BMC Ecology and Evolution

RESEARCH

DNA barcode reference library of the fish larvae and eggs of the South China Sea: taxonomic effectiveness and geographic structure

Changping Jiang[1†], Fengming Liu[1†], Jiao Qin[2*], Nicolas Hubert[3*], Bin Kang[4], Liangliang Huang[5] and Yunrong Yan[1,6*]

Abstract

Fish early-stages constitute useful indicators of the states of marine ecosystems, as well as important fishery resources. Given the spectacular phenotypic changes during ontogeny, and the paucity of diagnostic morphological characters at the species level, the identification of fish early-stages is a challenging task. DNA barcoding, the use of the mitochondrial gene of the cytochrome c oxidase subunit I (COI) as an internal species tag, opened new perspectives for the identifications of both larval fish and fish eggs. However, the accuracy of the identifications assisted by DNA barcoding are dependent of the completeness of the DNA barcode reference libraries used to assigned unknown sequences to known species. Here, we built a DNA barcode reference library for 113 species of larval fish and 85 species of fish eggs involving the production of 741 newly generated DNA barcodes from South China Sea (63 localities). Together with 514 DNA barcodes mined from Genbank for 116 species from the South China Sea regions, a reference library including 1255 DNA barcodes for 308 species (248 locations) was assembled. The present study emphasizes the importance of integrating DNA barcoding to large scale inventories of early stages, as DNA-based species delimitation analyses delimited 305 molecular operational taxonomic units (MOTUs) and multiple cases of discordance with morphological identifications were detected. Cryptic diversity is detected with 14 species displaying two MOTUs and a total of 23 species were lumped into 11 MOTUs due to low interspecific divergence and/or mixed lineages.

Keywords Fisheries management, Species identification, Species delimitation, Species divergence, Phylogeographical patterns

†Changping Jiang and Fengming Liu contribute equally to this paper and should be considered as co-first author.

*Correspondence:
Jiao Qin
qinjiao@zjblab.com
Nicolas Hubert
nicolas.hubert@ird.fr
Yunrong Yan
yryan_gdou@163.com
[1] College of Fisheries, Guangdong Ocean University, Zhanjiang 524088, China
[2] Mangrove Conservation and Research Center, Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), Zhanjiang 524000, China
[3] Institut de Recherche pour le Développement, UMR 226 ISEM (Univ. Montpellier, CNRS, IRD), Montpellier, France
[4] College of Fishery, Ocean University of China, Qingdao, China
[5] College of Environmental Science and Engineering, Guilin University of Technology, Guilin, China
[6] Guangdong Provincial Engineering and Technology Research Center of Far Sea Fisheries Management and Fishing of South China Sea, Guangdong Ocean University, Zhanjiang, China

Jiang *et al. BMC Ecology and Evolution*   (2024) 24:132

Page 2 of 18

Introduction

Early life history stages of the majority of marine fishes occur in the upper water column, where eggs and larvae are part of the plankton before they undergo multiple metaphormoses and settle in their respective adult habitats. Large-scale surveys of the ichthyoplankton are fundamental to understand ecosystem functioning and fish population dynamics [1, 2], and to help with the sustainable management of fisheries and the design of reserve networks to preserve ocean biodiversity [3, 4]. Survey data of the ichthyoplankton could be used for monitoring spawning habitats [5], detecting changes in phenology related to human-induced perturbations [6–9], and estimating the biomass of spawning adults [10]. The greatest challenge when collating larval fish data is the difference in taxonomic resolution among studies (*i.e.* species vs genus vs family level identification), which limits the ability to make large-scale comparisons of ichthyoplankton assemblages. Identifying fish early stages is a challenging task as few ichthyologists are able to identify larvae to the species level, and considering the progressive loss of taxonomic expertise (*i.e.* taxonomic impediment), the sustainability of this expertise is compromised in a near future [11].

Empirical studies of larval ecology face a major challenge which consists in accurately identifying larvae and eggs to the species level using morphological characters. This task is particularly challenging due to the high diversity of species usually encountered in marine ichthyoplankton swarms and the dramatic phenotypic changes during the fish life cycle [12]. As such, identifying fish larvae is no easy task. Consequently, there is an urgent need for a reliable and efficient approach to achieve accurate fish larvae identification. The use of standardized molecular approaches for species identification such as DNA barcoding can greatly improve the identification of ichthyoplankton [13], and potentially reduce the reliance on taxonomic experts [14]. It has been evidenced for instance that DNA barcoding could enhance the accuracy of species-level identification of marine ichthyoplankton by 70% [15]. By design, its accuracy is highly dependent on the completeness of the DNA barcode reference libraries used to assign unknown larval fish and eggs to known species [13]. These libraries turn surveys of Molecular Operational Taxonomic Units (MOTUs) into species surveys through the assignment of species names to MOTUs [16], hence giving meaning to molecular data for ecologists, evolutionary biologists and stakeholders [17]. Numerous studies have demonstrated the advantages of integrating morphological and molecular approaches in the ichthyological exploration of the world oceans. In the Pacific Ocean for instance, the Moorea Biocode project established the initial comprehensive DNA barcode reference library of the Pacific reef fishes [18]; and its application to the identification of marine larval fish assemblages in tropical and subtropical Pacific waters [13, 14].

The South China Sea (SCS) constitutes a major marine biodiversity hotspot with over 3300 fish species reported [19, 20], which are threatened by illegal fishing, disappearance of mangroves and wastewater emissions. Besides, the degradation of coral reefs in the SCS over the past few decades has posed a serious threat to the persistence of multiple fish populations. Hence, protecting fish resources in the SCS is an urgent imperative. Larval fishes and fish eggs, despite offering limited dispersal opportunities for many sedentary marine fish species, provide vital information on reproductive biology, including spawning ground and timing, and population recruitment success rates. Nonetheless, research on larval fishes and eggs remains significantly limited in the SCS, mostly scattered in the literature and plagued by misidentifications [21–23]. This situation is likely explained by the lack of a dedicated database to provide reliable references on larval fishes and eggs in the SCS.

Here, we present the result of a large-scale effort to DNA barcode larval fishes and fish eggs of the SCS with the objective to provide open resources for automated identifications of larval fishes and eggs, and promote studies of early life stages among marine fishes of the SCS (Fig. 1). Conducted between 2022 and 2023, a total of 63 sites were inventoried across the SCS (Fig. 2). In total, 741 specimens were identified, preserved, photographed and DNA barcoded to build an atlas guide representing at least 12 orders, 60 families, 113 genus and 188 species (Fig. 3) including 113 species of larval fish (Fig. 3A1-A2) and 85 species of fish eggs (Fig. 3B1-B2). Simultaneously leveraging data from current sampling efforts (188 species) and from DNA barcodes previously published in the literature [22–25] and retrieved from the NCBI and BOLD databases (166 species), we have constructed the first comprehensive larval fish and fish eggs DNA barcode reference library for the SCS. Our library includes 1255 sequences for 20 orders, 80 families, 193 genus and 308 species, including 471 DNA barcodes for 144 species which constitutes new records (Fig. 4). This comprehensive library for larval fishes and fish eggs is poised to be used by a diversity of users with varying interests, from fundamental to applied science, including fisheries management, functional ecology, taxonomy, and conservation. Additionally, the present library reveals numerous newly detected taxa for scientific exploration, along with complete collection data and DNA barcodes that will certainly facilitate their formal description as new species. Beyond shedding a new light on the fish species diversity of the SCS, this publicly available resource is anticipated
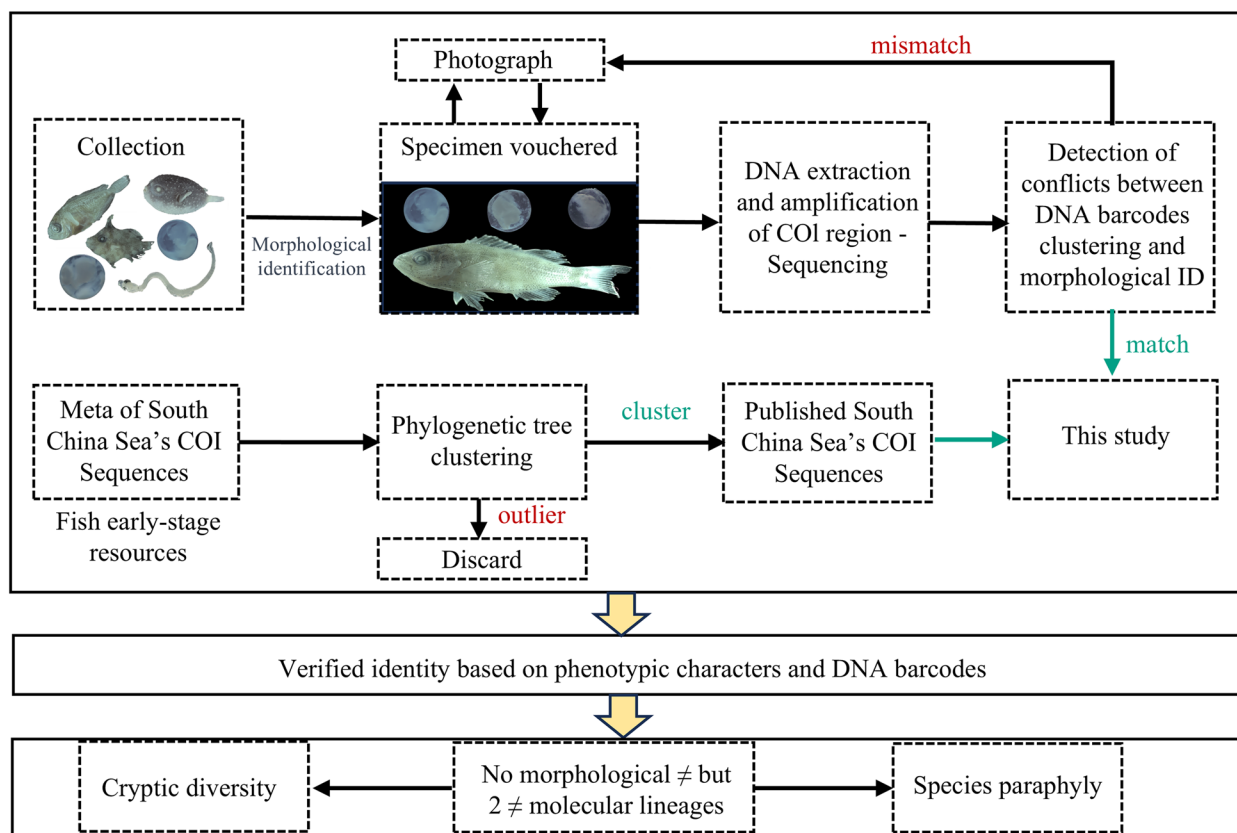
Jiang *et al. BMC Ecology and Evolution* (2024) 24:132

Page 3 of 18

Fig. 1 Overview of data generation. From specimen collection to the validation of data generation

to catalyse the development of DNA barcode reference libraries in the SCS. Furthermore, it is expected to enhance the accuracy of results for the growing number of studies utilizing DNA barcoding in the Western Pacific.

