* in¹⁵). We were unable to unequivocally assigned few specimens to a valid described species using available keys. In these cases, we used either “sp.” or “cf.”

We did not barcode five species listed in¹⁵: *Sardinella gibbosa*, *Zenarchopterus buffonis*, *Gerres macracanthus*, *Drepane longimana*, and *Johnius belangerii*, but we sequenced one specimen of *Cryptocentrus* sp., which was not listed in¹⁵. A total of 134 morphological species were considered in this study (Table 1).

Laboratory analyses. Genomic DNA was extracted using DNeasy Blood & Tissue kit (Qiagen, Germany) following the given protocol of animal tissue DNA extraction. The purity and concentration of the isolated DNA were measured using a microvolume UV spectrophotometer (Quawell Q300, Quawell, CA) and stored at $-20\text{ }^{\circ}\text{C}$ until further use. An approximately 650 bp fragment of the mitochondrial COI gene region was amplified using the combinations of the following primers previously designed by²²:

FishF1-5'-TCAACCAACCACAAAGACATTGGCAC-3';
 FishF2-5'-TCGACTAATCATAAAGATATCGGCAC-3';
 FishR1-5'-TAGACTTCTGGGTGGCCAAAGAATCA-3' and
 FishR2-5'-ACTTCAGGGTGACCGAAGAATCAGAA-3'.

Each sample was amplified in a final volume of 25 μL , containing 5.5 μL of 5x MyTaq™ Reaction Buffer Red (Bioline GmbH, Germany), 0.5 μL of each primer (100 ng/ μL), 0.25 μL 5U Taq polymerase (iNtRON Biotechnology Inc., Korea), 2.5 μL of genomic DNA (50 ng/ μL) and adequate nuclease-free water to complete the final reaction volume. Each amplification set was performed with the inclusion of a negative control (no template DNA) with thermal cycling conditions as follows: initial denaturation at $94\text{ }^{\circ}\text{C}$ for 4 min; followed by 35 cycles of denaturation at $94\text{ }^{\circ}\text{C}$ for 30 s, annealing at $48\text{ }^{\circ}\text{C}$ for 50 s, and extension at $72\text{ }^{\circ}\text{C}$ for 1 min; then a final extension at $72\text{ }^{\circ}\text{C}$ for 10 min. The PCR products were then fractionated by 2% gel electrophoresis to check for successful amplification. All positive amplifications were then sent for purification and sequencing to Apical Scientific Sdn. Bhd. (Selangor, Malaysia) operating the ABI PRISM 3730XL automated sequencer and the ABI PRISM BigDye terminator cycle sequencing kit v3.1 (Applied Biosystems, Foster City, CA). Bidirectional sequencing was employed to decrease the probability of sequencing errors.

Data analyses. Each generated chromatogram was manually screened prior to DNA alignment in MEGA X⁵⁸. The sequences were proofread and independently aligned and then inspected for deletions, insertions and stop codons using the same software.

A total of 350 COI sequences were determined in this study. To assess the taxon discrimination between all specimens, pairwise genetic distances were calculated within and between species, genera, and families based on the Kimura 2-parameter (K2P) distance model⁵⁹ using the analytical tools available in the BOLD system platform. To depict a graphical representation of the genetic relationships of the sequences, Bayesian Inference (BI) and Maximum Likelihood (ML) analyses were run in BEAST 2⁶⁰ and raxmlGUI 2.0⁶¹ program, respectively. The GTR+I+G substitution model was determined as the best one in PartitionFinder 2⁶², as implemented in the CIPRES portal⁶³. The BI tree was constructed with the GTR+I+G substitution model, empirical base frequencies with four gamma categories, employing a relaxed lognormal clock and the birth-death model. Two Markov Chain Monte Carlo (MCMC) chains of 40 million were run independently, sampled every 1000 generations and the first 20% were discarded as burn-in. Both run performances were then assessed for convergence (ESS > 200) using Tracer 1.7.1 and combined using LogCombiner 2.4.8 before the final tree was constructed using TreeAnnotator 2.4.7, within the BEAST 2 package⁶⁰. The ML tree was also built based on the GTR+I+G model with 1000 nonparametric bootstrap replicates. Both constructed trees were then viewed and edited in FigTree 1.4.4⁶⁴.

