Phylotranscriptomics analysis uncovers new insights into the phylogeny and evolutionary history of Epinephelidae

Qing-Hua Wang[1], Wei-Wei Zhang[1], Zhuo-Ying Weng[1], Yang-Fan Chen[1], Xi Wang[3], Le Wang[4], Xiao-Chun Liu[1,2], Yang Yang[5,*], Zi-Ning Meng[1, 2,*]

[1] School of Life Sciences, State Key Laboratory of Biocontrol / Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) / Guangdong Provincial Key Laboratory of Aquatic Economic Animals, Sun Yat-sen University, Guangzhou 510275, China

[2] China-ASEAN Belt and Road Joint Laboratory on Mariculture Technology, Guangzhou 510275, China

[3] Area of Ecology and Biodiversity, School of Biological Sciences, University of Hong Kong, Hong Kong SAR 999077, China

[4] Molecular Population Genetics Group, Temasek Life Sciences Laboratory, Singapore City 117604, Singapore

[5] Key Laboratory of Tropical Marine Fish Germplasm Innovation and Utilization, Ministry of Agriculture and Rural Affairs,Hainan Engineering Research Center for Germplasm Innovation and Utilization, Hainan Chenhai Aquatic Co., Ltd, Sanya 570000, China

[*] Corresponding authors, E-mail: yangy888@mail.sysu.edu.cn; mengzn@mail.sysu.edu.cn.

Zoological Research

ABSTRACT

Coral reef fishes, distinguished by remarkable species diversity and unique speciation mechanisms, offer an exceptional framework for understanding evolutionary diversification. As ecologically dominant predators, groupers (family Epinephelidae) exemplify high diversity in coral reef ecosystems, yet their phylogenetic relationships remain partially unresolved in previous studies based on mitochondrial and nuclear markers. Here, we performed a comprehensive phylotranscriptomic analysis of 92 specimens representing 32 species across seven genera of Epinephelidae. Leveraging concatenated nucleotide data analyzed with maximum likelihood and Bayesian inference, we firstly reconstructed a robust transcriptomic phylogeny with markedly enhanced resolution and nodal support compared to traditional marker-based frameworks. Our results revealed *Plectropomus* as the earliest diverging lineage, resolving a longstanding debate regarding the basal divergence within this family, and robustly supported the non-monophyly of *Epinephelus* and *Cephalopholis*. The origin and early diversification of Epinephelidae currently distributed across the Indian and Pacific Oceans may have occurred in the West Pacific during the Paleocene, followed by subsequent dispersal mediated by prevailing ocean currents. Furthermore, we found that the vicariance and speciation of major clades were likely driven by repeated episodes of sea level regression associated with glacial cycles. Enhanced upwelling and elevated primary productivity in recent periods may have further promoted the rapid radiation of groupers. To our knowledge, this represents the first application of phylotranscriptomics in the phylogenetic study of Epinephelidae, providing new insights into its phylogeny and evolutionary history.

Keywords: Biogeography; Epinephelidae; Evolutionary history; Phylogenetic comparison; Phylogeny; Phylotranscriptomics

INTRODUCTION

Coral reef ecosystems, though occupying only ~0.1% of the ocean's surface, support nearly one-third of all marine fish species and constitute one of the most diverse vertebrate assemblages (Bellwood and Hughes, 2001; Cowman, 2014). This remarkable biodiversity has arisen through intricate biogeographic dynamics, ecological adaptations, and repeated vicariant and sympatric events (Rocha et al., 2005; Siqueira et al., 2020). Groupers (Teleostei: Epinephelidae), comprising over 160 species across 16 genera, represent an ecologically pivotal and taxonomically diverse lineage and are emblematic of the dynamic diversification characteristic of coral reef ecosystems (Craig et al., 2011; Ma et al., 2016). As reef-dwelling fishes, groupers are widely distributed across tropical and subtropical coral reefs and continental shelves in the Atlantic, Indian, and Pacific Oceans (Craig et al., 2011). Groupers exhibit remarkable morphological and ecological diversity, including total length ranging from 30 to 270 cm, protogynous hermaphroditism with bidirectional sex reversal in some species, and a range of reproductive strategies characterized by

prolific spawning and complex mating systems (Heemstra and Randall, 1993; Erisman et al., 2009; Choat, 2012). Given these traits and ecological roles, groupers serve as great species for understanding the evolutionary underpinnings of reef fish diversification.

Phylogenetic relationships within Epinephelidae have been extensively studied, beginning with early work based on morphological taxonomy (Baldwin and Johnson, 1993). Initial molecular investigations utilized single-gene (16S ribosomal RNA gene, 16S) analyses to explore the phylogeny of groupers (Craig et al., 2011). Subsequent analyses expanded molecular datasets by incorporating mitochondrial markers such as Cytochrome *b* (Cyt *b*), 12S ribosomal RNA gene (12S), and Cytochrome c oxidase subunit I (COI) (Maggio et al., 2005; Ding et al., 2006; Craig and Hastings, 2007; Ma et al., 2016; Qu et al., 2018), as well as nuclear genes including Tyrosine kinase receptor C4 (TMO4C4), Histone H3 gene (Histone III), and Ryanodine receptor 3 (RYR3) (Craig and Hastings, 2007; Ma et al., 2016; Qu et al., 2018). Maggio et al. (2005) integrated Cyt *b* and 16S data to resolve relationships among eight Eastern Atlantic species, supporting a tripartite division of *Epinephelus*. Based on mitochondrial (12S and 16S) and nuclear genes (TMO4C4 and Histone III) across 155 species in 24 genera, Craig et al. (2007) ultimately warranted elevation of Epinephelidae from subfamily to family status. Building upon these foundations, Ma et al. (2016) reconstructed a more comprehensive phylogeny using four markers across 147 species (~87% of known diversity), illuminating the clade diversification patterns and processes. Nevertheless, some relationships remain unresolved, for example, the question of whether *Plectropomus* or *Variola* represents the earliest diverging lineage (Craig and Hastings, 2007; Ma et al., 2016; Ding et al., 2018). In addition, strong convergence in skeletal and external morphology, coupled with environmentally or physiologically induced shifts in body coloration and patterning, obscures species boundaries in groupers (Ding et al., 2018). Such complex traits often generate greater phenotypic variation within species than among them, driving persistent errors in taxonomic identification (Ding et al., 2018; Qu et al., 2018). It is thus very important to utilize high-resolution phylogenomic data.

Over the past three decades, technological breakthroughs in PCR and Sanger sequencing have altered the landscape of evolutionary biology, driving the transition from morphology-based taxonomy to molecular phylogenetics and phylogeographic frameworks. The emergence of high-throughput sequencing techniques has offered molecular systematists with powerful tools to analyze phylogenetic relationships in greater depth (da Fonseca et al., 2016). Phylogenomics can significantly mitigate sampling errors inherent in traditional phylogenetic analyses constrained by limited gene number and length (Delsuc et al., 2005). However, phylogenetic reconstruction remains challenging in taxa characterized by high biological diversity, complex biogeographic patterns, and ambiguous morphological characters, limiting the resolution of some phylogenetic relationships (Struck, 2013; Cox et al., 2014; Kocot et al., 2017). Phylogenetic topologies are highly sensitive to both dataset composition and inference methods. Although high-throughput

sequencing data can reduce stochastic error, accurate species-tree reconstruction requires rigorous locus selection, precise alignment curation, and inference-aware optimization to minimize systematic bias and ensure topological robustness (Du et al., 2019; Young and Gillung, 2020). As such, the careful selection of appropriate datasets and phylogenetic methods plays a critical role in the phylogenetic research of specific species.

RNA sequencing (RNA-seq) has been extensively employed in phylogenetic inference, with phylotranscriptomics emerging as a powerful and cost-effective approach, to uncover patterns of lineage diversification (Kadlec et al., 2017; Chen et al., 2022; Benítez-Álvarez et al., 2023). By enabling the reliable identification and extraction of orthologous genes, phylotranscriptomic data have demonstrated exceptional utility in resolving both shallow population-level differentiation and deep phylum-level phylogenetic divergence (Simion et al., 2017; Hughes et al., 2018). To date, phylotranscriptomics has been successfully applied in phylogenetic studies across diverse taxa, such as plants (Chen et al., 2022; Liu et al., 2022), bivalves (Li et al., 2020), and fishes (Tong et al., 2017; Ciezarek et al., 2019). In this study, we applied transcriptome data using multiple datasets and various phylogenetic methods to reconstruct a robust, highly supported, and well-resolved phylogenetic tree of family Epinephelidae. To ensure that our sampling was representative of the extant diversity within this family, we collected all publicly available transcriptome datasets for grouper species. By comparing our transcriptome-based phylogeny with previously reported phylogenies derived from limited mitochondrial and nuclear markers, we further clarified intergeneric relationships, divergence processes, and ancestral distributions within this family. Our study establishes a refined evolutionary framework for Epinephelidae and underscores the transformative potential of phylotranscriptomics in resolving the reef fish lineages.

MATERIALS AND METHODS

Ethics statement

All animal experiments were performed in accordance with the Guidelines for the Care and Use of Laboratory Animals in China. All experimental procedures and sample collection methods were approved by the Institutional Animal Care and Use Committee (IACUC) of School of Life Sciences, Sun Yat-sen University (approval No. SYSU-IACUS-2022-B0129).

Grouper sampling and data collection

A total of 92 grouper specimens were collected, representing 32 species of 7 genera of family Epinephelidae, and used for phylotranscriptomic analysis. Of these transcriptomes, 48 were assembled in our previous research (Yang et al., 2020; Zhang et al., 2024), 6 were downloaded from NCBI (Gomes et al., 2022), and 38 were newly sequenced (Supplementary Table S1 and S2). Most specimens were sampled from the West Pacific, with a smaller number collected from the Atlantic Ocean. According to habitat and distribution data compiled from Grouper of the

World: A Field and Market Guide (Craig et al., 2011), the West Pacific grouper included in this study are mainly distributed across the West Pacific and Indian Ocean regions. Fish tissues, including tail fin, skin, muscle, liver, spleen, brain, kidney, heart, spine, gill, gonads, intestine, and stomach, were collected from one to six individuals of each species, preserved in RNAlater, and stored at −80 ℃ until RNA extraction.