Materials and methods

Specimen collection

The collection of larval fish and fish eggs was conducted in accordance with the "Marine Survey Code" (GB12763.1-7-91), utilizing a large zooplankton net equipped with a mechanical flowmeter (HYDRO-BIOS). All the ichthyoplankton samples examined and analyzed here were newly collected during the course of the present study using zooplankton nets (80 cm diameter, 270 cm long, 505 μm mesh size, with a cod-end container mesh of 400 μm) deployed in vertical and horizontal trawls in open seas. Two simultaneous vertical and horizontal tows were used at fixed depth (15 min at 1.5–2.2 knots), using a flowmeter attached to the mouth of the net, which was used to standardise Ichthyoplankton counts to volume of water sampled. According to the specifications for oceanographic survey — Par6: marine biological survey (GB/T 12763.6—2007): the collected

samples were preserved separately in 75% ethanol–sea water solution and 5% formalin solution in seawater. These preserved samples were transported back to the laboratory for further analysis.

DNA barcode sequencing and data mining

In the laboratory, larval fishes and fish eggs from each station were examined using an Olympus SZX7 stereomicroscope, sorted, enumerated, and identified. Identification of species was done using key reference guides [26–29], larval fishes and fish eggs were then stored in ethanol for later reference, and a subset has been archived at the Guangdong Ocean University.

Each sample was first numbered, rehydrated in ultrapure water for 5–8 min for cleaning, and then photographed using a Zeiss microscope (Stemi 508) (Fig. 3). Total genomic DNA was extracted using an Aagen DNA Extraction Kit (Aagen, Guangdong, China) according to the manufacturer specifications. A partial fragment of the 5'-end of mitochondrial cytochrome c oxidase I gene (COI) of ~650 bp was amplified with the universal primers FishF$_1$(TCAACCAACCACAAAGACATTGGCAC) and FishR$_1$(TAGACTTCTGGGTGGCCAAAGAATCA) [30]. The polymerase chain reaction (PCR) employed

Jiang *et al. BMC Ecology and Evolution*   (2024) 24:132
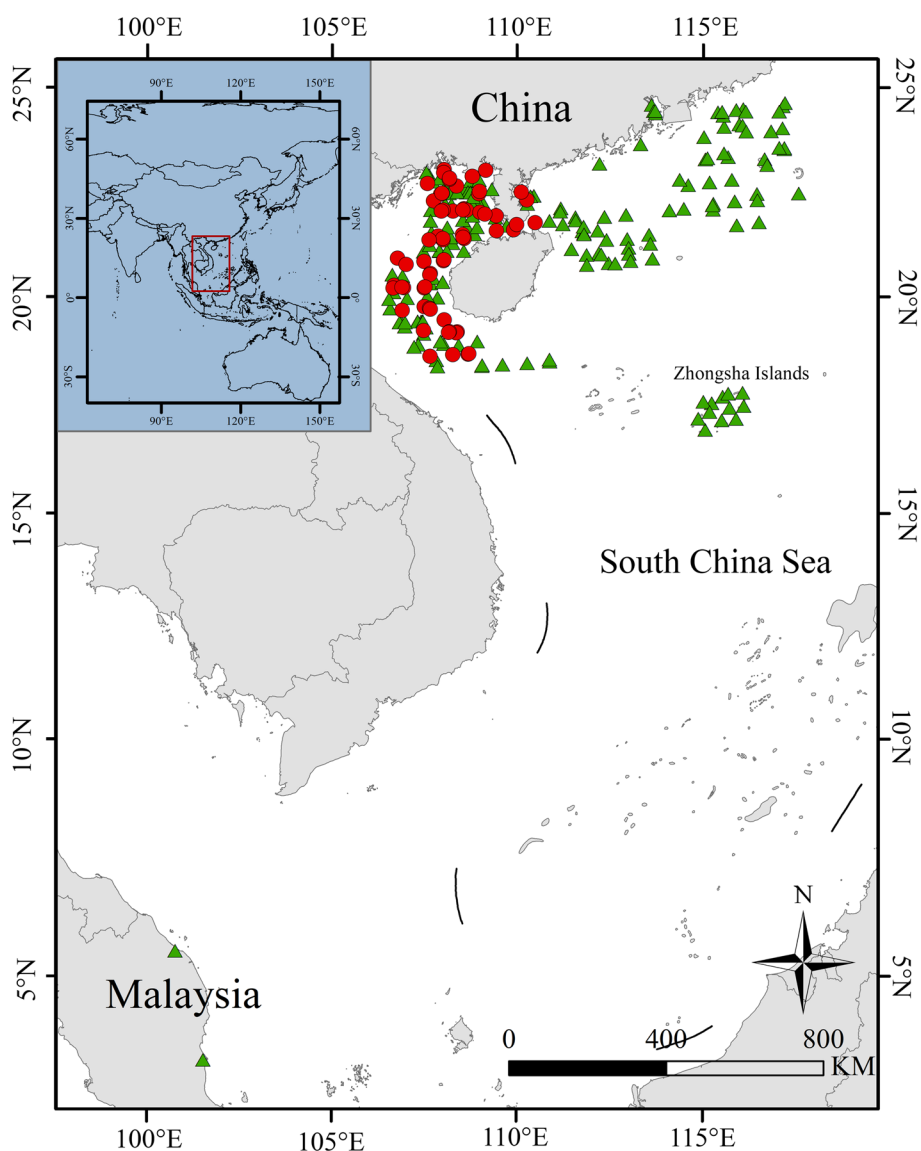
Page 4 of 18

Fig. 2 Map of sampling localities. Circles denote locations where zooplankton nets were used for capture, while triangles represent sampling sites with published sequences

the following thermocycling regime: 92°C for 5 min, 35 cycles at 92°C for 45 s, 49°C for 60 s, 72°C for 60 s, and a final extension at 72°C for 10 min. PCR amplifications were performed in a final volume of 25 μl containing 6.5 μL of ultrapure water, 2 μL of MgCl$_2$ (5 mM), 1 μL of forward and reverse primers, 12.5 μL of Taq polymerase and 2 μL of genomic DNA. Sanger sequencing was conducted at the Sangon Biotech company. Sequences were edited using Sequencher_5.4.5 (Gene Codes), aligned with MAFFT [31] and submitted to Genbank (accession numbers: PP354153-PP354861 and PP354116-PP354147; Table S1).