Three different sequence-based methods were used to delimit the Molecular Operational Taxonomic Units (MOTUs) from the analysed sequences—(1) Refined Single Linkage (RESL), (2) Automatic Barcode Gap Discovery (ABGD), and (3) Generalized Mixed Yule Coalescent (GMYC). The first analysis was done within the BOLD platform using the RESL algorithm⁶⁵ to assign sequences to a dedicated Barcode Index Numbers (BINs which are MOTUs). Next, the ABGD³⁹ analysis was run at the webserver (<https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html>) to census divergence within the analysed dataset for species delimitation. The ABGD analysis was run with the following settings: relative gap width $X=1.0$, intraspecific divergence (P) values range from 0.001 to 0.0059 for all the distance metrics, while all other parameter values were kept as default. Finally, the GMYC method⁶⁶ was employed with the fully resolved, BI ultrametric tree using only unique haplotypes (see above for the reconstruction method). The haplotype dataset used in the GMYC analysis was built in collapsing all 350 individual COI sequences into 258 unique haplotype sequences using ALTER⁶⁷. A single-threshold GMYC analysis was run in RStudio⁶⁸ with the 'splits' package⁶⁹.

Data availability

All the COI sequences determined in this study have been uploaded in BOLD⁵⁴ under the DBMR project (dx.doi.org/10.5883/DS-SRDBMR) and deposited in GenBank⁵⁵ (Accession nos. MW498499—MW498843).

Received: 24 February 2021; Accepted: 23 August 2021

Published online: 07 September 2021

References

1. Levin, L. A. *et al.* The function of marine critical transition zones and the importance of sediment biodiversity. *Ecosystems* **4**, 430–451 (2001).
2. Sarathchandra, C. *et al.* Significance of mangrove biodiversity conservation in fishery production and living conditions of coastal communities in Sri Lanka. *Diversity* **10**, 20 (2018).