Transcriptome sequencing and assembly

Total RNA was extracted from each sample using Trizol Reagent Kit (Invitrogen, CA, USA) according to the manufacturer's instructions. RNA concentrations and purity were measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, DE, USA). RNA integrity was assessed using RNA Nano 6000 Assay Kit on the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). One μg of RNA preparation was obtained by mixing equal aliquots of all tissues from each individual. RNA libraries were constructed using NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's protocols to obtain Illumina paired-end reads. Briefly, mRNA was enriched using poly-T oligo-attached magnetic beads (Qiagen, USA) and fragmented into short fragments using fragmentation buffer. First-strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase, followed by the synthesis of second-strand cDNA and NEBNext adaptor ligation. The cDNA library fragments and PCR products were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). All libraries were sequenced on the Illumina NovaSeq 6000 platform (Illumina, CA, USA) at BioMarker Biotechnology Co., Ltd (Beijing, China).

The raw sequencing RNA-seq reads were visualized and quality-checked using FastQC v0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads were first filtered using Trimmomatic v.0.39 (Bolger et al., 2014) with a minimum reads length (MINLEN) of 100 to obtain clean reads. Transcriptome assembly was accomplished using Trinity v.2.8.5 (Grabherr et al., 2011) with --CPU set to 20, --min_kmer_cov set to 2, and all other parameters set to default. All assemblies of each species were merged into one transcriptome. Preliminary redundancy reduction was implemented with CD-HIT-EST v.4.6 (Li and Godzik, 2006), implementing a sequence identity threshold of 0.95 and applying a word length of 10. The longest transcripts were selected using the "get_longest_isoform_seq_per_trinity_gene.pl" implemented in Trinity (Haas et al., 2013). Longest isoforms were used to predict open reading frames (ORFs) and coding sequences (CDS) using TransDecoder v.5.5.0 (Haas and Papanicolaou, 2019) against the NCBI Non-Redundant Protein Sequence (Nr), UniProt Knowledge Base (UniProtKB/Swiss-Prot), and Protein Family (Pfam) databases with an E-value threshold of $< 10^{-5}$. Local BLASTX searches were conducted for the predicted CDS to further reduce redundancy using DIAMOND (Buchfink et al., 2015) against the NCBI RefSeq (Pruitt et al., 2007) of all Actinopterygii datasets (189 species). Transcriptome completeness was assessed using BUSCO v.5.5.0 (Manni et al., 2021) according to the actinopterygii_odb10 database (highly conserved fish orthologs,

https://busco.ezlab.org). Assemblies with less than 50% complete BUSCOs were excluded, while those with 50% or higher completeness were retained for downstream ortholog identification.

Ortholog identification and sequence alignment

Orthogroups were identified using OrthoFinder v.2.5.5 (Emms and Kelly, 2015), applying an all-against-all BLAST algorithm to implement reciprocal best-hit BLASTs, using DIAMOND as the sequence similarity search engine, with the normalization of transcript length (Carruthers et al., 2018). Through ortholog identification, a total of 108 one-to-one single-copy orthologous genes (OGs) were identified and used for downstream analysis. All OGs were aligned using MAFFT v.7.453 (Katoh and Standley, 2013) with the L-INS-I algorithm. Subsequently, aligned OGs were trimmed using trimAl v.1.4 (Capella-Gutiérrez et al., 2009) with the --automated1 parameter.

Phylogenetic analyses

Phylogenetic trees of family Epinephelidae were reconstructed using two different dataset types (nucleotide and amino acid), five distinct phylogenetic inference methods (ML, BI, MSC, NJ, and ME), and two gene integration strategies (concatenation and coalescence).

(1) Phylogenetic analyses using two different dataset types

The corresponding nucleotide CDS were obtained and aligned using PAL2NAL v.14.0 (Suyama et al., 2006) against the amino acids of orthologs. Using the same phylogenetic method and gene integration strategy, the concatenated nucleotide and amino acid datasets were used to reconstruct a phylogenetic tree, respectively. Meanwhile, the average pairwise distance between species was calculated to further quantify the relationship within major clades for both nucleotide and amino acid datasets.

(2) Phylogenetic analyses using five distinct inference methods

Based on the same dataset and gene integration strategy, phylogenetic analyses of groupers were implemented using five distinct phylogenetic methods (Supplementary Table S3). First, maximum likelihood (ML) phylogenetic inference was performed using IQ-TREE v.2.1.2 (Nguyen et al., 2015) with 10,000 bootstrap replicates under the best-fit model determined by ModelFinder, which applied the ModelFinder Plus (MFP) option for -m parameter (Kalyaanamoorthy et al., 2017). Node support was evaluated by the SH-like approximate likelihood ratio test (SH-aLRT) with 10,000 replicates. Second, for the Bayesian inference (BI) analysis, the best-fit model was estimated by Akaike Information Criterion (AICc) score using PartitionFinder v.2.1.1 (Lanfear et al., 2017). The BI tree was inferred by MrBayes v.3.2.7 (Ronquist et al., 2012) with the parameter set as follows: Four Markov Chain Monte Carlo (MCMC) chains (including three heated and one cold chains) were run independently for 10 million generations and sampled every 1,000 generations. The first 20% of trees were discarded as burn-in for each MCMC run before convergence, and the subsequent trees were used to generate a phylogram. MCMC convergence was checked using Tracer v.1.7.1 (Rambaut et al., 2018), with

the effective sample sizes (ESS) of all parameters over 200. Third, species tree inference was conducted with the multispecies coalescent (MSC) model implemented in ASTRAL-Pro. Gene trees inferred from OGs were used as input for subsequent species tree estimation under the mixture model-based ML method in IQ-TREE. All analyses were conducted with default ASTRAL-Pro parameters. The individual gene trees were inferred using DensiTree v.3.1.0 (Bouckaert and Heled, 2014) to visualize the gene tree discrepancy. Each tree was individually rooted and ordered using Newick-utils (Junier and Zdobnov, 2010), followed by imposing ultrametric and dichotomous constraints by the phytools package v.2.3.0 (Revell, 2012) in R v.4.4.3 (https://r-project.org/). Finally, neighbour-joining (NJ) and minimum-evolution (ME) analyses were implemented using MEGA v.11.0 (Tamura et al., 2021) with 10,000 pseudo replicates. All phylogenetic information was exported in newick format and visualized using FigTree v.1.4.4 (Rambaut, 2018).

(3) Phylogenetic analyses using two gene integration strategies

According to the same dataset and phylogenetic method, the concatenated dataset was used to reconstruct a concatenation-based phylogenetic tree. The files of the concatenated dataset were obtained with the software PhyloSuite v.1.2.3 (Xiang et al., 2023). In addition, the concordance and conflict among gene trees across datasets were evaluated by mapping individual gene trees onto the coalescent-based tree using PhyParts (Smith et al., 2015). The pie chart of each internal branch (node) was summarized and visualized using the phypartspiecharts.py script (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts).

Divergence time estimation

Divergence time estimation was implemented using MCMCtree program in PAML v.4.10.7 (Yang, 2007) using the approximate likelihood method, based on the nucleotide dataset and the ML tree inferred from it. With the root age set to 65 million years (Ma), referring to the divergence time estimates for Epinephelidae and Serranidae by Ma et al. (2016), the overall substitution rate was roughly estimated to be 0.086, using BASEML under the GTR model. Given the limited fossil record of Epinephelidae, calibration was based on a notable *Epinephelus* fossils: *E. casottii*, represents the oldest species of *Epinephelus*, dated to the Badenian stage (also known as the mid-Miocene, 16.5–13 Ma) (Schultz, 2000), establishing the minimum age for the common ancestor of *Epinephelus*. The divergence of *Epinephelus* was constrained by a hard minimum age of 13 Ma for its most recent common ancestor (TMRCA) (Ma et al., 2016). Divergence time was also calibrated according to the recorded divergence time range in TimeTree database (http://timetree.org/) of two genera pairs: *Plectropomus* & *Variola* (22.3–62.3 Ma) and *Variola* & *Cephalopholis* (26.7–54.7 Ma). Bayesian inference was carried out for 10 million generations, with parameters sampled at intervals of 100 generations. The first 20% of samples were removed as burn-in, and statistical robustness was verified by assessing ESS (> 200) in Tracer (Rambaut et al., 2018).

Ancestral area reconstruction

Ancestral area reconstruction and biogeographic inference were performed on the time tree excluding the outgroup using the R package BioGeoBEARS (Matzke, 2013). According to Craig et al. (2011) and Ma et al. (2016), a total of seven sea areas were defined, including West Atlantic, East Atlantic, West Indian, North Indian, South Indian, West Pacific, and Central Pacific Oceans (**Figure 1**; Supplementary Table S4). The geographic distribution of each species was labeled according to the monograph of Grouper of the World: A Field and Market Guide (Craig et al., 2011). Each species was assigned to one to five sea areas. Given that *E. marginatus* is the only Atlantic species represented in the samples of this study, its inclusion could potentially bias biogeographic inference. Therefore, this species was excluded from the analysis. The BAYAREALIKE, DEC, and DIVALIKE models were used to infer the ancestral geographic ranges (Matzke, 2014), and the best-fitting model was determined based on the corrected AICc score.