The taxonomic coverage of the DNA barcode reference library was further expanded by mining COI sequences for missing taxa in international repositories such as GenBank and BOLD. We employed two retrieval strategies: (1) In GenBank, we searched and downloaded COI sequences using a combination of keywords including 'South China Sea', 'fish larvae', 'fish eggs' and 'fish larvae and eggs'; (2) We searched for studies in Web of Science matching the aforementioned keywords to identify and review all relevant publications, and compiled all available COI sequences.
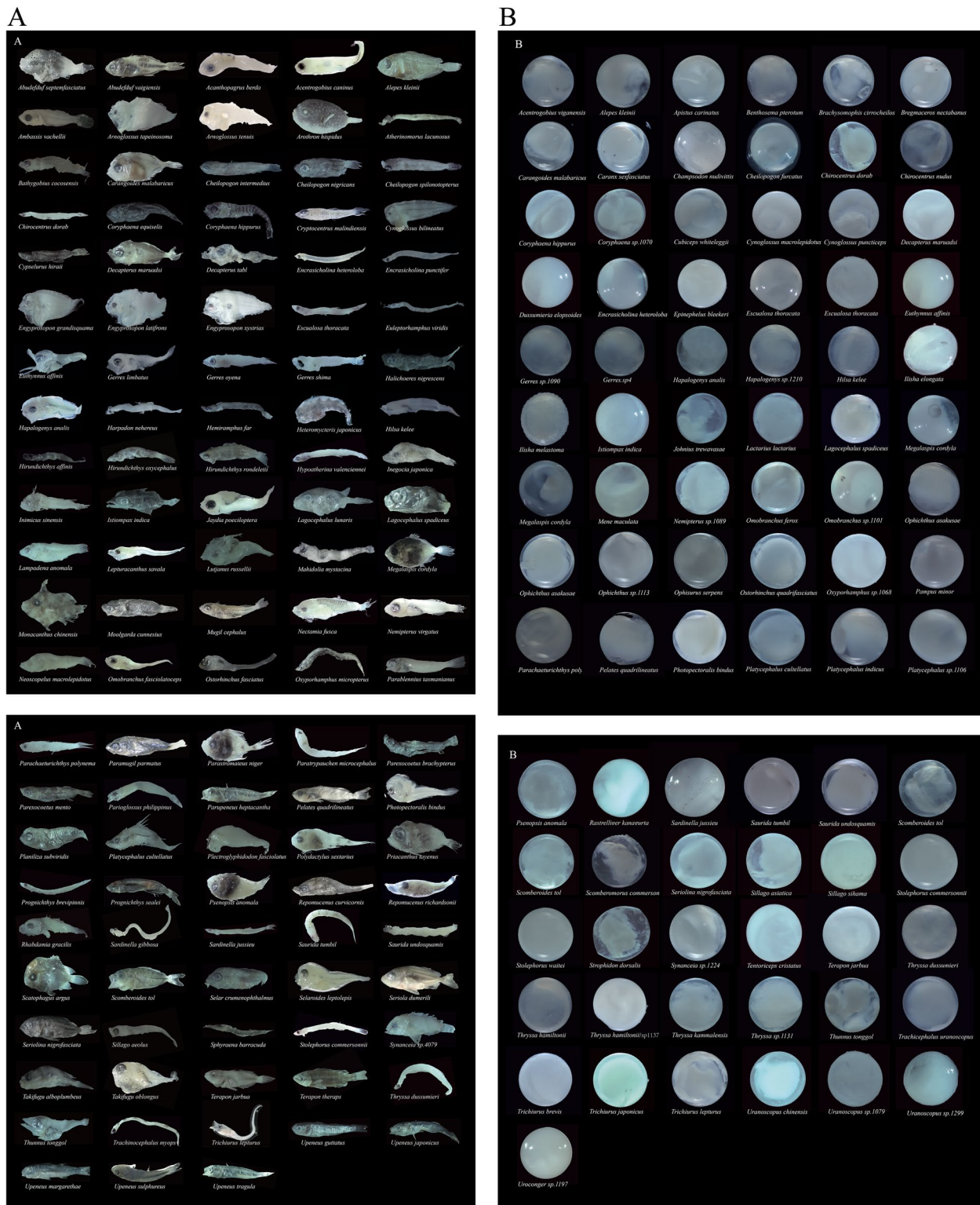
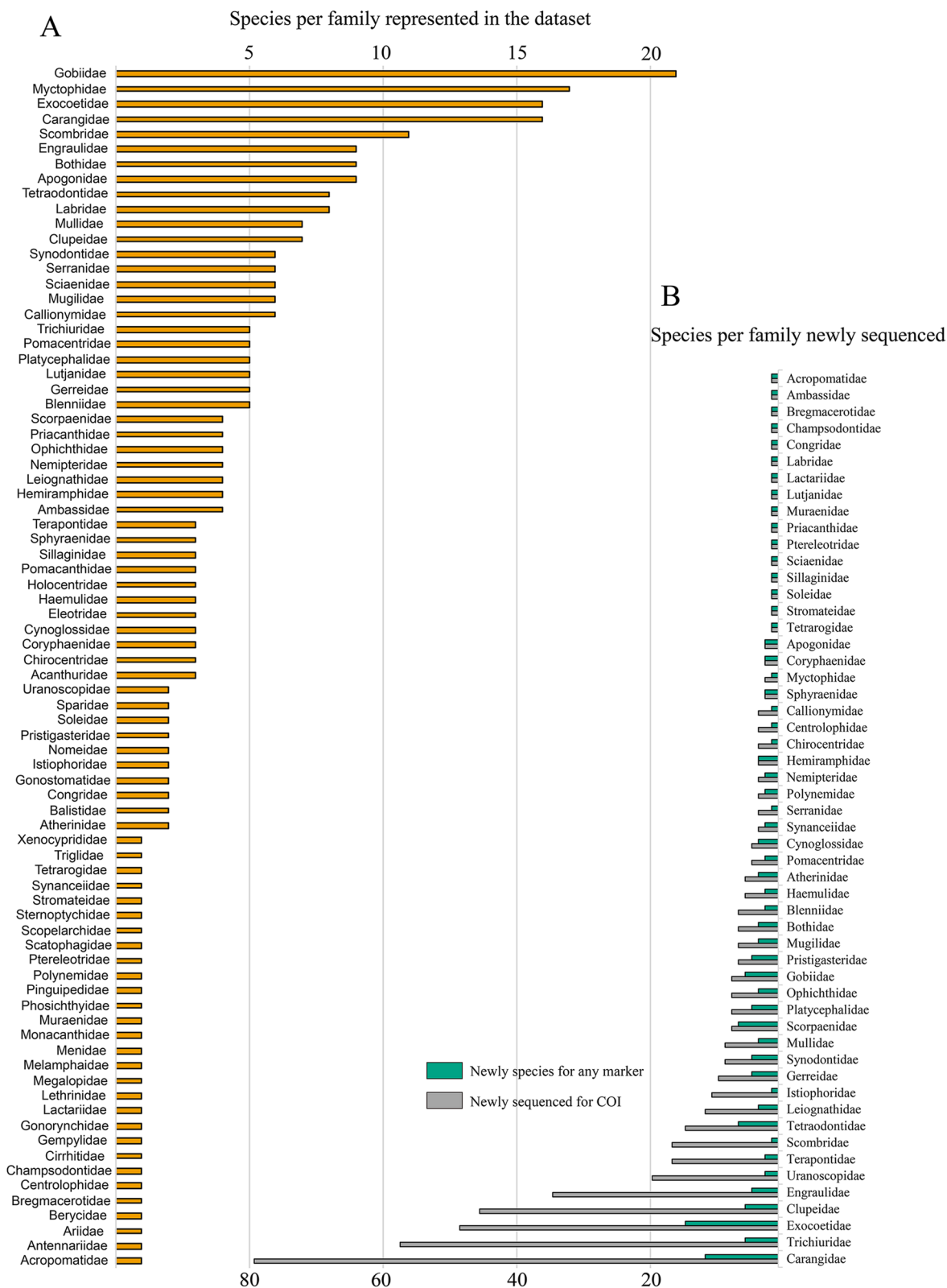**Fig. 3** Photographs of larval fish and fish eggs. **A** larval fish; **B** fish eggs

Fig. 4 Species diversity included in the Biodiversity of South China Sea dataset. **A** Number of species by family; **B** Number of newly sequenced species in each family in present dataset, New species records for any mitochondrial gene (green bars); New species records for COI marker in the South China Sea(grey bars)

Genetic species delimitation

Several methods have been suggested for delineating species based on DNA sequences [17, 32, 33]. Each of these methods possesses distinct properties, especially in handling singletons (i.e., delimited lineages represented by a single sequence) or heterogeneous speciation rates among lineages [34]. A combination of different approaches is increasingly used to overcome potential pitfalls arising from uneven sampling [35–38]. We used six different sequence-based methods of species delimitation to identify Molecular Operational Taxonomic Units (MOTUs): (i) Refined single linkage (RESL) as implemented in BOLD and used to generate Barcode index numbers (BIN) [17], (ii) Assemble Species by Automatic Partitioning (ASAP) [33], (iii) Poisson tree process (PTP) in its single (sPTP) and multiple rates version (mPTP) as implemented in the stand-alone software mPTP_0.2.3 [32] and (iv) general mixed yule-coalescent (GMYC) in its simple (sGMYC) and multiple rate version (mGMYC) as implemented in the R package Splits 1.019) [39]. The final delimitation scheme was established by deriving a majority-rule consensus from the six delimitation analyses conducted.

Both Refined Single Linkage (RESL) and ASAP utilize DNA alignments as input, with sequences being submitted to the BOLD and ASAP web servers respectively, for delimitation analysis. For PTP analyses, a Maximum Likelihood (ML) tree was generated using the R package Phangorn 2.8.1 [40] with a GTR + F + R10 substitution model. Subsequently, the ultrametric and fully resolved tree required by GMYC analyses was reconstructed using the Bayesian approach implemented in BEAST 2.4.8 [41]. Two Markov chains of 50 million each were run independently employing a Yule pure birth model tree prior, a strict-clock of 1.2% of genetic distance per million years [42], and a GTR + F + R10 substitution model. Trees were sampled every 10,000 states after an initial burn-in period of 10 million. Both runs were first checked for statistical robustness (ESS > 200) using Tracer 1.7.1 and further combined using LogCombiner 2.4.8. The maximum credibility tree was established using TreeAnnotator 2.4.7 [41]. Sequences were collapsed into haplotypes prior to Bayesian analyses using the ALTER webserver (http://www.sing-group.org/ALTER/) [43]. Subsequently, we used the R package P2C2M.GMYC 1.0 (https://github.com/P2C2M) to assess the fit of the GMYC model to our constructed dataset.

The taxonomic coverage of present sampling was further examined at both species and MOTUs levels by generating a sequence accumulation curve using the R package iNEXT [44].

Specimen identification and genetic distances

Morphological identifications were refined by comparisons to sequence-based identifications performed with the blast engines of NCBI nucleotide database and the Barcode of life datasystem (BOLD). Blast results were collected for both best match and interspecific best match following Hubert et al. [13]. Specimen identifications to the species level were collected when the best match was above, and the interspecific best match below, a similarity threshold of 98%.

Kimura 2-parameter (K2P) [45] pairwise genetic distances were calculated utilizing the R package Ape 4.1 [46]. Maximum intraspecific and nearest-neighbor genetic distances were calculated using the matrix of pairwise K2P genetic distances and the R package Spider 1.5 [47]. We checked for the presence of a barcode gap, *i.e.* the lack of overlap between the distribution of the maximum intraspecific and nearest neighbour genetic distances, by plotting both distances and examining their relationships on an individual basis rather than comparing both distributions independently [48]. The barcode gap was examined for both species and MOTUs, and a neighbour-joining (NJ) tree, constructed based on K2P distances, was also generated for a visual inspection of genetic distances and DNA barcode clusters. We used the K2P distance in all distance-related metrics in order to account for biased transition/transversion ratios and make our study comparable with similar metrics in the literature as K2P is a widely used model to compute genetic distances in DNA barcoding studies.

Finally, haplotype networks were reconstructed using the statistical parsimony network approach to visually examine cases of closely related species exhibiting haplotype sharing and/or mixed genealogies with the software Network 4.6 [49]. The haplotype networks were reconstructed including newly generated sequences as well as additional sequences mined from public database.

Results

A total of 741 COI sequences were produced based on the samples collected from the 63 sites visited in the SCS between 2022 and 2023 (Table S1; Fig. 2). All sequences were above 600 bp long and had no stop codons or insertions/deletions, indicating that the collected sequences represent functional coding regions. Alongside newly generated DNA barcodes, 514 sequences belonging to 116 species, 129 genera, 63 families and 19 orders within 185 collection sites were mined from GenBank and BOLD. After aligning newly generated and mined sequences, the final alignment consisted of 1255 sequences of 648 bp from 248 sites in the SCS.