3. Brown, C. J. *et al.* The assessment of fishery status depends on fish habitats. *Fish Fish.* **20**, 1–14 (2019).
4. Kathiresan, K. & Bingham, B. L. Biology of mangroves and mangrove ecosystems. *Adv. Mar. Biol.* **40**, 84–254 (2001).
5. De La Morinière, E. C., Pollux, B., Nagelkerken, I. & Van der Velde, G. Post-settlement life cycle migration patterns and habitat preference of coral reef fish that use seagrass and mangrove habitats as nurseries. *Estuar. Coast. Shelf Sci.* **55**, 309–321 (2002).
6. Asaad, I., Lundquist, C. J., Erdmann, M. V. & Costello, M. J. Delineating priority areas for marine biodiversity conservation in the Coral Triangle. *Biol. Conserv.* **222**, 198–211 (2018).
7. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853 (2000).
8. Chong, V. C., Lee, P. K. & Lau, C. M. Diversity, extinction risk and conservation of Malaysian fishes. *J. Fish Biol.* **76**, 2009–2066. <https://doi.org/10.1111/j.1095-8649.2010.02685.x> (2010).
9. Wong, S. L. *Matang Mangroves: A Century of Sustainable Management* (Sasyaz Holdings Private Ltd., Forestry Department Peninsular Malaysia, 2004).
10. Ong, J. *et al.* Hutan paya laut Merbok, Kedah: Pengurusan hutan, persekitaran fizikal dan kepelbagaian flora. In *Siri Kepelbagaian Biologi Hutan* Vol. 23 (eds Ku-Aman, K. A. *et al.*) 21–33 (Jabatan Perhutanan Semenanjung Malaysia, 2015).
11. Jamaluddin, J. A. F. *et al.* DNA barcoding of shrimps from a mangrove biodiversity hotspot. *Mitochondrial DNA Part A* **30**, 618–625. <https://doi.org/10.1080/24701394.2019.1597073> (2019).
12. Mansor, M., Mohammad-Zafrizal, M., Nur-Fadhilah, M., Khairun, Y. & Wan-Maznah, W. Temporal and spatial variations in fish assemblage structures in relation to the physicochemical parameters of the Merbok estuary, Kedah. *J. Nat. Sci. Res.* **2**, 110–127 (2012).
13. Hookham, B., Shau-Hwai, A. T., Dayrat, B. & Hintz, W. A baseline measure of tree and gastropod biodiversity in replanted and natural mangrove stands in Malaysia: Langkawi Island and Sungai Merbok. *Trop. Life Sci. Res.* **25**, 1 (2014).
14. Mansor, M., Najamuddin, A., Mohammad-Zafrizal, M., Khairun, Y. & Siti-Azizah, M. Length-weight relationships of some important estuarine fish species from Merbok estuary, Kedah. *J. Nat. Sci. Res.* **2**, 8–19 (2012).
15. Zainal Abidin, D. H. *et al.* Ichthyofauna of Sungai Merbok Mangrove Forest Reserve, northwest Peninsular Malaysia, and its adjacent marine waters. *Check List* **17**, 601–631 (2021).
16. Lim, H. C., Zainal Abidin, M., Pulungan, C. P., de Bruyn, M. & Mohd Nor, S. A. DNA barcoding reveals high cryptic diversity of the freshwater halfbeak genus *Hemirhamphodon* from Sundaland. *PLoS ONE* **11**, e0163596 (2016).
17. Mennesson, M. I., Bonillo, C., Feunteun, E. & Keith, P. Phylogeography of *Eleotris fusca* (Teleostei: Gobioidi: Eleotridae) in the Indo-Pacific area reveals a cryptic species in the Indian Ocean. *Conserv. Genet.* **19**, 1025–1038 (2018).
18. Gomes, L. C., Pessali, T. C., Sales, N. G., Pompeu, P. S. & Carvalho, D. C. Integrative taxonomy detects cryptic and overlooked fish species in a neotropical river basin. *Genetica* **143**, 581–588 (2015).
19. Iyiola, O. A. *et al.* DNA barcoding of economically important freshwater fish species from north-central Nigeria uncovers cryptic diversity. *Ecol. Evol.* **8**, 6932–6951 (2018).
20. Stern, N., Rinkevich, B. & Goren, M. Integrative approach revises the frequently misidentified species of *Sardinella* (Clupeidae) of the Indo-West Pacific Ocean. *J. Fish Biol.* **89**, 2282–2305 (2016).
21. Hebert, P. D., Ratnasingham, S. & De Waard, J. R. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **270**, S96–S99 (2003).
22. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. & Hebert, P. D. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1847–1857 (2005).