RESULTS

Transcriptome assembly summary and ortholog identification

Transcriptome assembly from 92 Epinephelidae individuals yielded 14.29 million Trinity transcripts and 11.83 million unigenes (Supplementary Table S2), with a mean GC content of 46.47% (range from 43.41% to 48.77%). Following the retention of only the longest isoforms, the assembly yielded 8.35 Gb of bases, achieving an N50 of 1,272 bp and contig lengths with a median of 366 bp and a mean of 723 bp. After merging all assemblies of each species into one transcriptome, the merged transcripts were filtered for redundancy reduction, resulting in an average of 226,854 unigenes each species, with an average unigene length of 787 bp (Supplementary Table S5). An average of 152,922 longest unigenes per sample were used for ORF and CDS prediction, identifying an average of 80,686 candidate coding sequences. Functional annotation mapped the average of 41,770, 41,121, and 40,944 unigenes to NR, Pfam, and Swiss-Prot databases, respectively. After aligning to NCBI RefSeq data from 189 fishes, an average of 28,494 final identified unigenes were used for completeness assessment and OGs search. BUSCO assessment identified an average of 69.92% complete orthologs, with an average of 64.92% as complete and single-copy genes, supporting the completeness and robustness of the final transcriptomes (Supplementary Table S6).

The phylogenetic framework of Epinephelidae was resolved using transcriptomic data from 32 species of the family and one outgroup from Serranidae (*Centropristis striata*; NCBI accession number: GCF_030273125). The ortholog identification recovered 108 OGs. To assess whether these OGs were strongly linked to each other that may bias phylogenetic inference, we mapped all 108 genes onto the chromosome-level genome assembly of *Epinephelus coioides* (NCBI accession number: ASM5131402v1). The OGs were broadly distributed across the genome, and even the

most spatially proximate pair (gene 25 and gene 28 on CM120195.1) were separated by 3,431 bp (Supplementary Figure S1). This pattern confirmed that the markers do not exhibit strong linkage. Accordingly, all 108 OGs used to infer the phylogenetic relationships of Epinephelidae.

Phylogenetic comparisons based on different dataset types, inference methods, and gene integration strategies

(1) Phylogenetic comparison using nucleotide and amino acid datasets

Phylogenetic trees generated from 32 species of family Epinephelidae revealed several inconsistencies in support values and topologies for some clades when using nucleotide and amino acid datasets. ML analysis yielded identical topologies using the concatenated nucleotide and amino acid datasets, indicating a relatively consistent phylogeny (**Figure 2A**). However, for BI, NJ, and ME analyses, some inconsistencies in topologies were observed between nucleotide and amino acid datasets, particularly in the phylogenetic relationships within *Epinephelus* and between *Epinephelus* and *Anyperodon* (Supplementary Figure S2). Notably, when constructed using nucleotide sequences, the mean support values for these four phylogenetic trees were higher compared to those based on amino acid sequences, suggesting stronger support for phylogenetic inferences derived from nucleotide data.

To further evaluate the discordance between nucleotide and amino acid datasets, individual gene trees were visualized in DensiTree (**Figure 2B**). Here, the nucleotide data showed a clearer pattern and higher informativeness than amino acid data. All individual gene trees were successfully constructed using nucleotide data, while two individual gene trees constructed using amino acid data failed to be generated. The successfully constructed individual gene trees were used to obtain a MSC tree (**Figure 2C**). For MSC analysis, some inconsistent topologies and support values were also observed in the phylogenetic relationships within *Epinephelus* and between *Epinephelus* and *Anyperodon* based on nucleotide and amino acid data. Furthermore, the values of final normalized quartet score of 0.85 and 0.67 for nucleotide and amino acid data, respectively, which suggested that approximately 85% and 67% of the quartet trees in those individual gene trees were consistent with the MSC trees generated from these datasets. In addition, MSC trees constructed from nucleotide sequences exhibited higher mean support values than those inferred from amino acid data. Thus, the results of MSC analysis also showed a higher informativeness and more robust phylogenetic construction for nucleotide data.

Genetic distance was also used to evaluate phylogenetic relationships (Supplementary Figure S3). The genetic distances estimated from nucleotide data were consistent with the reconstructed phylogenetic tree. Species with closer relationships exhibited smaller genetic distances, whereas more distantly related species showed larger genetic distances. In contrast, the genetic distances calculated from amino acid data did not correspond to the phylogenetic relationships inferred from the tree, and no clear linear pattern was observed between phylogenetic relationships and genetic distance. Together, these results indicated that nucleotide datasets provided a more reliable basis

for reconstructing the phylogeny of groupers.

(2) Phylogenetic comparison using ML, BI, MSC, NJ, and ME methods

Based on the above results, the analysis primarily focused on the phylogenetic reconstruction using nucleotide data. Phylogenetic comparisons among different inference methods revealed discrepancies in both support values and tree topologies for certain clades (**Figure 3**). The topologies of phylogenetic trees resulting from ML and BI analyses were consistent with each other but differed from those derived from MSC, NJ, and ME methods, with the main differences observed within *Epinephelus* and between *Epinephelus* and *Anyperodon*. Moreover, BI trees yielded higher mean support values than those generated by ML, MSC, NJ, and ME analyses, suggesting that BI provides more reliable support for nucleotide-based phylogenetic reconstruction of groupers.

(3) Phylogenetic comparison using concatenation-based and coalescent-based methods

For all phylogenetic reconstructions based on nucleotide data, ML and BI yielded identical topologies, and their mean support values were higher than those from NJ, ME, and MSC analyses. Accordingly, the results of ML and BI analyses were used to infer the concatenation-based phylogeny, as summarized in the ML tree (**Figure 4A**). The coalescent-based tree exhibited a topology closely resembling that of the concatenation-based method, with differences mainly confined to *Epinephelus* (**Figure 4B**). However, gene conflict analysis of the coalescent-based tree revealed 14 nodes with both major and minor conflicts (consistency support < 50%), whereas the concatenation-based tree showed relatively high topological consensus with robust support across most nodes, suggesting a more reliable topology for the concatenation-based method.

Phylogenetic analysis

Informed by the results of the phylogenetic comparisons above, a phylogenetic tree of 32 species of the family Epinephelidae was reconstructed using a nucleotide dataset, ML/BI analysis, and a concatenation-based method, yielding a more robust topology with enhanced support and resolving five main clades (**Figure 4A**). The first divergence was observed in the *Plectropomus* clade, followed by the *Variola* clade, each consisting of only a single genus. Then, a more complex clade, the *Cephalopholis-Aethaloperca* clade, comprised five species of *Cephalopholis* and a single species of *Aethaloperca*, *Ae. rogaa*. Finally, two sister clades were defined: the *Epinephelus-Chromileptes-Anyperodon* clade and the *Epinephelus* clade. One sister clade consisted of eight species of *Epinephelus* and two single species of genera, *A. leucogrammicus* and *Ch. altivelis*, while the other clade was composed solely of *Epinephelus*.

Divergence time estimation

According to the benthic foraminifera isotopic records of Waelbroeck et al. (2002) and Miller et al. (2005), global sea levels were redrawn, as shown in **Figure 5A**. From the divergence tree, the Epinephelidae and Serranidae split 62.5 Mya (52.7–70.3 Mya) in the Paleocene (**Figure 5B**).

The crown diversification of Epinephelidae started 58.7 Mya (48.3–67.3 Mya) with the split of the *Plectropomus* clade from the node uniting all other clades. Then, the *Variola* clade separated from the remaining taxa at 47.8 Mya (39.2–55.9 Mya). Subsequently, the *Cephalopholis-Aethaloperca* clade split at 31.7 Mya (25.9–38.7 Mya) in the Oligocene, and species formation occurred at 27.6 Mya (22.1–34.2 Mya). In the end, the separation of the *Epinephelus-Chromileptes-Anyperodon* clade and the *Epinephelus* clade took place 18.8 Mya (14.2–24.1 Mya), resulting in several species groups distributed in the West Atlantic, East Atlantic, West Indian, North Indian, South Indian, West Pacific, and Central Pacific Oceans. Interestingly, the divergences and speciation of *Variola*, *Cephalopholis-Aethaloperca*, and *Epinephelus* occurred in synchrony with pronounced sea level drops during the middle Eocene (48–42 Mya), the late Oligocene (24–28 Mya), and the early to middle Miocene (19–13 Mya), respectively, underscoring the role of paleoceanographic change in shaping lineage splits. In addition, the divergence of all species within each genus occurred after ~15 Mya, emphasizing the recent and rapid diversification of Epinephelidae (**Figure 5B**).

Ancestral area reconstruction

Ancestral area reconstruction was performed with BioGeoBEARS using a nucleotide dataset from 136 OGs of 31 Epinephelidae species excluding the outgroup and *E. marginatus*. Biogeographic inference under the BAYAREALIKE model yielded a significantly better likelihood (–ln L = 71.21) compared to the DEC (–ln L = 72.58) and DIVALIKE (–ln L = 72.10) models (**Figure 6A**; Supplementary Figure S4). Consequently, subsequent analyses and discussion were based on the BAYAREALIKE model. The model inferred that the most recent common ancestor of extant Epinephelidae currently distributed across the Indian and Pacific Oceans originated in the West Pacific Ocean (**Figure 6A**), followed by a radiation of five major clades driven by a series of vicariance events. The basal *Plectropomus* clade was suggested to have initially dispersed across the South Indian, West Pacific, and Central Pacific Oceans. The *Variola* and *Cephalopholis-Aethaloperca* clades exhibited rapid and extensive dispersal, spanning the West, North, and South Indian Oceans as well as the West and Central Pacific Oceans. Furthermore, the *Epinephelus-Chromileptes-Anyperodon* and *Epinephelus* clades, both of West Pacific Ocean origin, experienced intricate evolutionary histories marked by multiple dispersal and vicariance events, highlighting the dynamic biogeographic processes underlying lineage diversification in Epinephelidae.