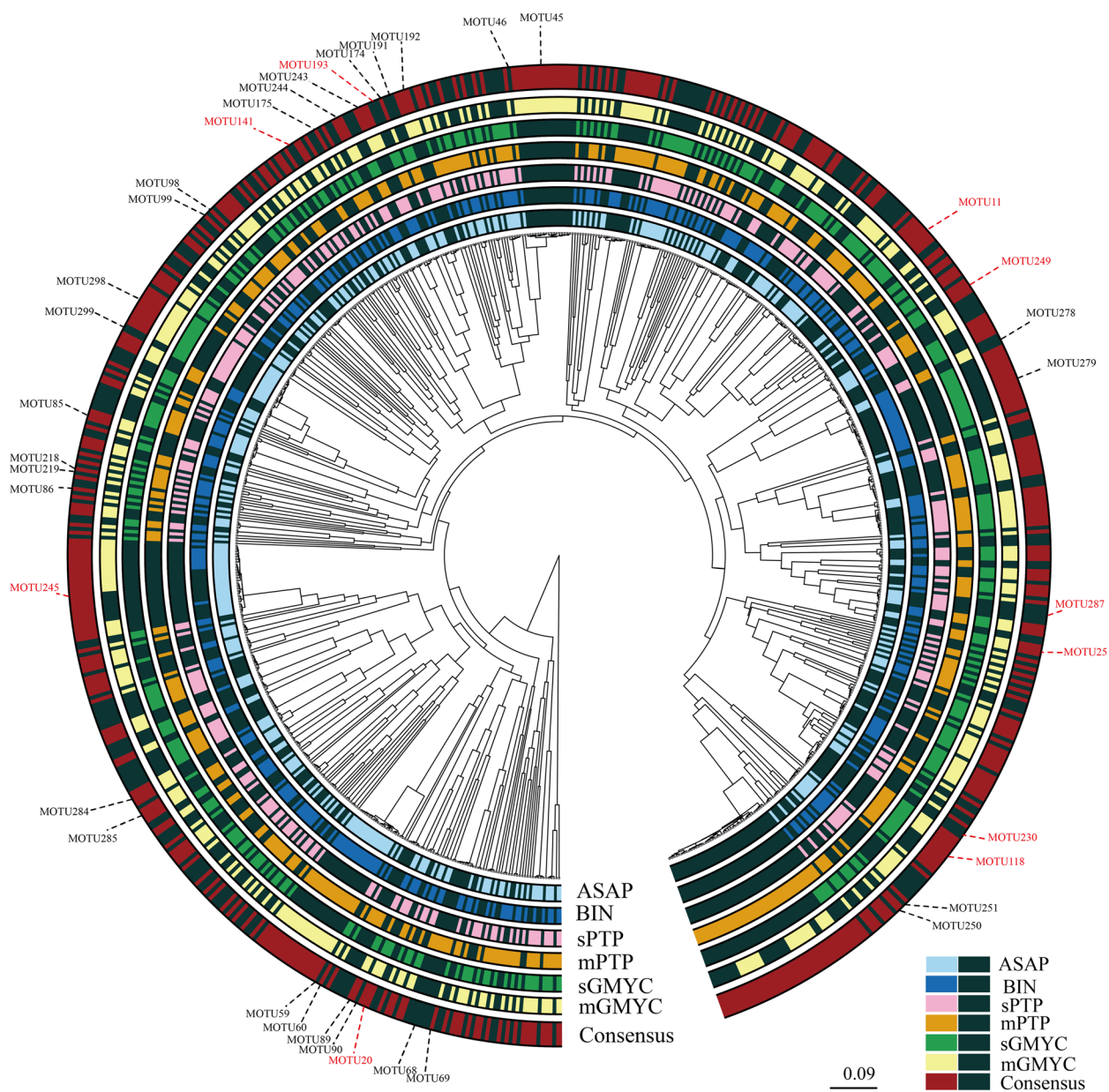MOTU delimitation analyses yielded varying numbers of MOTUs according to methods with 302, 287, 308,197,

Fig. 5 Bayesian chronograms based on a 1.2% of genetic divergence per million years including DNA-based species delimitation derived from ASAP, BIN, sPTP, mPTP, sGMYC, mGMYC and final delimitation schemes based on majority rule consensus among the six methods for larval fish and fish eggs

304 and 316 MOTUs delimited by BIN, ASAP, sPTP, mPTP, sGMYC and mGMYC respectively. The results of the P2C2M.GMYC analysis indicate that our dataset did not violate GMYC assumptions ($p = 0.7$), implying that the GMYC model is applicable to our dataset. The final consensus, based on a majority rule, consisted of 305 MOTUs (Fig. 5). BLAST analyses in NCBI and BOLD were congruent and a total of 131 sequences could be unambiguously identified to the species level using a 98%

similarity threshold (Table S2; Fig. 6, case I), corresponding to 58 species. A total of 583 sequences presented could be identified to the species level by the 98% similarity threshold, however, the best interspecific match was below 98% similarity as well (Fig. 6, case II). Finally, 27 sequences could not be identified to the species level.

When combining morphological and molecular identifications, a total of 188 species belonging to 113 genera, 60 families and 12 orders were identified among the 741
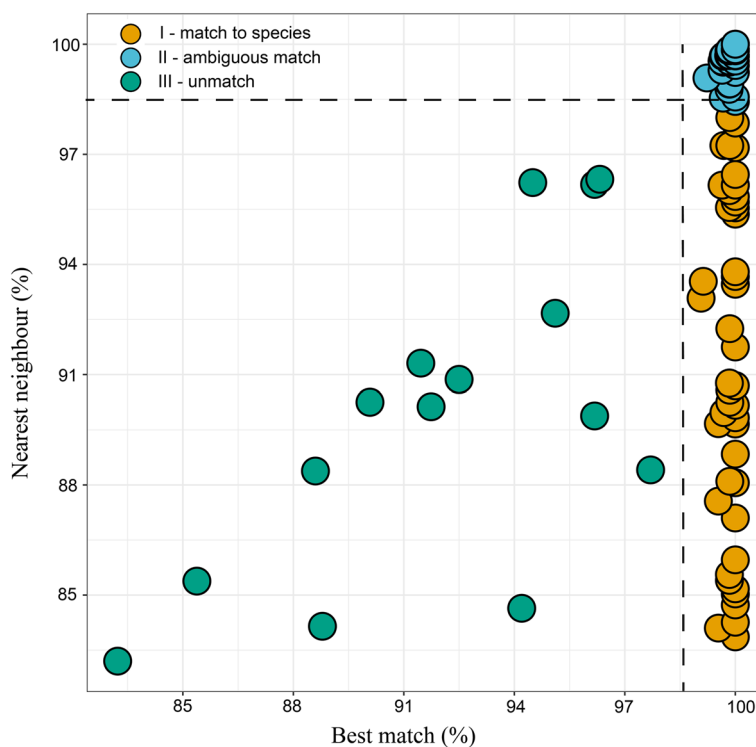
Fig. 6 Identifications using the BOLD database; best match compared with nearest neighbour (similarity percentage) for each specimen in BOLD

newly generated DNA barcodes (Table S1, Fig. 4). Among these 188 species, 121 were represented by larval fish samples and 99 species by fish eggs samples. Together with the 514 DNA barcodes mined from Genbank and BOLD, the 1255 DNA barcodes collected here belong to 308 species, 193 genera,80 families and 20 orders (Table 1). The number of sequences per species varied from 1 to 72, with an average of 4 sequences per species. Mean genetic divergence was 17.68% (0% −34.03%) within family and 8.50% (0% − 25.57%) between genera within family (Table 2). Maximum intra specific distances ranged between 0% and 21.40%, and nearest neighbour distances ranged from 0 to 23.55% (Table 2). Nearest neighbour distance was 36.71-fold higher than maximum intraspecific distance on average, with an index ratio ranging between 0 and 122.94.

Plotting maximum intraspecific and nearest neighbor K2P genetic distances revealed the absence of a barcode gap, as maximum intraspecific genetic distances surpassed distances to the nearest neighbour in several cases (Fig. 7). However, a barcode gap was observed for MOTUs (Fig. 7A, B, D). A total of 9 species displayed lower nearest neighbour K2P distance than their maximum intraspecific distances (*Thunnus albacares, Ostorhinchus kiensis, Hypoatherina valenciennei, Arnoglossus polyspilus, Coryphaena hippurus, Sardinella*

jussieu, Thryssa kammalensis, Nectamia fusca, Rhabdamia gracilis) and nearest neighbour K2P distances below 1 per cent of pairwise distance were observed for 27 species with 1 species displaying a K2P genetic distance of 0 to their nearest phylogenetic relative (*Thryssa hamiltonii*). In addition, 11 species displayed maximum intraspecific distances above 2%, including *Ostorhinchus fasciatus* (3.22%), *Ceratoscopelus warmingii* (3.97%), *Scatophagus argus* (6.68%), *Terapon jarbua* (7.59%), *Platycephalus indicus* (8.10%), *Engyprosopon latifrons* (8.42%), *Cynoglossus macrolepidotus* (14.33%), *Eleotris oxycephala* (19.03%), *Thryssa kammalensis* (20.65%), *Nectamia fusca* (20.90%) and *Rhabdamia gracilis* (21.40%) (Table S3).

Upon visually examining the NJ tree constructed using K2P genetic distances, it appears that several discrepancies between species and MOTUs are detected with 14 species displaying multiple MOTUs (Table 3) and 23 species displaying mixed genealogies within 11 MOTUs shared by more than one species (Table 4; Fig. 5). Accumulation curves indicate that the newly generated set and the entire dataset are far from reaching a plateau for larval fish and fish eggs, suggesting that the number of species recovered in this study underestimates the true early-stage resources diversity in the South China Sea (Fig. 8).