23. Xu, L. *et al.* Assessment of fish diversity in the South China Sea using DNA taxonomy. *Fish. Res.* **233**, 105771 (2020).
24. Lakra, W. *et al.* DNA barcoding Indian marine fishes. *Mol. Ecol. Resour.* **11**, 60–71 (2011).
25. Hubert, N. *et al.* Cryptic diversity in Indo-Pacific coral-reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. *PLoS ONE* **7**, e28987 (2012).
26. Adibah, A. & Darlina, M. Is there a cryptic species of the golden snapper (*Lutjanus johnii*)?. *Genet. Mol. Res.* **13**, 8094–8104 (2014).
27. Bakar, A. A. *et al.* DNA barcoding of Malaysian commercial snapper reveals an unrecognized species of the yellow-lined *Lutjanus* (Pisces: Lutjanidae). *PLoS ONE* **13**, e0202945 (2018).
28. Farhana, S. N. *et al.* Exploring hidden diversity in Southeast Asia's *Dermogenys* spp. (Beloniformes: Zenarchopteridae) through DNA barcoding. *Sci. Rep.* **8**, 1–11 (2018).
29. Jaafar, T. N. A. M., Taylor, M. I., Nor, S. A. M., de Bruyn, M. & Carvalho, G. R. DNA barcoding reveals cryptic diversity within commercially exploited Indo-Malay Carangidae (Teleostei: Perciformes). *PLoS ONE* **7**, e49623 (2012).
30. Azmir, I., Esa, Y., Amin, S., Salwany, M. & Zuraina, M. DNA barcoding analysis of larval fishes in Peninsular Malaysia. *J. Environ. Biol.* **41**, 1295–1308 (2020).
31. Chu, C. *et al.* Using DNA barcodes to aid the identification of larval fishes in tropical estuarine waters (Malacca Straits, Malaysia). *Zool. Stud.* **58**, e30 (2019).
32. Hubert, N., Delrieu-Trottin, E., Irisson, J.-O., Meyer, C. & Planes, S. Identifying coral reef fish larvae through DNA barcoding: A test case with the families Acanthuridae and Holocentridae. *Mol. Phylogenet. Evol.* **55**, 1195–1203 (2010).
33. Ko, H.-L. *et al.* Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *PLoS ONE* **8**, e53451 (2013).
34. Chin, T. C., Adibah, A., Hariz, Z. D. & Azizah, M. S. Detection of mislabelled seafood products in Malaysia by DNA barcoding: Improving transparency in food market. *Food Control* **64**, 247–256 (2016).
35. Hubert, N. *et al.* Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE* **3**, e2490 (2008).
36. Landi, M. *et al.* DNA barcoding for species assignment: The case of Mediterranean marine fishes. *PLoS ONE* **9**, e106135 (2014).
37. Russell, D., Thuesen, P. & Thomson, F. A review of the biology, ecology, distribution and control of Mozambique tilapia, *Oreochromis mossambicus* (Peters 1852) (Pisces: Cichlidae) with particular emphasis on invasive Australian populations. *Rev. Fish Biol. Fish.* **22**, 533–554 (2012).
38. Hebert, P. D., Cywinska, A. & Ball, S. L. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 313–321 (2003).
39. Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* **21**, 1864–1877 (2012).
40. Meier, R., Zhang, G. & Ali, F. The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification. *Syst. Biol.* **57**, 809–813 (2008).
41. Ortiz, D. & Francke, O. F. Two DNA barcodes and morphology for multi-method species delimitation in Bonnetina tarantulas (Araneae: Theraphosidae). *Mol. Phylogenet. Evol.* **101**, 176–193 (2016).
42. Hajibabaei, M., Singer, G. A., Hebert, P. D. & Hickey, D. A. DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* **23**, 167–172 (2007).
43. Mecklenburg, C. W., Møller, P. R. & Steinke, D. Biodiversity of arctic marine fishes: taxonomy and zoogeography. *Mar. Biodivers.* **41**, 109–140 (2011).
44. Puckridge, M., Andreakis, N., Appleyard, S. A. & Ward, R. D. Cryptic diversity in flathead fishes (Scorpaeniformes: Platycephalidae) across the Indo-West Pacific uncovered by DNA barcoding. *Mol. Ecol. Resour.* **13**, 32–42 (2013).
45. Thirumaraiselvi, R. & Thangaraj, M. Genetic diversity analysis of Indian Salmon, *Eleutheronema tetradactylum* from South Asian countries based on mitochondrial COI gene sequences. *Not. Sci. Biol.* **7**, 417–422 (2015).