DISCUSSION

Establishing a robust phylogenetic framework for Epinephelidae using transcriptomic data

Groupers act as ecologically critical predators and apex regulators in coral reef ecosystems, shaping community structure and sustaining resilience across the subtropical and tropical oceans (Skinner et al., 2020). To elucidate the complex evolutionary relationships within Epinephelidae, we conducted a comprehensive phylotranscriptomic analysis and systematically comparing dataset

types (nucleotide and amino acid), inference methods (ML, BI, MSC, NJ, and ME), and gene integration strategies (concatenation and coalescence). These comparisons are particularly relevant for recently radiated taxa, where subtle differences in phylogenetic signal can result in markedly divergent topologies. Our multi-level assessment identified the optimal analytical framework for reconstructing a robust phylogeny of groupers.

The choice between nucleotide and amino acid sequences for phylogenetic reconstruction can influence phylogenetic resolution, particularly for groups with recent radiations (Townsend et al., 2008; Kapli et al., 2023). Our results consistently suggested that nucleotide datasets outperformed amino acid datasets for Epinephelidae, showing higher node support across ML, BI, NJ, and ME analyses, greater gene tree congruence in DensiTree visualizations, and stronger support in MSC trees (**Figure 2**; Supplementary Figure S2). These results reflects the greater information content of nucleotide sequences for recently diverged taxa, whereas amino acid sequences, being more conserved, are more suitable for studying long-term evolutionary patterns (Nei and Kumar, 2000); (Townsend et al., 2008; Kapli et al., 2023). Given that Epinephelidae primarily diversified from the Paleogene (Near et al., 2013; Ma et al., 2016), nucleotide data provide increased resolution for reconstructing their recent evolutionary dynamics. Similar patterns have been observed in other taxa with recent radiations, such as yeast species (Ren et al., 2005) and the planarian genus (Benítez-Álvarez et al., 2023), highlighting the effectiveness of nucleotide data in these cases.

Both ML and BI methods are grounded in probabilistic models in phylogenetic analysis. ML identifies the tree that best explains the data under a given model, while BI accounts for parameter uncertainty by integrating over priors to generate posterior distributions (Kolaczkowski and Thornton, 2009; McCormack et al., 2013; Susko, 2015). In our study, ML and BI approaches yielded congruent topologies and consistently higher nodal support compared to the MSC, NJ, and ME methods (**Figure 3**). This suggests that, despite their theoretical differences and potential limitations such as long-branch attraction, ML and BI both offer higher phylogenetic resolution and robustness when applied to transcriptomic-scale datasets in groupers.

Our results showed that the concatenation-based method provides a more robust phylogenetic framework for groupers compared with the coalescent-based approach (**Figure 4**). Concatenation integrates thousands of loci into a supermatrix, enhancing phylogenetic signal and nodal support (Rokas et al., 2003; Jarvis et al., 2014). Coalescent-based method explicitly accommodates gene tree heterogeneity and incomplete lineage sorting (ILS), but often show reduced resolution and lower nodal support during rapid radiations (Degnan and Rosenberg, 2009). Given that groupers represent a recently diversified lineage (Near et al., 2013; Ma et al., 2016), concatenation is more suitable for reconstructing their phylogeny. Similar patterns have been observed in other rapidly radiating taxa, such as cichlid fishes (Brawand et al., 2014), Darwin's finches (Lamichhaney et al., 2015), and the planarian genus *Dugesia* (Benítez-Álvarez et al., 2023), where concatenation yields stable and well-supported phylogenies.

Our comprehensive phylotranscriptomic comparisons highlighted the critical importance of selecting optimal dataset types and analytical frameworks in reconstructing phylogenies of recent rapidly diversified taxa. Nucleotide sequences provided higher nodal supports and greater genetic informativeness than amino acid sequences across all methods. ML and BI methods yielded more robust topologies than MSC, NJ, and ME approaches. Furthermore, concatenation outperformed coalescent-based method, especially in recently diverged clades. Accordingly, we reconstructed a robust Epinephelidae phylogeny using nucleotide dataset, ML/BI inference, and a concatenation method, offering a stable and well-resolved framework for downstream evolutionary analysis.

First Phylotranscriptomic insights revealing the phylogeny of Epinephelidae

In the present study, we applied phylotranscriptomic analysis to family Epinephelidae for the first time, establishing a complete analytical process and strategy. Using this approach, we recovered a highly supported phylogenetic topology across all nodes based on 108 single-copy orthologs (totaling 146,787 bp) derived from coding regions. Many previous molecular studies have sought to resolve the phylogenetic relationships of Epinephelidae. The earliest work relied on a single marker, the mitochondrial 16S rRNA gene, providing initial insights into grouper phylogeny (Craig et al., 2001). Despite its widespread application in fish phylogenetics (Near et al., 2004; Crottini et al., 2008; Li et al., 2008), the limited sequence length (590 bp) offered insufficient resolution. Subsequent studies incorporated two additional mitochondrial genes (12S, COI) and two nuclear genes (TMO4C4 and Histone H3), increasing the alignment length to over 2,000 bp and yielding improved phylogenetic inference (Craig and Hastings, 2007; Ma et al., 2016). These advances highlighted the importance of both marker number and molecular characteristics. Even if previous phylogenetic studies revealed that intergeneric relationships are typically well supported, some interspecific relationships within genera exhibit comparatively reduced resolution, indicating that the evolutionary signal captured by existing loci is limited in certain clades. This limitation underscored the need for higher-resolution molecular markers to more accurately reconstruct the complex diversification patterns of Epinephelidae. Transcriptomic approaches provide orders of magnitude more phylogenetic information than traditional markers, substantially improving phylogenetic resolution across clades (Dunn et al., 2013; Smith et al., 2015). Though initial analytical processing, including quality control, data cleaning, and ortholog identification, is intensive, standardized pipelines allow scalable and repeatable analyses (Benítez-Álvarez et al., 2023). The advent of large-scale genomic datasets has driven the shift from fixed partition strategies to site-heterogeneous mixture models that infer substitution patterns directly from the data, markedly improving phylogenetic accuracy (Holder and Lewis, 2003; Schrempf et al., 2020). In this study, based on the transcriptomic data, we established a more resolved phylogeny of Epinephelidae compared to previous molecular phylogenetic approaches, particularly in resolving relationships among closely related species. Phylogenetic reconstruction using nucleotide dataset yielded highly supported relationships at both the genus and species

levels (most ML > 90%, all BI = 1). Our findings highlighted the power of phylotranscriptomics in resolving evolutionary histories and provided a robust molecular framework for understanding grouper diversification.

The advent and application of high-throughput transcriptomic sequencing have enabled the generation of a more robust and highly resolved phylogeny for the Epinephelidae, characterized by enhanced nodal support and improved topological stability. Although earlier phylogenetic studies have provided important insights, persistent topological conflicts remain regarding the placement of certain genera and species. For example, recent molecular studies have yielded inconsistent results concerning which lineage represents the earliest divergence within the family. Most prior analyses have suggested that *Plectropomus* represents the earliest diverging lineage in the Epinephelidae (Craig et al., 2001; Ding et al., 2006; Craig and Hastings, 2007; Qu et al., 2018; Liang et al., 2020), while a molecular dating analysis of Ma et al. (2016) revealed that *Variola* is the first divergent genus within the *Variola-Plectropomus-Saloptia-Gonioplectrus* clade. In the present study, our phylotranscriptomic analysis supported *Plectropomus* as the first divergent lineage, followed by *Variola*. Developmental observations reveal that *Plectropomus* occupies a basal position within Epinephelidae, serving as the sister group to all other taxa, with dorsal fin spine formation providing key morphological support (Leis, 1986). The occurrence of 7 or 8 spines is considered the ancestral condition for Epinephelidae, whereas the evolution of 10 or 11 spines reflects a derived trait (Leis, 1986; Johnson, 1988). Morphological records indicate that *Plectropomus* bears 7 or 8 slender dorsal-fin spines, in contrast to the 9 spines in *Variola* (Craig et al., 2011). Combined molecular and morphological evidence thus identifies *Plectropomus* as the earliest diverging lineage and the sister group to the remaining Epinephelidae. The phylogenetic relationships among genera and species within the *Cephalopholis-Aethaloperca* and *Epinephelus-Chromileptes-Anyperodon* clades were consistent with the results of Ma et al. (2016). However, topological conflicts were observed within the *Epinephelus* clade (II, **Figure 4A**) between our phylotranscriptomic tree and that of Ma et al. (2016), despite some nodes receiving over 64% support under the ML framework. The observed incongruence likely reflects a recent burst of diversification within *Epinephelus*. In addition, our phylotranscriptomic results strongly supported the non-monophyly of *Cephalopholis* and *Epinephelus*: *Cephalopholis* grouped closely with *Aethaloperca*, while *Epinephelus* clustered with *Chromileptes* and *Anyperodon*. Early morphological evidence has also suggested a sister-group relationship between *Cephalopholis* and *Aethaloperca*. Both genera are nine-spined groupers that share trisegmental pterygiophores and homologous dorsal-fin spines, while uniquely retaining the ancestral fused scalelet (Leis, 1986; Smith-Vaniz et al., 1988; Craig et al., 2001). In this study, our phylotranscriptomic results placed *Epinephelus*, *Chromileptes*, and *Anyperodon* in a paraphyletic relationship, in agreement with previous molecular phylogenetic studies (Craig and Hastings, 2007; Ma et al., 2016; Qu et al., 2018; Liang et al., 2020). Morphological evidence indicates that these genera share numerous

common traits (Baldwin et al., 1993; Craig et al., 2011), yet *Chromileptes* displays a high dorsal head profile and *Anyperodon* lacks palatine teeth (Smith-Vaniz et al., 1988; Baldwin and Johnson, 1993). Together, these findings suggest that *Chromileptes* and *Anyperodon* represent morphologically specialized lineages within *Epinephelus* (Ding et al., 2006; Craig and Hastings, 2007).