Jiang *et al. BMC Ecology and Evolution* (2024) 24:132

Page 10 of 18

Table 1 Taxonomic coverage across orders, families, genera and species of the present study

Order	Family	Genus	Species	No
Anguilliformes	3	6	7	12
Atheriniformes	1	2	2	18
Aulopiformes	2	5	7	31
Beloniformes	2	10	20	55
Beryciformes	2	3	4	6
Clupeiformes	4	10	20	175
Cypriniformes	1	1	1	2
Elopiformes	1	1	1	1
Gadiformes	1	1	1	1
Gonorynchiformes	1	1	1	1
Lamniformes	1	1	1	1
Lophiiformes	1	1	1	1
Mugiliformes	1	3	5	83
Myctophiformes	1	11	17	49
Perciformes	44	114	178	719
Pleuronectiformes	3	6	14	43
Scorpaeniformes	4	7	12	31
Stephanoberyciformes	1	1	1	1
Stomiiformes	3	3	4	4
Tetraodontiformes	3	6	11	21
Total	80	193	308	1255

Table 2 Summary statistics of pairwise K2P genetic distances among sequences at the species, genus, and family level

Level	N taxa	Min dist	Mean dist	Max dist
Within species	308	0	0.0127	0.2140
Within genus	193	0	0.0850	0.2557
Within family	80	0	0.1768	0.3403

Haplotype networks were reconstructed for 23 species displaying shallow genetic divergence and/or haplotype sharing. Scattered haplotypes across haplotype networks were observed for the MOTU including *Alepes djedaba*, *Alepes kleinii* and *Selaroides leptolepis* (MOTU11), the MOTU including *Hirundichthys affinis* and *Hirundichthys oxycephalus* (MOTU118), the MOTU including *Prognichthys brevipinnis* and *Prognichthys sealei* (MOTU230), the MOTU including S*ardinella gibbose* and *Sardinella jussieu* (MOTU245), the MOTU including *Saurida tumbil* and *Saurida undosquamis* (MOTU249), and the MOTU including *Thunnus albacares* and *Thunnus tonggol* (MOTU287). A single case of haplotype sharing was observed between *Thryssa hamiltonii* and *Thryssa kammalensis* (MOTU284), with a haplotype placed in central position in haplotype network (Fig. 9).

The maximum credibility tree reconstructed with BEAST 2.4.8 for the 14 species displaying multiple and deeply diverging MOTUs revealed a diversity of the phylogeographic patterns (Fig. 10). Two lineages were detected in *Ceratoscopelus warmingii*, *Coryphaena equiselis*, *Cynoglossus macrolepidotus*, *Eleotris oxycephala*, *Engyprosopon latifrons*, *Eviota shimadai*, *Nectamia fusca*, *Ostorhinchus fasciatus*, *Platycephalus indicus*, *Rhabdamia gracilis*, *Scatophagus argus*, *Terapon jarbua*, *Thryssa kammalensis* and *Upeneus japonicus*. In most of the cases where two MOTUs are detected within species, a South–North or an East–West differentiation was observed, as exemplified by *S. argus*, *E. oxycephala* and *T. kammalensis* (South–North) (Fig. 10E, J, N), as well as *E. latifrons*, *C. equiselis* and *R. gracilis* (East–West) (Fig. 10B, C, D). In *E. latifrons* and *C. equiselis*, one lineage was distributed in the Northwest SCS including the Beibu Guif, and Near Hainan Island, and another lineage is occurring largely in Zhongsha Islands or Near Hainan Island (Fig. 10B, C). In *R. gracilis*, the two lineages are located in the Beibu Gulf and the Pearl River estuary, respectively (Fig. 10D), the two lineages of *S. argus*, *E. oxycephala* and *T. kammalensis* exhibit distinct geographic isolation along the Sunda Shelf (Fig. 10E, J, N). However, the co-occurrence of haplotypes from distinct MOTUs is observed in some regions of the Beibu Gulf and near the Hainan Island. Alternatively, allopatric distributions of conspecific MOTUs involved different patterns with: (1) the Beibu Gulf vs. Pearl River and Paracel Islands for the MOTUs within *R. gracilis* and *E. latifrons* (Fig. 10D, B), (2) Malaysia vs. the Beibu Gulf and Pearl River for *S. argus*, *E. oxycephala* and *T. kammalensis* (Fig. 10E, J, N). The estimated divergence time suggest that most MOTU divergence events originated during the Pliocene, but a few noticeable exceptions are detected within *C. macrolepidotus*, *E. latifrons*, *P. indicus* and *U. japonicus* with MOTU divergence happening before the Pliocene (Fig. 10).

Discussion

The present study provides the first comprehensive assessment of DNA barcoding for the identification of larval fish and fish eggs assemblages of the SCS, including individual photographs of the 113 species of larval fish and the 85 species of fish eggs sampled here, as well as a DNA barcode reference library for 308 species. This study also provides a compilation of published larval fish and fish egg COI sequences from the SCS [22–25], providing the largest DNA barcode reference library published so far for SCS fish early stages. Besides, major regions of the SCS were covered including the Beibu Gulf, Near Hainan Island, Near, Qiongzhou Strait, Guangdong coastal, Pearl River, Zhongsha Islands and Malaysia, which allows
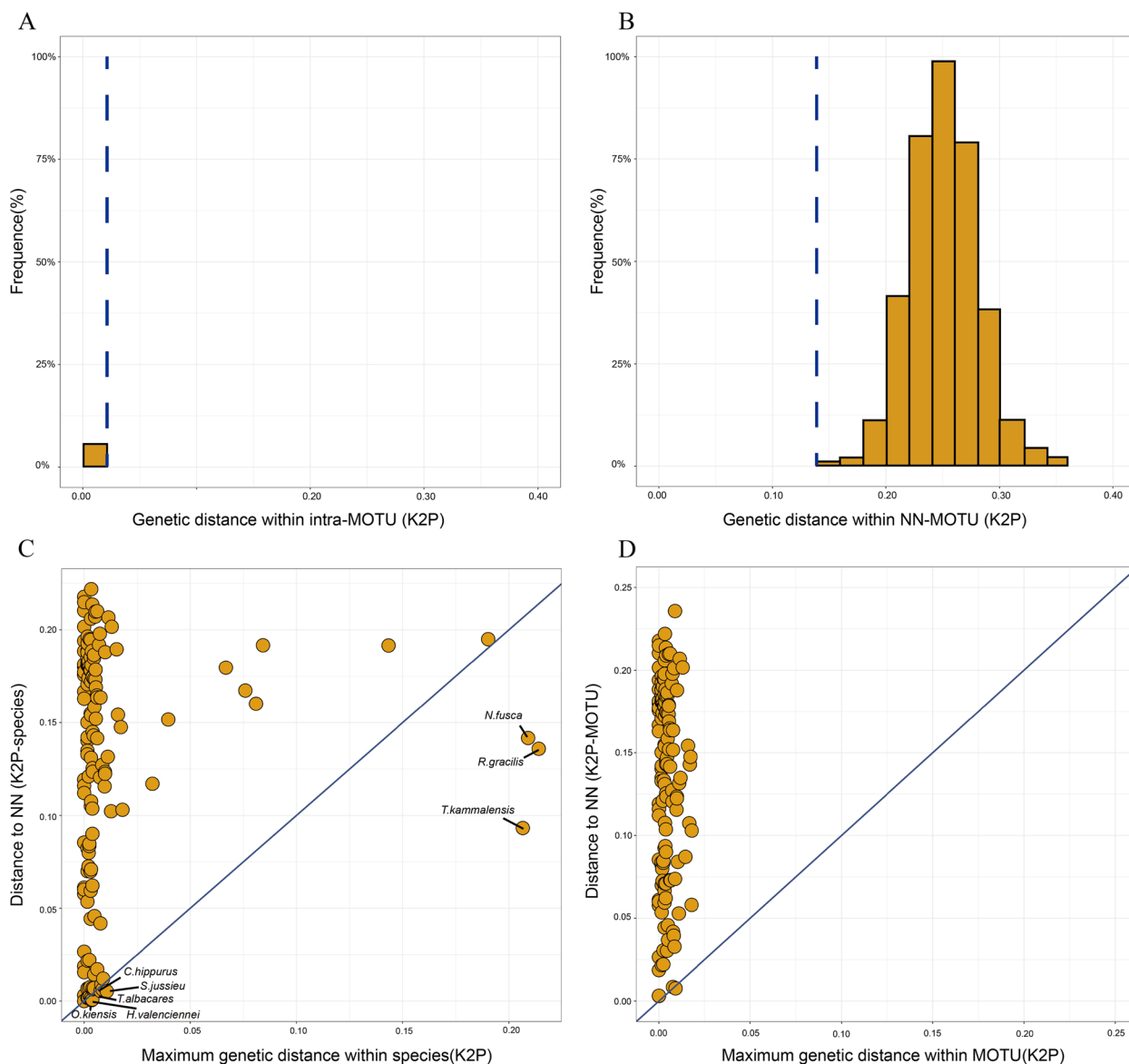
Fig. 7 Distribution of K2P genetic distances. **A** distribution of the maximum K2P genetic distances within MOTUs; **B** distribution of the minimum K2P genetic distances to the nearest MOTU. The dashed line highlights the 'barcoding gap' between the distributions of maximum intra-MOTU and maximum inter-MOTUs distances; **C** relationships between the maximum intraspecific and nearest-neighbour (NN) K2P genetic distances, **D** relationships between the maximum intra-MOTU and nearest-neighbour (NN) for MOTU K2P genetic distances

examining the impact of geographic structure on the performance of DNA barcoding.