46. Delrieu-Trottin, E. *et al.* Biodiversity inventory of the grey mullets (Actinopterygii: Mugilidae) of the Indo-Australian Archipelago through the iterative use of DNA-based species delimitation and specimen assignment methods. *Evol. Appl.* **13**, 1451–1467 (2020).
47. Durand, J.-D., Hubert, N., Shen, K.-N. & Borsa, P. DNA barcoding grey mullets. *Rev. Fish Biol. Fish.* **27**, 233–243 (2017).
48. Alavi-Yeganeh, M. S., Khajavi, M. & Kimura, S. A new ponyfish, *Deveximentum mekranensis* (Teleostei: Leiognathidae), from the Gulf of Oman. *Ichthyol. Res.* **68**, 437–444. <https://doi.org/10.1007/s10228-020-00794-y> (2021).
49. Carpenter, K. E. & Niem, V. *FAO Species Identification Guide for Fishery Purposes. The Living Marine Resources of the Western Central Pacific. Bony Fishes Part 4 (Labridae to Latimeriidae), Estuarine Crocodiles, Sea Turtles, Sea Snakes and Marine Mammals* Vol. 6 (FAO Library, 2001).
50. Chen, W., Ma, X., Shen, Y., Mao, Y. & He, S. The fish diversity in the upper reaches of the Salween River, Nuijiang River, revealed by DNA barcoding. *Sci. Rep.* **5**, 1–12 (2015).
51. Guimarães-Costa, A. J. *et al.* Fish diversity of the largest deltaic formation in the Americas—a description of the fish fauna of the Parnaíba Delta using DNA Barcoding. *Sci. Rep.* **9**, 1–8 (2019).
52. Hupało, K. *et al.* An urban Blitz with a twist: Rapid biodiversity assessment using aquatic environmental DNA. *Environ. DNA* **3**, 200–213 (2020).
53. Zainal Abidin, D. H. & Noor Adelyna, M. A. *Universities as Living Labs for Sustainable Development* 211–225 (Springer, 2020).
54. Ratnasingham, S. & Hebert, P. D. BOLD: The barcode of life data system. *Mol. Ecol. Notes* **7**, 355–364 (2007).
55. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **46**, D41–D47 (2018).
56. Mansor, M. I. *et al.* *Field Guide to Important Commercial Marine Fishes of the South China Sea* (SEAFDEC/MFRDMD, 1998).
57. Nuruddin, A. A. & Isa, S. M. *Trawl Fisheries in Malaysia—Issues, Challenges and Mitigating Measures* (Fisheries Research Institute, Department of Fisheries Malaysia, 2013).
58. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
59. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
60. Bouckaert, R. *et al.* BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
61. Edler, D., Klein, J., Antonelli, A. & Silvestro, D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol. Evol.* **12**, 373–377 (2021).
62. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
63. Miller, M. A., Pfeiffer, W. & Schwartz, T. In *Proceedings of the 2011 TeraGrid Conference: Extreme digital discovery* 1–8 (2011).
64. Rambaut, A. FigTree v1.4.4. Available from: <http://tree.bio.ed.ac.uk/software/figtree/> (2018).
65. Ratnasingham, S. & Hebert, P. D. A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS ONE* **8**, e66213 (2013).
66. Pons, J. *et al.* Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **55**, 595–609 (2006).
67. Glez-Pena, D., Gomez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F. & Posada, D. ALTER: Program-oriented conversion of DNA and protein alignments. *Nucleic Acids Res.* **38**, W14–W18 (2010).
68. Team, R. *RStudio: integrated development for R* (RStudio Inc., 2015).
69. Fujisawa, T. & Barraclough, T. G. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: A revised method and evaluation on simulated data sets. *Syst. Biol.* **62**, 707–724 (2013).

Acknowledgements

We are very grateful to Norli Fauzani, Jamsari Amirul Firdaus, Siti Amalia Aisyah, Siti Zuliana, and Noor Zahidah for their help in sample collection. Special thanks to all the local fishermen of Merbok Estuary for their enthusiastic support in the sampling activities. This study was funded by the Ministry of Education (Department of Higher Education) through the Fundamental Research Grant Scheme (FRGS) [203/PCGSS/6711567], Universiti Sains Malaysia's Bridging Grant [304/PCGSS/6316237], and Ministry of Water, Land and Natural Resources Grant [UMT/IMB/2018/53307].

Author contributions

Design and conception of experiments: D.H.Z.A., N.A.M.A., S.A.M.N., and S.L.; data collection and specimens processing: D.H.Z.A. and S.L.; barcoding data analyses: D.H.Z.A.; production and review of the manuscript: D.H.Z.A., S.A.M.N., S.L., M.A.R., N.A.J., and N.A.M.A.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97324-1>.

Correspondence and requests for materials should be addressed to D.H.Z.A., S.A.M.N. or N.A.M.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021