Although phylotranscriptomic approaches generally provide higher-resolution phylogenetic topologies and stronger nodal support compared with traditional methods, topological conflicts still persist among certain species within specific genera. Topological conflicts among the phylogenetic trees were primarily observed within *Epinephelus*-*Chromileptes*-*Anyperodon* clade and *Epinephelus* clade. Gene tree discordance across the coalescent-based topology highlighted pervasive topological heterogeneity at internal nodes within these two clades, underscoring pronounced genealogical variation (Rancilhac et al., 2021). The relatively recent diversification of groupers may have contributed to the observed topological conflicts or incongruence, potentially due to incomplete lineage sorting, interspecific hybridization, gene duplication events, and horizontal gene transfer (Mallo and Posada, 2016; Arcila et al., 2017). Furthermore, topological incongruence among genes may also arise from estimation errors within individual gene alignments, such as missing sequence data, phylogenetic noise, and long-branch attraction (Mallo and Posada, 2016; Meng et al., 2021).

Phylotranscriptomics providing new insights into the evolutionary history of Epinephelidae

The evolutionary history of coral reef fishes exhibits remarkable diversity, shaped by the intricate interplay of ecological dynamics, geological events, and biogeographic processes (Cowman and Bellwood, 2011; Siqueira et al., 2020). Our study suggested that the Epinephelidae originated in the Paleocene (~58.7 Mya), a timing older than the estimate proposed by Near et al. (2013) (~50 Mya, early Eocene) and Ma et al. (2016) (~36 Mya, late Eocene). The origin and early diversification of the family appear to have been shaped by increased ecological opportunity and elevated primary productivity during the Paleogene (Ma et al., 2016). On one hand, the global climatic shift from a greenhouse to an icehouse Earth system, resulting in a substantial decline in atmospheric CO_2, the formation of glaciation, and declining sea levels (Zachos et al., 2001; Livermore et al., 2007; Lear et al., 2008). These changes led to widespread extinction of warm-water adapted marine taxa, as reflected by dramatic declines in fossil diversity and supported by phylogenetic evidence from several coral reef fish families (Rohde and Muller, 2005; Cowman and Bellwood, 2011; Ma et al., 2016), thereby providing substantial niches for Epinephelidae. On the other hand, intensified oceanic upwelling during this period drove a marked increase in marine primary productivity, which likely promoted greater trophic complexity within marine ecosystems (Scher and Martin, 2006; Livermore et al., 2007; Norris et al., 2013), further favouring the speciation and diversification of the Epinephelidae lineage.

Interestingly, our findings suggested that the diversification and formation of major lineages

within the Epinephelidae may have been closely associated with repeated sea level declines from the Eocene to the Miocene. Previous studies have demonstrated that historical barriers, particularly those resulting from eustatic sea level regressions, can play a pivotal role in promoting population isolation by disrupting connectivity among marine basins (McManus, 1985). During glacial periods, dramatic drops in sea level exposed continental shelves, giving rise to land barriers and creating refugial habitats (Renema et al., 2008; Pellissier et al., 2014). Coral reef fishes likely dispersed into these isolated refugia, resulting in population fragmentation and intergeneric diversification (Pellissier et al., 2014). From the Eocene to the Miocene, global sea level declines exceeding 60 m occurred approximately 48–42 Mya, 24–28 Mya, and 19–13 Mya (**Figure 5A**) (Waelbroeck et al., 2002; Miller et al., 2005), leading to the fragmentation of reef systems and the emergence of marked biogeographic discontinuities across shallow marine habitats. These recurrent episodes of sea level regression likely served as key historical drivers of vicariance, facilitating the divergence of multiple major lineages within the Epinephelidae.

The pronounced diversification of Epinephelidae after ~18 Mya coincided with major paleoceanographic and ecological shifts in the Indo-Pacific. Following the Middle Miocene Climate Transition (MMCT, ~14 Mya), global cooling and sea level fluctuations restructured shallow marine environments, while tectonic processes associated with the formation of the Indo-Australian Archipelago (IAA) created an unparalleled network of coral reef habitats (Renema et al., 2008; Bellwood and Meyer, 2009). These changes opened extensive ecological niches, promoting resource partitioning and adaptive radiation among early grouper lineages. Concurrently, the establishment of the Indonesian Throughflow and intensification of the South Equatorial Current enhanced larval dispersal across the Indo-Pacific, whereas intermittent barriers facilitated allopatric divergence (Cowman and Bellwood, 2013; Makarim et al., 2019). After the transient warming of the Middle Miocene Climatic Optimum (MMCO, ~17–15 Mya), Earth entered the MMCT, a phase marked by progressive cooling and expansion of the Antarctic ice sheet (Shevenell et al., 2004). Reconstructions of oceanic circulation have revealed a sustained decline in tropical Pacific sea surface temperatures since ~16 Mya, without apparent MMCO warming, likely due to intensified upwelling (Kamikuri and Moore, 2017). Previous paleoceanographic records reveal a pronounced rise in primary productivity during ~16–14 Mya (Shipboard Scientific Party, 1989; Lyu et al., 2023), expanding ecological capacity and setting the stage for diversification within coral reef ecosystems. Collectively, these climatic and oceanographic shifts boosted primary productivity and expanded reef habitats, fostering ecological opportunities and driving the rapid radiation of groupers.

Our biogeographic reconstruction analysis, based on the grouper samples collected in this study, suggests that Epinephelidae lineages currently distributed across the Indian and Pacific Oceans may have originated in the West Pacific Ocean **(Figure 6)**, a region known for its remarkable biological richness, particularly in the IAA (Renema et al., 2008; Bellwood and

Meyer, 2009). The high diversity and climatic stability of the IAA likely served as an evolutionary reservoir, enabling the persistence and diversification of early grouper lineages (Hoeksema, 2007). Consistent with our findings, previous studies have also shown the IAA as a major center of origin and diversification for reef fishes, including wrasses (Labridae), damselfishes (Pomacentridae), and butterflyfishes (Chaetodontidae) (Cowman and Bellwood, 2013). The biogeographic distribution of Epinephelidae is likely shaped by repeated long-distance dispersal events, mediated by prevailing ocean currents throughout the Indo-Pacific region (Cowman and Bellwood, 2013). Originating in the West Pacific, the *Plectropomus*, *Variola*, and *Cephalopholis-Aethaloperca* clades underwent early lineage diversification and spatial expansion, likely driven by prevailing oceanic currents (**Figure 6B**). The westward South Equatorial Current and Indonesian Throughflow likely facilitated larval export into the Indian Ocean basin, supporting range extension into the East and West Indian Oceans (Gordon et al., 2012; Pang et al., 2021). Simultaneously, the splitting of the South Equatorial Current into the East Australian and New Guinea Coastal Currents may have driven eastward dispersal into the Central Pacific Ocean (Cetina-Heredia et al., 2014; Bull et al., 2017). Coupled with the extended pelagic larval durations typical of groupers (Victor, 1986), these oceanographic corridors enabled extensive connectivity and genetic exchange across vast seascapes. The intricate dispersal and vicariance history of the *Epinephelus* and *Epinephelus-Chromileptes-Anyperodon* clades further highlights the dynamic role of current systems in driving phylogeographic complexity. Overall, ocean currents may function both as pathways and isolating forces, underpinning the complex biogeographic structure and diversification of the Epinephelidae lineage (Cowman and Bellwood, 2011).

CONCLUSIONS

This study represents the first application of phylotranscriptomic analysis to Epinephelidae, establishing a comprehensive analytical process and strategy. Leveraging transcriptomic data, phylogenetic reconstruction using the concatenated nucleotide dataset with ML/BI methods yielded substantially enhanced resolution, stronger nodal support, and improved topological stability compared with previous single- or multi-locus molecular studies. Our results corroborate several well-supported phylogenetic relationships reported in earlier studies, including the non-monophyly of *Epinephelus* and *Cephalopholis*, while further clarifying their sister-group relationships. By integrating transcriptomic evidence with key morphological characters, we resolve longstanding conflict regarding the earliest diverging lineage of Epinephelidae, robustly supporting *Plectropomus* as the basal lineage. We infer that the origin and early diversification of groupers currently distributed across the Indian and Pacific Oceans may have occurred in the West Pacific during the Paleocene, followed by global radiation facilitated by major oceanic circulation systems. Moreover, repeated glacial sea-level fluctuations appear to have played a central role in shaping lineage divergence and geographic isolation of major clades, while subsequent

enhancement of upwelling intensity and primary productivity may have promoted recent episodes of rapid diversification. In conclusion, our study refines the evolutionary and biogeographic framework of Epinephelidae, resolves key phylogenetic uncertainties, and demonstrates the distinct power of phylotranscriptomics for disentangling complex evolutionary histories in reef fishes.

FUNDING

This work was supported by the National Natural Science Foundation of China (32273136), Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (SML2024SP001), Science and Technology Planning Project of Guangdong Province (2023B1212060023), Agriculture Research System of China (ARS-47), 2025 Guangdong Provincial Dedicated Funds for State Key Laboratories, Natural Science Foundation of Hainan Province (324QY575), Hainan Province "Nan hai xin xing" Science and Technology Innovation Talent Platform Project (NHXXRCXM202374), and Hainan Agricultural Production Research System (HNARS-03-Z01).