Accuracy of DNA barcoding for the SCS larval fish and fish eggs

Since the inception of DNA barcoding by Hebert et al. [50], it has been increasingly used as a standardize molecular methods of species identification and numerous studies have demonstrated how DNA barcoding can help accelerate the pace of species discovery [13, 35–38,

51]. Our study confirms the benefits of integrating DNA barcoding into the taxonomic workflow of a biodiversity inventory in species-rich, yet complex biotas. In the case of complex species assemblages where identifications are challenging and taxonomic controversies are present, a combination with detailed morphological comparisons is necessary [52]. Besides, the accuracy of a DNA barcode reference library is also tightly dependent of the accuracy of morphological identifications [16]. Here, we ensured the accuracy of our library by including four steps during

Jiang *et al. BMC Ecology and Evolution*   (2024) 24:132

Page 12 of 18

Table 3 List of potential cryptic species including their barcode index number in BOLD, maximum K2P genetic distances within a MOTU (D_M), and K2P distance to their nearest MOTU (D_N). The nearest-neighbour distance for each MOTU corresponds to the distance to the closest MOTU in this study

	D_M	D_N
Ceratoscopelus warmingii	0.0397	0.1517
MOTU45(BIN:AAC3130)	0.0081	0.0395
MOTU46(BIN:AAT9399)	–	0.0363
Coryphaena equiselis	0.0032	0.1051
MOTU59(BIN:AAY5000)	0.0000	0.0031
MOTU60(BIN:ACK8197)	0.0000	0.0032
Cynoglossus macrolepidotus	0.1433	0.1916
MOTU68(BIN:AAC4819)	0.0040	0.1429
MOTU69(BIN:ACB6037)	–	0.1405
Eleotris oxycephala	0.1903	0.1950
MOTU85(BIN:ADC9175)	0.0022	0.1902
MOTU86(BIN:ADB3148)	–	0.1884
Engyprosopon latifrons	0.0842	0.1917
MOTU89(BIN:AET9781)	0.0015	0.0839
MOTU90(BIN:ADL7646)	–	0.0830
Eviota shimadai	0.0155	0.1936
MOTU98(BIN:ACC0615)	–	0.0155
MOTU99(BIN:ACM3493)	–	0.0155
Nectamia fusca	0.2090	0.1417
MOTU174(BIN:AAD9453)	–	0.1088
MOTU175(BIN:AAF8240)	–	0.1746
Ostorhinchus fasciatus	0.0322	0.1170
MOTU191(BIN:N/A)	0.0025	0.0306
MOTU192(BIN:AAC1243)	0.0044	0.0304
Platycephalus indicus	0.0810	0.1602
MOTU218(BIN:AAB2372)	–	0.0810
MOTU219(BIN:AAB2372)	–	0.0810
Rhabdamia gracilis	0.2140	0.1359
MOTU243(BIN:ACS5956)	0.0103	0.0841
MOTU244(BIN:AAC9583)	0.0054	0.1791
Scatophagus argus	0.0668	0.1796
MOTU250(BIN:AAB3531)	0.0031	0.0668
MOTU251(BIN:AAB3530)	–	0.0668
Terapon jarbua	0.0759	0.1673
MOTU278(BIN:AAA9351)	0.0038	0.0710
MOTU279(BIN:AAA9354)	0.0088	0.0738
Thryssa kammalensis	0.2065	0.0932
MOTU284(BIN:AAD4535)	0.0050	0.0369
MOTU285(BIN:ADB1427)	0.0040	0.1841
Upeneus japonicus	0.0127	0.1022
MOTU298(BIN:AAB9326)	0.0091	0.0077
MOTU299(BIN:AAB9327)	0.0077	0.0085

the identification procedures. First, larval fish and eggs were sorted according to their morphological attributes using the most updated field guides available. Second, previously published DNA barcode for SCS fish early stages were mined from BOLD and NCBI. Third, DNA barcodes of individual larval fishes and eggs were blasted in NCBI and Genbank to collect a molecular identification following the methodology previously proposed by Hubert et al. [13]. Blast results were sorted according to a threshold of 99% of similarity, and three categories were determined including match to species (case I), ambiguous match to species (case II) and unmatched (case III) [13, 22]. Cases of ambiguous match were further examined by comparison with sequences mined from Genbank and BOLD, and the origin of the ambiguity was determined *i.e.* species synonymy, shallow divergence or lineage sorting. Fourth, blast results were finally compared to morphological identification and morphological attributes were finally re-examined at the light of DNA barcoding. Here, unambiguous match to species accounted for only 63.29% of the samples, but our iterative procedure helped in improving overall identification results.

Species divergence in the SCS
This study provides molecular evidence for the presence of 305 MOTUs whose delimitation was corroborated by most DNA-based delimitation methods applied. Several instances of large conflicts between mPTP and other algorithms were associated to cases of multiple MOTUs displaying small genetic distances among them (*e.g. Abudefduf septemfasciatus* and *A. vaigiensis* MOTUs). This is a known trend of mPTP which tend to underestimated diversity if diversification rates largely varies among the lineages analyzed [35, 37]. These findings affirm the advantages of integrating multiple species delimitation methods and opting for a consensus approach instead of relying on a single method to avoid artifacts [35, 36, 38, 52]. These methods helped characterized the diversity among larval fish and eggs of the SCS, and highlight multiple cases of shallow divergence among closely related species or multiple cryptic MOTUs within species related to the geographic structure of the SCS.

The accuracy of DNA barcoding in identifying unknown specimen to the species level is tightly linked to the taxonomic coverage of the fauna under scrutiny and the spatial coverage of genetic diversity for widespread species [35]. Spatial scale is particularly important for the detection of a barcode gap, as increasing spatial scale may result in increasing maximum intraspecific genetic distances but also result in decreasing the distance to the nearest neighbour by increasing taxonomic coverage [36]. Comparisons with other large-scale

Jiang *et al. BMC Ecology and Evolution*   (2024) 24:132

Page 13 of 18

Table 4 List of potential MOTUs shared by multiple species including their maximum K2P genetic distances within a MOTU (D_M), K2P distance to their nearest MOTU (D_N), and a list of species detected for each MOTU. The nearest-neighbour distance for each MOTU corresponds to the distance to the closest MOTU, disregarding species boundaries

	D_M	D_N	**Species**
MOTU11	0.0169	0.1429	*Alepes djedaba, Alepes kleinii, Selaroides leptolepis*
MOTU20	0.0077	0.1517	*Arnoglossus polyspilus, Engyprosopon grandisquama*
MOTU25	0.0061	0.0731	*Auxis thazard, Sarda orientalis*
MOTU118	0.0109	0.0529	*Hirundichthys affinis, Hirundichthys oxycephalus*
MOTU141	0.0031	0.0924	*Lagocephalus spadiceus, Lagocephalus wheeleri*
MOTU193	0.0027	0.1333	*Ostorhinchus kiensis, Ostorhinchus quadrifasciatus*
MOTU230	0.0179	0.0581	*Prognichthys brevipinnis, Prognichthys sealei*
MOTU245	0.0118	0.1348	*Sardinella gibbose, Sardinella jussieu*
MOTU249	0.0083	0.2013	*Saurida tumbil, Saurida undosquamis*
MOTU284	0.0050	0.0369	*Thryssa hamiltonii, Thryssa kammalensis*
MOTU287	0.0035	0.0936	*Thunnus albacares, Thunnus tonggol*
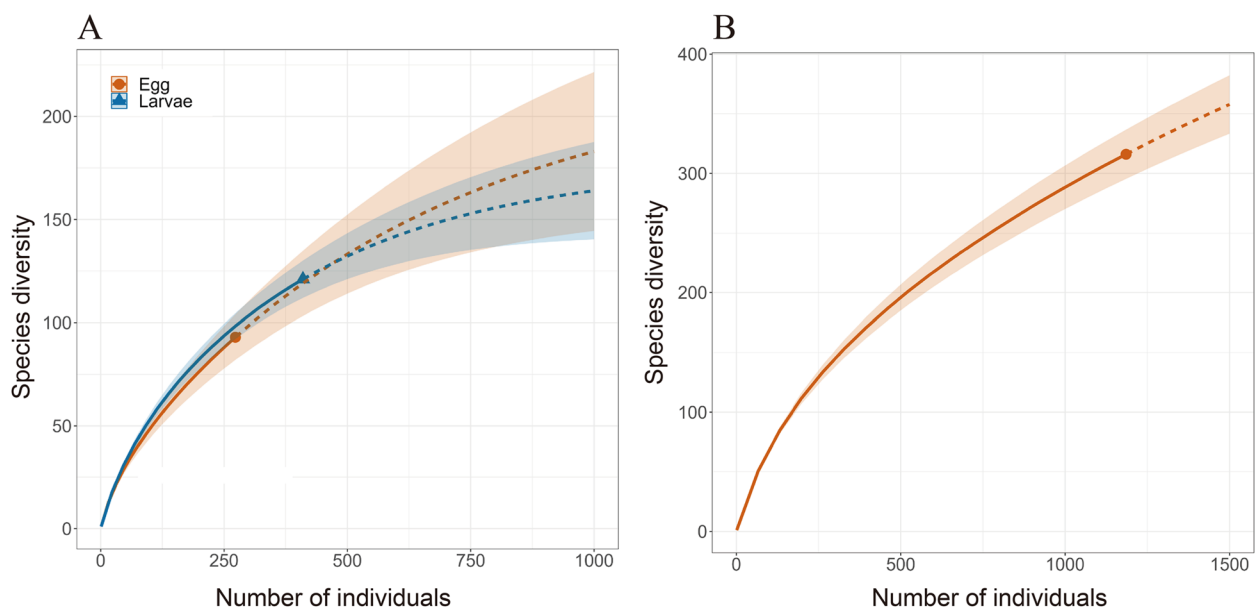
Fig. 8 Accumulation curves for species richness in the present sampling. **A** accumulation curve for larval fish and eggs for the newly generated record, **B** Fish early-stage resources for the entire dataset. Solid and dotted lines represent current and extrapolated trends, respectively

DNA barcoding campaign are consistent as the average intraspecific genetic distance estimated here (0.13%) is lower than observed elsewhere at wider spatial scale such as the Mediterranean (0.39%), Australian shores (0.39%) and the Indo-Pacific (> 1%). However, with an average genetic distance among congeneric species of 8.5%, the SCS exhibits shallower divergence among congeneric species than elsewhere in the Mediterranean (8.91%), Australian waters (9.93%), or Indo-Pacific (> 14%) [18, 53, 54]. This trend suggests that the SCS hosts more closely related species than other DNA-barcoded regions of the Pacific. The variations observed between genera may be

indicative of the average age of congeneric species divergence, as some species are younger than others within genera, and also in comparison to other genera [30]. This trend suggests that the SCS served as a diversification hotspot during the Pliocene, with increased speciation rates compared to other marine fish assemblages.

Shallow divergence and haplotype sharing

DNA-based species delimitation analyses converged with specimen identifications in 285 species where a single MOTU was delimited within a nominal species. This indicates a success rate of 93.44%, which was comparable
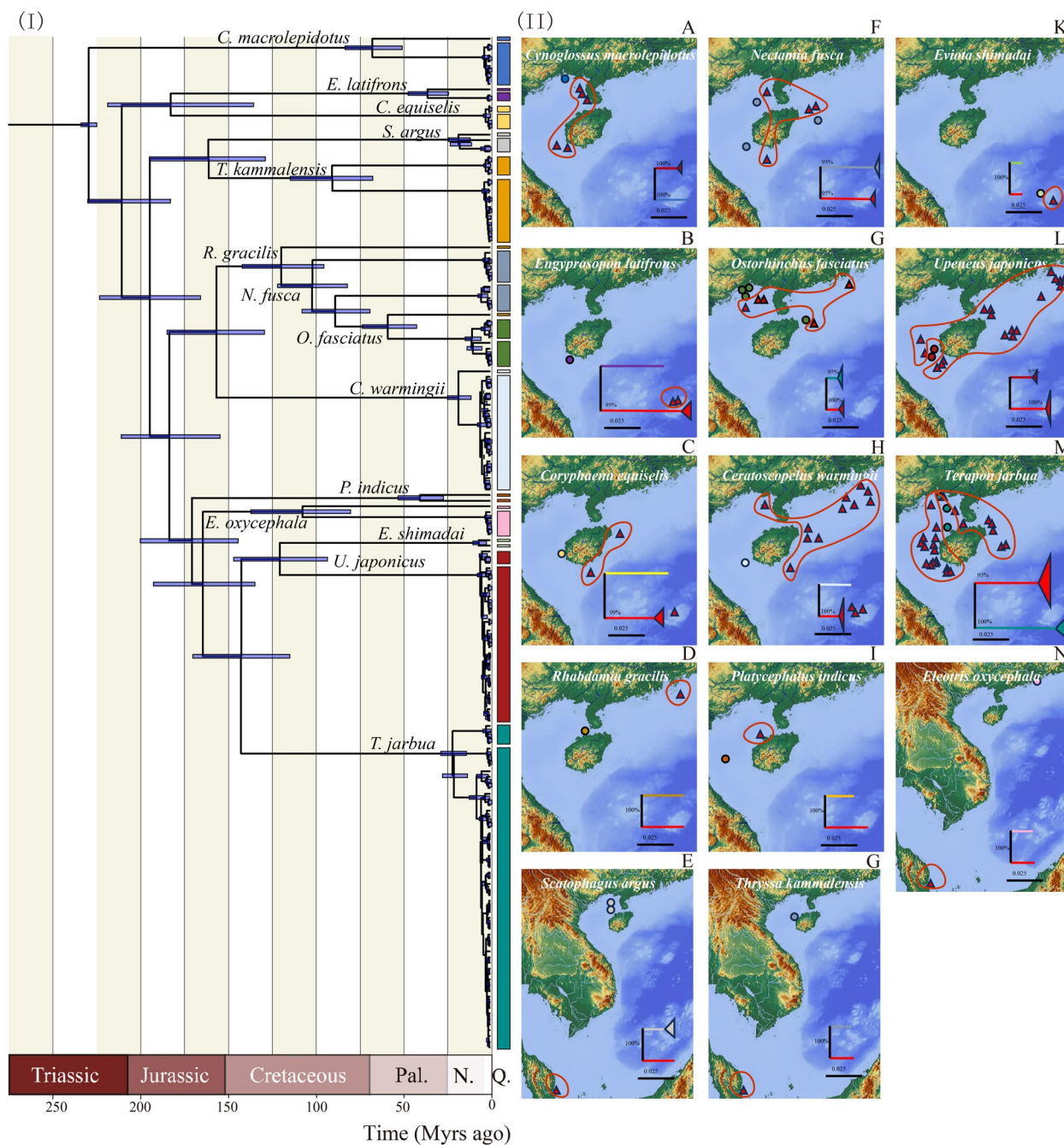
Fig. 9 Median-joining networks for species groups with mixed genealogies and/or sharing haplotypes for COI sequences

Fig. 10 Phylogeographic patterns among selected groups of species with multiple MOTUs. MOTUs are represented according to the final delimitation schemes based on majority rule consensus among the 6 methods. Different colours represent different species, from top to bottom as *C. macrolepidotus*, *E. latifrons*, *C. equiselis*, *S. argus*, *T. kammalensis*, *R. gracilis*, *N. fusca*, *O. fasciatus*, *C. warmingii*, *P. indicus*, *E. oxycephala*, *E. shimadai*, *U. japonicu* and *T. jarbua*, respectively, and on the right side, the geographical patterns correspond to the colours of the respective species. Trees at the bottom right of each map are neighbour-joining trees of the corresponding species using K2P distances. Different colour and symbol (circle vs triangle) represent different lineages and/or MOTUs. Scale bars correspond to K2P genetic distances(the geographical location of one of the lineages has been circled)

with other large-scale studies [51, 55]. In total, 305 MOTUs were detected among the 308 species detected, and 11 MOTUs displayed more than a single species with

23 species. Taking into account the maternal inheritance of mitochondrial genes and the shallow genealogies observed here, the mixing of species genealogies in those

11 cases can be attributed to either recent divergence and incomplete lineage sorting or historical introgressive hybridization [36, 37]. Shallow divergence predominantly occurs among closely related species within the same genus (*e.g. Alepes, Prognichthys, Saurida, Thryssa, Hirundichthys, Sardinella, Thunnus*) but a few cases involving genus paraphyly were detected (*Alepes* and *Selaroides*). This shallow genetic divergence was accompanied by an apparent morphological similarity among species. This trend is consistent with the role of the SCS as a hotspot of diversification for marine fishes [56].

A single case of haplotype sharing is observed between *Thryssa hamiltonii* and *T. kammalensis*, with a single shared haplotype in a central position of their reconstructed haplotype network, suggesting incomplete lineage sorting instead of introgressive hybridization. This shared ancestral haplotype is located in the Beibu Gulf, while more recently derived haplotypes occurring in Malaysia for both species were not shared, suggesting a common origin in the Beibu gulf. However, this hypothesis calls for a broader study of this species pair using molecular markers of biparental inheritance.

Cryptic diversity and phylogeographical patterns

Several cases of multiple, highly divergent MOTUs were detected in 14 species, including *C. warmingii, C. equiselis, C. macrolepidotus, E. oxycephala, E. latifrons, E. shimadai, N. fusca, O. fasciatus, P. indicus, R. gracilis, S. argus, T. jarbua, T. kammalensis* and *U. japonicu* with a total of 28 MOTUs (Table 4). Most intraspecific lineage divergences are dated to the Pliocene, suggesting the influence of historical geological events (Fig. 10I). In *C. warmingii, C. macrolepidotus, E.oxycephala, E.latifrons, N.fusca, O.fasciatus, P.indicus, R.gracilis, S.argus, T.jarbua* and *T.kammalensis,* high divergence (> 3% maximum intraspecific genetic distance) between MOTUs is observed. In *C. macrolepidotus, O. fasciatus, N. fusca, U. japonicu, P. indicus, C. warmingii* and *T. jarbua,* intraspecific MOTUs display alternative geographical distributions in the Beibu Gulf, or the Beibu Gulf and the coastal areas in eastern Guangdong and the waters near Hainan Island. By contrast, *E. shimadai* has two distinct MOTUs observed in the coral reefs of the Zhongsha Islands, which are known to host a high diversity of species [23, 57]. For *E. latifrons, C. equiselis* and *R. gracilis*, two MOTUs were identified from the Beibu Gulf and the coastal areas along the eastern sides of Guangdong and Hainan sampling sites, with one MOTU in the Beibu Gulf and the other in the coastal areas along the eastern sides of Guangdong and Hainan. These distribution patterns are consistent with the influence of the terrestrial isolation, elevation of which created a biogeographical barrier

separating the Beibu Gulf and the coastal areas along the eastern sides of Guangdong and Hainan, a scenario previously proposed by the phylogeographical studies of multiple fish taxa [58]. Hainan Island started to separate and drift away from the northern part of the South China Sea (Beibu Gulf) approximately 65 million years ago (ma) [58], a time frame matching our divergence time estimates among these MOTUs, dated at 32.99–41.92 Ma (*C. equiselis* ~ 32.99ma, *E. latifrons* ~ 41.92ma). Likewise, the two MOTUs detected in *S. argus, E. oxycephala* and *T. kammalensis* show marked geographical distribution patterns, with one MOTU located in Malaysia and another one situated in the northern part of the South China Sea. This pattern is consistent with the influence of the Mid-Indian Ocean Barrier (MIOB) on the divergence of MOTUs within widespread species [59, 60]. The MIOB has been identified as a strong barrier to gene flow for marine organisms, which also indicates that within a highly connected marine environment, older geographical barriers can promote divergence among clades.

Conclusions

Our study provides the most comprehensive DNA barcoding campaign for larval fish and fish eggs assemblages to date in the SCS, with several findings challenging current taxonomic knowledge. To ensure the accuracy of our identifications, a multiple step procedure was implemented which largely helped in improving the resolution of the identification procedure. The present study underscores the integrity and accuracy of the database in the SCS and highlights the significance of this reference library, as well as confirms the utility of standardized DNA-based species delimitation methods in aiding biodiversity inventories and species identification. Conflicts detected between species boundaries and sequence-based species delimitation methods point to relatively shallow inter-specific differences among species for 23 species. At the other end of the spectrum, the detection of multiple, and highly divergent, MOTUs in 14 species suggests that the diversity of SCS fishes is currently underestimated and potentially new species are awaiting a formal description. This pattern point to the influence of several geographical barriers, fostering the emergence of cryptic diversity and unrecognized divergence events [18, 61]. We confirmed the influence of geographic structure on this recent diversification, a trend that further points to the need to improve our knowledge of this biodiversity-rich region. As such, this study warrants further research developments and provides guidelines for future taxonomic studies, sustainable management and conservation of fishery resources in the SCS.

Jiang *et al. BMC Ecology and Evolution*   (2024) 24:132

Page 17 of 18

Supplementary Information

Data availability
Sequences were submitted to Genbank, and are publicly available with accession numbers PP354153-PP354861 and PP354116-PP354147 (Table S1).