DATA AVAILABILITY

All raw data were deposited in the NCBI Sequence Read Archive (SRA) with accession number PRJNA1344424. All scripts and detailed bioinformatic methodologies used in this study are publicly available at https://github.com/QinghuaWang-sysu/Phylotranscriptomics-of-groupers.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Q.H.W.: Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft. W.W.Z.: Investigation, Formal analysis, Visualization. Z.Y.W.: Investigation, Formal analysis, Visualization. Y.F.C.: Investigation, Data curation. X.W.: Investigation, Formal analysis. L.W.: Investigation, Formal analysis. Y.Y.: Investigation, Data curation, Funding acquisition, Writing – review & editing. Z.N.M.: Conceptualization, Funding acquisition, Supervision, Investigation, Resources, Writing – review & editing. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We gratefully thank the reviewers for their important suggestions to improve the paper. We

sincerely acknowledge all study participants who contributed to this study. We thank Tong Wang, Duo Li, Lei-Lei Zeng, Qi-Chuang Wei, and Chu Lin for their advice on phylogenetic analysis; Xin Wen, Li-Na Wu, Wen-Hua Hunag, Yu-Xin Zhang, Fan-Ming Guo, Qing-Xin Ruan, and Si-Jie Hua for their help on sample collection.

REFERENCES

Arcila D, Ortí G, Vari R, Armbruster JW, Stiassny MLJ, Ko KD, et al. 2017. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nature Ecology & Evolution*, **1**: 0020.

Baldwin C, Johnson GD. 1993. Phylogeny of the Epinephelinae (Teleostei: Serranidae). *Bulletin of Marine Science*, **52**: 240-283.

Bellwood DR, Hughes TP. 2001. Regional-scale assembly rules and biodiversity of coral reefs. *Science*, **292**: 1532-1535.

Bellwood DR, Meyer CP. 2009. Searching for heat in a marine biodiversity hotspot. *Journal of Biogeography*, **36**: 569-576.

Benítez-Álvarez L, Leria L, Fernández R, Mateos E, El Ouanighi Y, Bennas N, et al. 2023. Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean. *Molecular Phylogenetics and Evolution*, **178**: 107649.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**: 2114-2120.

Bouckaert RR, Heled J. 2014. DensiTree 2: seeing trees through the forest. *bioRxiv*: 012401.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**: 375-381.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**: 59-60.

Bull CYS, Kiss AE, Jourdain NC, England MH, van Sebille E. 2017. Wind forced variability in eddy formation, eddy shedding, and the separation of the East Australian Current. *Journal of Geophysical Research: Oceans*, **122**: 9980-9998.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**: 1972-1973.

Carruthers M, Yurchenko AA, Augley JJ, Adams CE, Herzyk P, Elmer KR. 2018. De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics*, **19**: 32.

Cetina-Heredia P, Roughan M, van Sebille E, Coleman MA. 2014. Long-term trends in the East Australian Current separation latitude and eddy driven transport. *Journal of Geophysical Research: Oceans*, **119**: 4351-4366.

Chen L, Jin WT, Liu XQ, Wang XQ. 2022. New insights into the phylogeny and evolution of

682 Podocarpaceae inferred from transcriptomic data. *Molecular Phylogenetics and Evolution*,
683 **166**: 107341.
684 Choat JH. 2012. Spawning aggregations in reef fishes; Ecological and Evolutionary Processes. *Reef*
685 *Fish Spawning Aggregations: Biology, Research and Management*: 85-116.
686 Ciezarek AG, Osborne OG, Shipley ON, Brooks EJ, Tracey SR, McAllister JD, et al. 2019.
687 Phylotranscriptomic insights into the diversification of endothermic *Thunnus* tunas. *Molecular*
688 *Biology and Evolution*, **36**: 84-96.
689 Cowman PF. 2014. Historical factors that have shaped the evolution of tropical reef fishes: a review of
690 phylogenies, biogeography, and remaining questions. *Frontiers in Genetics*, **5**: 394.
691 Cowman PF, Bellwood DR. 2011. Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic
692 extinction events, and the development of biodiversity hotspots. *Journal of Evolutionary*
693 *Biology*, **24**: 2543-2562.
694 Cowman PF, Bellwood DR. 2013. The historical biogeography of coral reef fishes: global patterns of
695 origination and dispersal. *Journal of Biogeography*, **40**: 209-224.
696 Cox CJ, Li B, Foster PG, Embley TM, Civáň P. 2014. Conflicting phylogenies for early land plants are
697 caused by composition biases among synonymous substitutions. *Systematic Biology*, **63**:
698 272-279.
699 Craig M, Sadovy DMY, Heemstra P. 2011. Groupers of the World: A Field and Market Guide.
700 Grahamstown: NISC (Pty) Ltd.
701 Craig MT, Hastings PA. 2007. A molecular phylogeny of the groupers of the subfamily Epinephelinae
702 (Serranidae) with a revised classification of the Epinephelini. *Ichthyological Research*, **54**:
703 1-17.
704 Craig MT, Pondella DJ, Franck JPC, Hafner JC. 2001. On the status of the Serranid fish genus
705 *Epinephelus*: evidence for paraphyly based upon 16S rDNA sequence. *Molecular*
706 *Phylogenetics and Evolution*, **19**: 121-130.
707 Crottini A, Marotta R, Barbuto M, Casiraghi M, Ferraguti M. 2008. The world in a river? A
708 preliminary analysis of the 16S rDNA variability of *Tubifex* species (Clitellata: Tubificidae)
709 from the Lambro River. *Molecular Phylogenetics and Evolution*, **48**: 1189-1203.
710 da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, et al. 2016.
711 Next-generation biology: Sequencing and data analysis approaches for non-model organisms.
712 *Marine Genomics*, **30**: 3-13.
713 Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies
714 coalescent. *Trends in Ecology & Evolution*, **24**: 332-340.
715 Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life.
716 *Nature Reviews Genetics*, **6**: 361-375.
717 Ding S, Zhuang X, Guo F, Wang J, Su Y, Zhang Q, et al. 2006. Molecular phylogenetic relationships
718 of China Seas groupers based on cytochrome b gene fragment sequences. *Science China Life*

Sciences, **49**: 235-242.

Ding SX, Liu QH, Wu HH, Qu M. 2018. A review of research advances on the biology and artificial breeding of groupers. *Journal of Fishery Sciences of China*, **25**: 737-752.

Du Y, Wu S, Edwards SV, Liu L. 2019. The effect of alignment uncertainty, substitution models and priors in building and dating the mammal tree of life. *BMC Evolutionary Biology*, **19**: 203.

Dunn CW, Howison M, Zapata F. 2013. Agalma: an automated phylogenomics workflow. *BMC Bioinformatics*, **14**: 330.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, **16**: 157.

Erisman BE, Craig MT, Hastings PA. 2009. A Phylogenetic Test of the Size-Advantage Model: Evolutionary Changes in Mating Behavior Influence the Loss of Sex Change in a Fish Lineage. *The American Naturalist*, **174**: 83-99.

Gomes AS, Passos LS, Rocha Aride PH, Chisté B, Gomes LC, Boldrini-França J. 2022. Gene expression changes in *Epinephelus marginatus* (Teleostei, Serranidae) liver reveals candidate molecular biomarker of iron ore contamination. *Chemosphere*, **303**: 134899.

Gordon AL, Huber BA, Metzger EJ, Susanto RD, Hurlburt HE, Adi TR. 2012. South China Sea throughflow impact on the Indonesian throughflow. *Geophysical Research Letters*, **39**.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**: 644-652.

Haas B, Papanicolaou A. 2019. TransDecoder 5.5.0. Retrieved from https://github.com/TransDecoder/TransDecoder/wiki.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**: 1494-1512.

Heemstra PC, Randall J. 1993. Groupers of the world. FAO Species Catalogue 16. Food and Agriculture Organization of the United Nations, Rome 16, Rome.

Hoeksema BW. 2007. Delineation of the Indo-Malayan centre of maximum marine biodiversity: The Coral Triangle. In Biogeography, Time and Place: Distributions, Barriers and Islands. *Springer*: 117-178.

Holder M, Lewis PO. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, **4**: 275-284.

Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, et al. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences*, **115**: 6249-6254.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**: 1320-1331.

Johnson GD. 1988. *Niphon spinosus*, a primitive epinepheline serranid: corroborative evidence from the larvae. *Japanese Journal of Ichthyology*, **35**: 7–18.

Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, **26**: 1669-1670.

Kadlec M, Bellstedt DU, Le Maitre NC, Pirie MD. 2017. Targeted NGS for species level phylogenomics: "made to measure" or "one size fits all"? *PeerJ*, **5**: e3569.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**: 587-589.

Kamikuri S-i, Moore TC. 2017. Reconstruction of oceanic circulation patterns in the tropical Pacific across the early/middle Miocene boundary as inferred from radiolarian assemblages. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **487**: 136-148.

Kapli P, Kotari I, Telford MJ, Goldman N, Yang Z. 2023. DNA sequences are as useful as protein sequences for inferring deep phylogenies. *Systematic Biology*, **72**: 1119-1135.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**: 772-780.

Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, et al. 2017. Phylogenomics of lophotrochozoa with consideration of systematic error. *Systematic Biology*, **66**: 256-282.

Kolaczkowski B, Thornton JW. 2009. Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS One*, **4**: e7891.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**: 371-375.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**: 772-773.

Lear CH, Bailey TR, Pearson PN, Coxall HK, Rosenthal Y. 2008. Cooling and ice growth across the Eocene-Oligocene transition. *Geology*, **36**: 251-254.

Leis JM. 1986. Larval development in four species of Indo-Pacific coral trout *Plectropomus* (Pisces: Serranidae: Epinephelinae) with an analysis of the relationships of the genus. *Bulletin of Marine Science*, **38**: 525-552.

Li J, Wang X, Kong X, Zhao K, He S, Mayden RL. 2008. Variation patterns of the mitochondrial 16S rRNA gene with secondary structure constraints and their application to phylogeny of cyprinine fishes (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution*, **47**: 472-487.

Li J, Lemer S, Kirkendale L, Bieler R, Cavanaugh C, Giribet G. 2020. Shedding light: a phylotranscriptomic perspective illuminates the origin of photosymbiosis in marine bivalves. *BMC Evolutionary Biology*, **20**: 50.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**: 1658-1659.

Liang R, Chen M, Liao G, Zhang Z, Zhang G, Chen Y, et al. 2020. Molecular phylogenetic relationships of 40 species of Epinephelinae based on the partial sequences of 16S rRNA and COI genes. *Haiyang Xuebao*, **42**: 9-19.

Liu XQ, Xia XM, Chen L, Wang XQ. 2022. Phylogeny and evolution of Cupressaceae: Updates on intergeneric relationships and new insights on ancient intergeneric hybridization. *Molecular Phylogenetics and Evolution*, **177**: 107606.

Livermore R, Hillenbrand C-D, Meredith M, Eagles G. 2007. Drake Passage and Cenozoic climate: An open and shut case? *Geochemistry, Geophysics, Geosystems*, **8**.

Lyu J, Auer G, Bialik OM, Christensen B, Yamaoka R, De Vleeschouwer D. 2023. Astronomically-paced changes in paleoproductivity, winnowing, and mineral flux over broken ridge (Indian Ocean) since the early Miocene. *Paleoceanography and Paleoclimatology*, **38**: e2023PA004761.

Ma KY, Craig MT, Choat JH, van Herwerden L. 2016. The historical biogeography of groupers: Clade diversification patterns and processes. *Molecular Phylogenetics and Evolution*, **100**: 21-30.

Maggio T, Andaloro F, Hemida F, Arculeo M. 2005. A molecular analysis of some Eastern Atlantic grouper from the *Epinephelus* and *Mycteroperca* genus. *Journal of Experimental Marine Biology and Ecology*, **321**: 83-92.

Makarim S, Sprintall J, Liu Z, Yu W, Santoso A, Yan X-H, et al. 2019. Previously unidentified Indonesian Throughflow pathways and freshening in the Indian Ocean during recent decades. *Scientific Reports*, **9**: 7364.

Mallo D, Posada D. 2016. Multilocus inference of species trees and DNA barcoding. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **371**: 20150335.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. *Current Protocols*, **1**: e323.

Matzke NJ. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, **5**: 242–248.

Matzke NJ. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, **63**: 951-970.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One*, **8**: e54848.

McManus J, 1985. Marine speciation, tectonics and sea-level changes in Southeast Asia. *In: Proceedings of the Fifth International Coral Reef Congress*, 133–138.

Meng R, Luo LY, Zhang JY, Zhang DG, Nie ZL, Meng Y. 2021. The deep evolutionary relationships

of the morphologically heterogeneous Nolinoideae (Asparagaceae) revealed by transcriptome data. *Frontiers in Plant Science*, **11**: 584981.

Miller KG, Kominz MA, Browning JV, Wright JD, Mountain GS, Katz ME, et al. 2005. The Phanerozoic record of global sea-level change. *Science*, **310**: 1293-1298.

Near TJ, Pesavento JJ, Cheng C-HC. 2004. Phylogenetic investigations of Antarctic notothenioid fishes (Perciformes: Notothenioidei) using complete gene sequences of the mitochondrial encoded 16S rRNA. *Molecular Phylogenetics and Evolution*, **32**: 881-891.

Near TJ, Dornburg A, Eytan RI, Keck BP, Smith WL, Kuhn KL, et al. 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **110**: 12738-12743.

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press Inc., New York.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Molecular Biology and Evolution*, **32**: 268-274.

Norris RD, Turner SK, Hull PM, Ridgwell A. 2013. Marine ecosystem responses to Cenozoic global change. *Science*, **341**: 492-498.

Pang X, Bassinot F, Sepulcre S. 2021. Indonesian Throughflow variability over the last two glacial-interglacial cycles: Evidence from the eastern Indian Ocean. *Quaternary Science Reviews*, **256**: 106839.

Pellissier L, Leprieur F, Parravicini V, Cowman PF, Kulbicki M, Litsios G, et al. 2014. Quaternary coral reef refugia preserved fish diversity. *Science*, **344**: 1016-1019.

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, **35**: D61-D65.

Qu M, Tang W, Liu Q, Wang D, Ding S. 2018. Genetic diversity within grouper species and a method for interspecific hybrid identification using DNA barcoding and RYR3 marker. *Molecular Phylogenetics and Evolution*, **121**: 46-51.

Rambaut A. 2018. Figtree, a graphical viewer of phylogenetic trees. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**: 901-904.

Rancilhac L, Irisarri I, Angelini C, Arntzen JW, Babik W, Bossuyt F, et al. 2021. Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. *Molecular Phylogenetics and Evolution*, **155**: 106967.

Ren F, Tanaka H, Yang Z. 2005. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. *Systematic Biology*, **54**: 808-818.

867 Renema W, Bellwood DR, Braga JC, Bromfield K, Hall R, Johnson KG, et al. 2008. Hopping hotspots:
868 global shifts in marine biodiversity. *Science*, **321**: 654-657.
869 Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things).
870 *Methods in Ecology and Evolution*, **3**: 217-223.
871 Rocha LA, Robertson DR, Roman J, Bowen BW. 2005. Ecological speciation in tropical reef fishes.
872 *Proceedings of the Royal Society B*, **272**: 573-579.
873 Rohde RA, Muller RA. 2005. Cycles in fossil diversity. *Nature*, **434**: 208-210.
874 Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving
875 incongruence in molecular phylogenies. *Nature*, **425**: 798-804.
876 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. 2012. MrBayes 3.2:
877 efficient Bayesian phylogenetic inference and model choice across a large model space.
878 *Systematic Biology*, **61**: 539-542.
879 Scher HD, Martin EE. 2006. Timing and climatic consequences of the opening of Drake Passage.
880 *Science*, **312**: 428-430.
881 Schrempf D, Lartillot N, Szöllősi G. 2020. Scalable empirical Mixture Models that account for
882 across-site compositional heterogeneity. *Molecular Biology and Evolution*, **37**: 3616-3631.
883 Schultz O. 2000. Ein Zackenbarsch (*Epinephelus*, Serranidae, Pisces) aus dem Mittel-Miozän von
884 Retznei, Steiermark. *Joannea Geol. Paläont.*, **2**: 5-56.
885 Shevenell AE, Kennett JP, Lea DW. 2004. Middle Miocene southern ocean cooling and Antarctic
886 cryosphere expansion. *Science*, **305**: 1766-1770.
887 Shipboard Scientific Party. 1989. Site 752. In J. Peirce, J. Weissel, E. Taylor, J. Alt, J. Dehn, N.
888 Driscoll, et al. (Eds.), Ocean Drilling Program. *Proceedings of the ODP, Initial Reports*, **121**:
889 111–169.
890 Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, et al. 2017. A Large and
891 Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals.
892 *Current Biology*, **27**: 958-967.
893 Siqueira AC, Morais RA, Bellwood DR, Cowman PF. 2020. Trophic innovations fuel reef fish
894 diversification. *Nature Communications*, **11**: 2669.
895 Skinner C, Mill AC, Newman SP, Alsagoff SN, Polunin NVC. 2020. The importance of oceanic atoll
896 lagoons for coral reef predators. *Marine Biology*, **167**: 19.
897 Smith-Vaniz WF, Johnson GD, Randall JE. 1988. Redescription of *Gracila albomarginata* (Fowler and
898 Bean) and *Cephalopholis polleni* (Bleeker) with comments on the generic limits of selected
899 Indo-Pacifi c groupers (Pisces: Serranidae: Epinephelinae). *The Proceedings of the Academy*
900 *of Natural Sciences of Philadelphia*, **140**: 1-23.
901 Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict,
902 concordance, and gene duplications with examples from animals and plants. *BMC*
903 *Evolutionary Biology*, **15**: 150.

Struck TH. 2013. The Impact of Paralogy on Phylogenomic Studies – A Case Study on Annelid Relationships. *PLoS One*, **8**: e62892.

Susko E. 2015. Bayesian long branch attraction bias and corrections. *Systematic Biology*, **64**: 243-255.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**: W609-W612.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, **38**: 3022-3027.

Tong C, Fei T, Zhang C, Zhao K. 2017. Comprehensive transcriptomic analysis of Tibetan Schizothoracinae fish *Gymnocypris przewalskii* reveals how it adapts to a high altitude aquatic life. *BMC Evolutionary Biology*, **17**: 74.

Townsend JP, López-Giráldez F, Friedman R. 2008. The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. *Journal of Molecular Evolution*, **67**: 437-447.

Victor BC. 1986. Larval settlement and juvenile mortality in a recruitment-limited coral reef fish population. *Ecological Monographs*, **56**: 145-160.

Waelbroeck C, Labeyrie L, Michel E, Duplessy JC, McManus JF, Lambeck K, et al. 2002. Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quaternary Science Reviews*, **21**: 295-305.

Xiang CY, Gao F, Jakovlić I, Lei HP, Hu Y, Zhang H, et al. 2023. Using PhyloSuite for molecular phylogeny and tree-based analyses. *iMeta*, **2**: e87.

Yang Y, Wu LN, Chen JF, Wu X, Xia JH, Meng ZN, et al. 2020. Whole-genome sequencing of leopard coral grouper (*Plectropomus leopardus*) and exploration of regulation mechanism of skin color and adaptive evolution. *Zoological Research*, **41**: 328-340.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**: 1586-1591.

Young AD, Gillung JP. 2020. Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, **45**: 225-247.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**: 686-693.

Zhang WW, Weng ZY, Wang X, Yang Y, Li D, Wang L, et al. 2024. Genetic mechanism of body size variation in groupers: Insights from phylotranscriptomics. *Zoological Research*, **45**: 314-328.