Declarations

Ethics approval and consent to participate
All the sampling sites were not privately owned or protected, and field sampling did not involve protected species. The fish collection process complied with the guidelines of Guangdong Ocean University. This study was approved by the Laboratory Animal Ethics Committee of Guangdong Ocean University. All animals and experiments were conducted in accordance with the "Guidelines for Experimental Animals" of the Ministry of Science and Technology (Beijing, China).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Hsieh C, Reiss CS, Hunter JR, Beddington JR, May RM, Sugihara G. Fishing elevates variability in the abundance of exploited species. Nature. 2006;443:859–62.
2. Kritzer JP, Sale PF. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish Fish. 2004;5:131–40.
3. Beger M, Linke S, Watts M, Game E, Treml E, Ball I, et al. Incorporating asymmetric connectivity into spatial decision making for conservation. Conserv Lett. 2010;3:359–68.
4. Gell FR, Roberts CM. Benefits beyond boundaries: the fishery effects of marine reserves. Trends Ecol Evol. 2003;18:448–55.
5. Muhling BA, Lamkin JT, Alemany F, García A, Farley J, Ingram GW, et al. Reproduction and larval biology in tunas, and the importance of restricted area spawning grounds. Rev Fish Biol Fish. 2017;27:697–732.
6. Asch RG. Climate change and decadal shifts in the phenology of larval fishes in the California current ecosystem. Proc Natl Acad Sci. 2015;112:E4065–74.
7. Hobday AJ, Pecl GT. Identification of global marine hotspots: sentinels for change and vanguards for adaptation action. Rev Fish Biol Fish. 2014;24:415–25.
8. Last PR, White WT, Gledhill DC, Hobday AJ, Brown R, Edgar GJ, et al. Long-term shifts in abundance and distribution of a temperate fish fauna: a response to climate change and fishing practices. Glob Ecol Biogeogr. 2011;20:58–72.
9. Wernberg T, Smale DA, Tuya F, Thomsen MS, Langlois TJ, De Bettignies T, et al. An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. Nat Clim Change. 2013;3:78–82.
10. Smith PE, Moser HG. Long-term trends and variability in the larvae of Pacific sardine and associated fish species of the California Current region. Deep Sea Res Part II Top Stud Oceanogr. 2003;50:2519–36.
11. Leis JM. Taxonomy and systematics of larval Indo-Pacific fishes: a review of progress since 1981. Ichthyol Res. 2015;62:9–28.
12. Choat JH. A comparison of towed nets, purse seine, and light-aggregation devices for sampling larvae and pelagic juveniles of coral reef fishes. Fish BullUS. 1993;91:195–209.
13. Hubert N, Espiau B, Meyer C, Planes S. Identifying the ichthyoplankton of a coral reef using DNA barcodes. Mol Ecol Resour. 2015;15:57–67.
14. Smith JA, Miskiewicz AG, Beckley LE, Everett JD, Garcia V, Gray CA, et al. A database of marine larval fish assemblages in Australian temperate and subtropical waters. Sci Data. 2018;5:1–8.
15. Ko H-L, Wang Y-T, Chiu T-S, Lee M-A, Leu M-Y, Chang K-Z, et al. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS One. 2013;8:e53451.
16. Schmidt S, Schmid-Egger C, Morinière J, Haszprunar G, Hebert PD. DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). Mol Ecol Resour. 2015;15:985–1000.
17. Ratnasingham S, Hebert PD. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS One. 2013;8:e66213.
18. Hubert N, Meyer CP, Bruggemann HJ, Guerin F, Komeno RJ, Espiau B, 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. 2012;7:e28987.
19. Alien GR, Amaoka K, Anderson WD Jr, Bellwood DR, Bohlke EB, Bradbury MG, et al. A checklist of the fishes of the South China Sea. Raffles Bull Zool. 2000;8:569–667.
20. Huang D, Licuanan WY, Hoeksema BW, Chen CA, Ang PO, Huang H, et al. Extraordinary diversity of reef corals in the South China Sea. Mar Biodivers. 2015;45:157–68.
21. Azmir IA, Esa Y, Amin SMN, Md Yasin IS, Md Yusof FZ. Identification of larval fish in mangrove areas of Peninsular Malaysia using morphology and DNA barcoding methods. J Appl Ichthyol. 2017;33:998–1006.
22. Hou G, Chen Y, Wang J, Pan C, Lin J, Feng B, et al. Molecular identification of species diversity using pelagic fish eggs in spring and late autumn-winter in the Eastern Beibu Gulf, China. Front Mar Sci. 2022;8:806208.
23. Huang D, Chen J, Xu L, Wang X, Ning J, Li Y, et al. Larval fish assemblages and distribution patterns in the Zhongsha Atoll (Macclesfield Bank, South China Sea). Front Mar Sci. 2022;8:787765.
24. Hou G, Xu Y, Chen Z, Zhang K, Huang W, Wang J, et al. Identification of eggs and spawning zones of hairtail fishes *Trichiurus* (Pisces: Trichiuridae)

Jiang *et al. BMC Ecology and Evolution*   (2024) 24:132

Page 18 of 18

in Northern South China Sea, using DNA barcoding. Front Environ Sci. 2021;9:703029.

25. Hou G, Chen Y, Wang S, Wang J, Chen W, Zhang H. Formalin-fixed fish larvae could be effectively identified by DNA barcodes: a case study on thousands of specimens in south China Sea. Front Mar Sci. 2021;8:634575.

26. Okiyama M. An Atlas of the early stage stage fishes in Japan. 1989.

27. Neira FJ, Miskiewicz AG. Larvae of temperate Australian fishes: laboratory guide for larval fish identification. Nedlands: University of Western Australia Press; 1998.

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

29. Moser HG, Watson W. Preliminary guide to the identification of the early life history stages of myctophiform fishes of the western central Atlantic. NOAA Technical Memorandum NMFS-SEFSC-453; 2001.

30. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. DNA barcoding Australia's fish species. Philos Trans R Soc B Biol Sci. 2005;360:1847–57.

31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

32. Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, et al. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics. 2017;33:1630–8.

33. Puillandre N, Brouillet S, Achaz G. ASAP: assemble species by automatic partitioning. Mol Ecol Resour. 2021;21:609–20.

34. Luo A, Ling C, Ho SY, Zhu C-D. Comparison of methods for molecular species delimitation across a range of speciation scenarios. Syst Biol. 2018;67:830–46.

35. Arida E, Ashari H, Dahruddin H, Fitriana YS, Hamidy A, Irham M, et al. Exploring the vertebrate fauna of the Bird's Head Peninsula (Indonesia, West Papua) through DNA barcodes. Mol Ecol Resour. 2021;21:2369–87.

36. Chen W, Hubert N, Li Y, Xiang D, Cai X, Zhu S, et al. Large-scale DNA barcoding of the subfamily Culterinae (Cypriniformes: Xenocyprididae) in East Asia unveils a geographical scale effect, taxonomic warnings and cryptic diversity. Mol Ecol. 2022;31:3871–87.

37. Shen Y, Hubert N, Huang Y, Wang X, Gan X, Peng Z, et al. DNA barcoding the ichthyofauna of the Yangtze River: Insights from the molecular inventory of a mega-diverse temperate fauna. Mol Ecol Resour. 2019;19:1278–91.

38. Jiang C, Yi M, Luo Z, He X, Lin H, Hubert N, et al. DNA barcoding the ichthyofauna of the Beibu Gulf: Implications for fisheries management in a seafood market hub. Ecol Evol. 2023;13:e10822.

39. Fujisawa T, Barraclough TG. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. Syst Biol. 2013;62:707–24.

40. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27:592–3.

41. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014;10:e1003537.

42. Bermingham E. Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. Mol Syst Fishes. 1997:113–28.

43. Glez-Pena D, Gnmez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D. ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Res. 2010;38:W14–8.

44. Hsieh TC, Ma KH, Chao A, McInerny G. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol Evol. 2016;7:1451–6.

45. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16:111–20.

46. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20:289–90.

47. Brown SD, Collins RA, Boyer S, Lefort M-C, Malumbres-Olarte J, Vink CJ, et al. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour. 2012;12:562–5.

48. Blagoev GA, Dewaard JR, Ratnasingham S, Dewaard SL, Lu L, Robertson J, et al. Untangling taxonomy: a DNA barcode reference library for Canadian spiders. Mol Ecol Resour. 2016;16:325–41.

49. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.

50. Hebert PD, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci. 2003;270:313–21.

51. April J, Mayden RL, Hanner RH, Bernatchez L. Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci. 2011;108:10602–7.

52. Wu Y-H, Hou S-B, Yuan Z-Y, Jiang K, Huang R-Y, Wang K, et al. DNA barcoding of Chinese snakes reveals hidden diversity and conservation needs. Mol Ecol Resour. 2023;23:1124–41.

53. Hubert N, Dettai A, Pruvost P, Cruaud C, Kulbicki M, Myers RF, et al. Geography and life history traits account for the accumulation of cryptic diversity among Indo-West Pacific coral reef fishes. Mar Ecol Prog Ser. 2017;583:179–93.

54. Landi M, Dimech M, Arculeo M, Biondo G, Martins R, Carneiro M, et al. DNA barcoding for species assignment: the case of Mediterranean marine fishes. PLoS One. 2014;9:e106135.

55. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, et al. Identifying Canadian freshwater fishes through DNA barcodes. PLoS One. 2008;3:e2490.

56. Allen GR. Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. Aquat Conserv Mar Freshw Ecosyst. 2008;18:541–56.

57. Hodge JR, Bellwood DR. The geography of speciation in coral reef fishes: the relative importance of biogeographical barriers in separating sister-species. J Biogeogr. 2016;43:1324–35.

58. Pellissier L, Leprieur F, Parravicini V, Cowman PF, Kulbicki M, Litsios G, et al. Quaternary coral reef refugia preserved fish diversity. Science. 2014;344:1016–9.

59. Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MCA, Mahardika GN, Manjaji-Matsumoto BM, et al. Comparative phylogeography of the Coral Triangle and implications for marine management. J Mar Sci. 2011;2011:396982.

60. Hung K-W, Russell BC, Chen W-J. Molecular systematics of threadfin breams and relatives (Teleostei, Nemipteridae). Zool Scr. 2017;46:536–51.

61. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. PLoS Biol. 2004;2:e312.

Publisher's Note