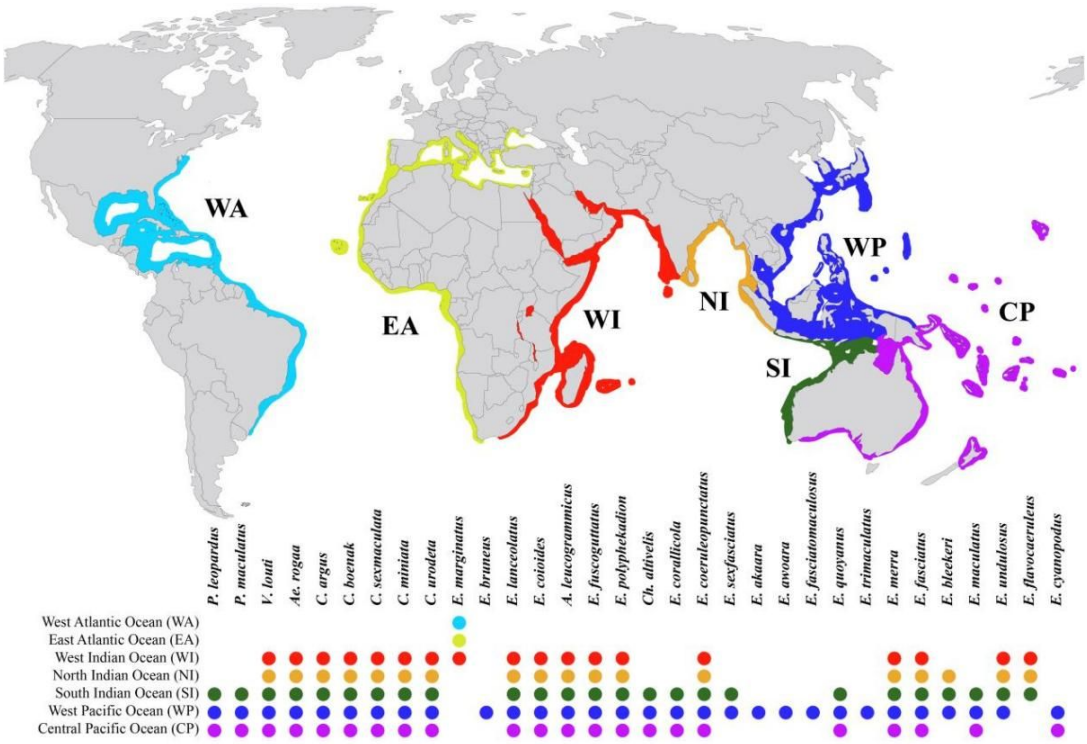
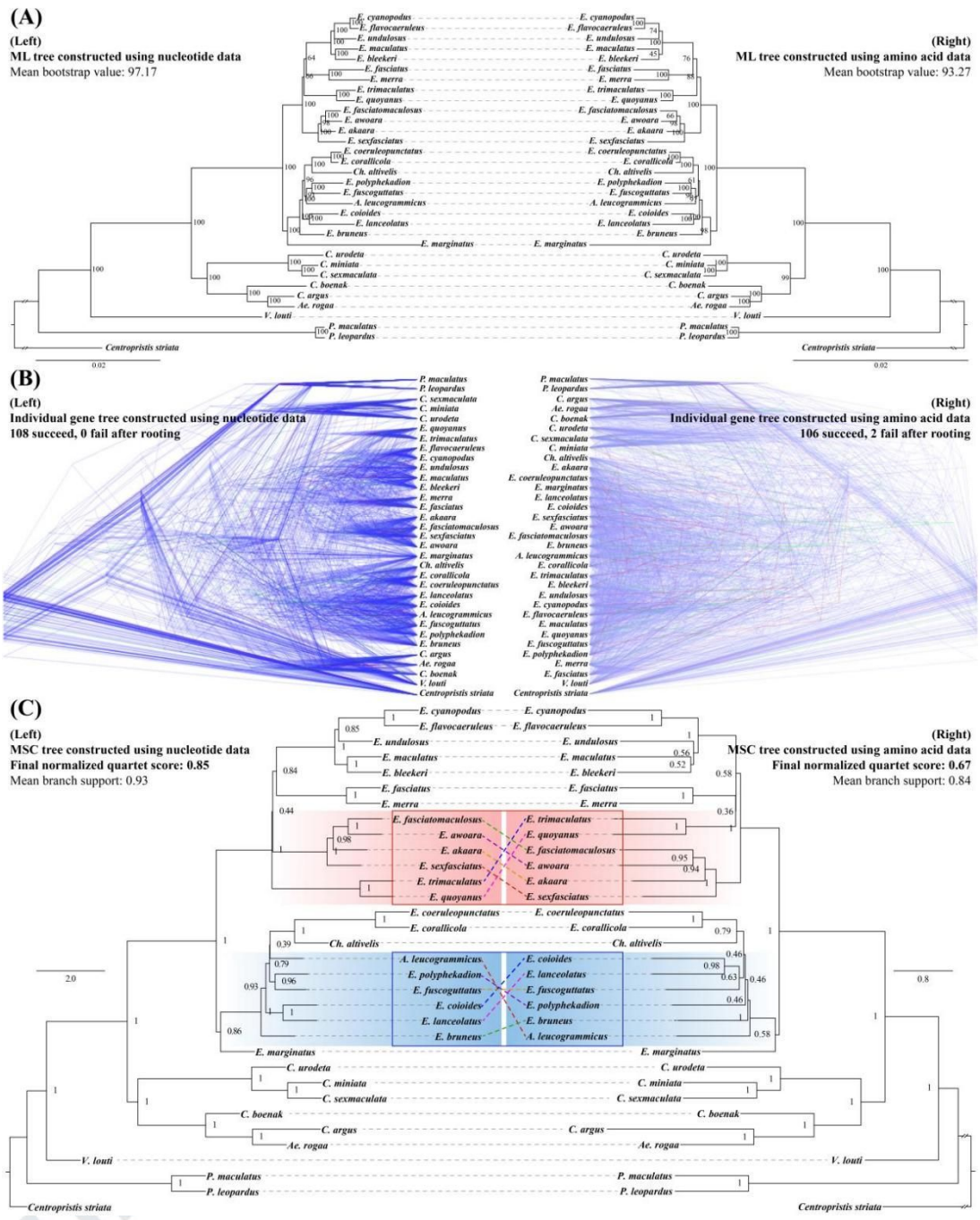


Figure 1 Biogeographic distribution of Epinephelidae. According to Craig et al. (2011) and Ma et al. (2016), seven sea areas were defined, including West Atlantic Ocean (WA), East Atlantic Ocean (EA), West Indian Ocean (WI), North Indian Ocean (NI), South Indian Ocean (SI), West Pacific Ocean (WP), and Central Pacific Ocean (CP).



943

944 **Figure 2** Phylogenetic comparison based on the same inference method using concatenated
945 nucleotide (left) and amino acid (right) datasets. (A) ML method, number on the branch showing
946 the bootstrap value. (B) DensiTree cloudogram inferred from 108 gene trees. (C) MSC tree,
947 number on the branch showing the branch support.

948

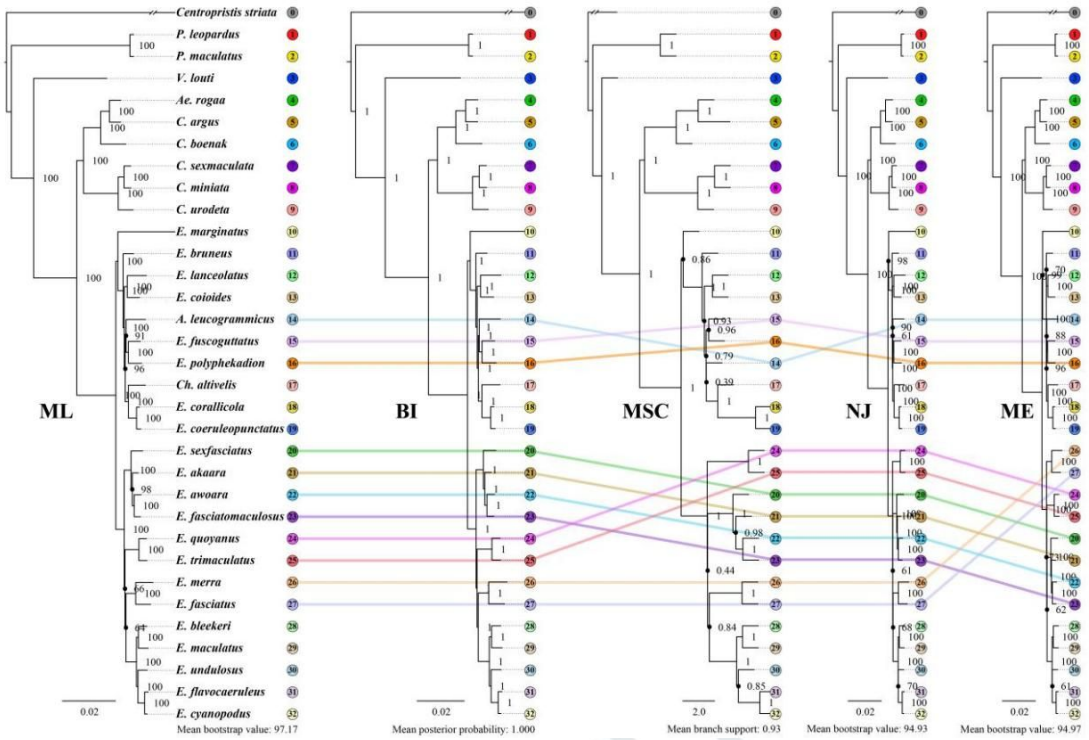


Figure 3 Phylogenetic comparison based on the concatenated nucleotide dataset using different inference methods, including ML, BI, MSC, NJ, and ME. Node circles indicate bootstrap support values below 100.

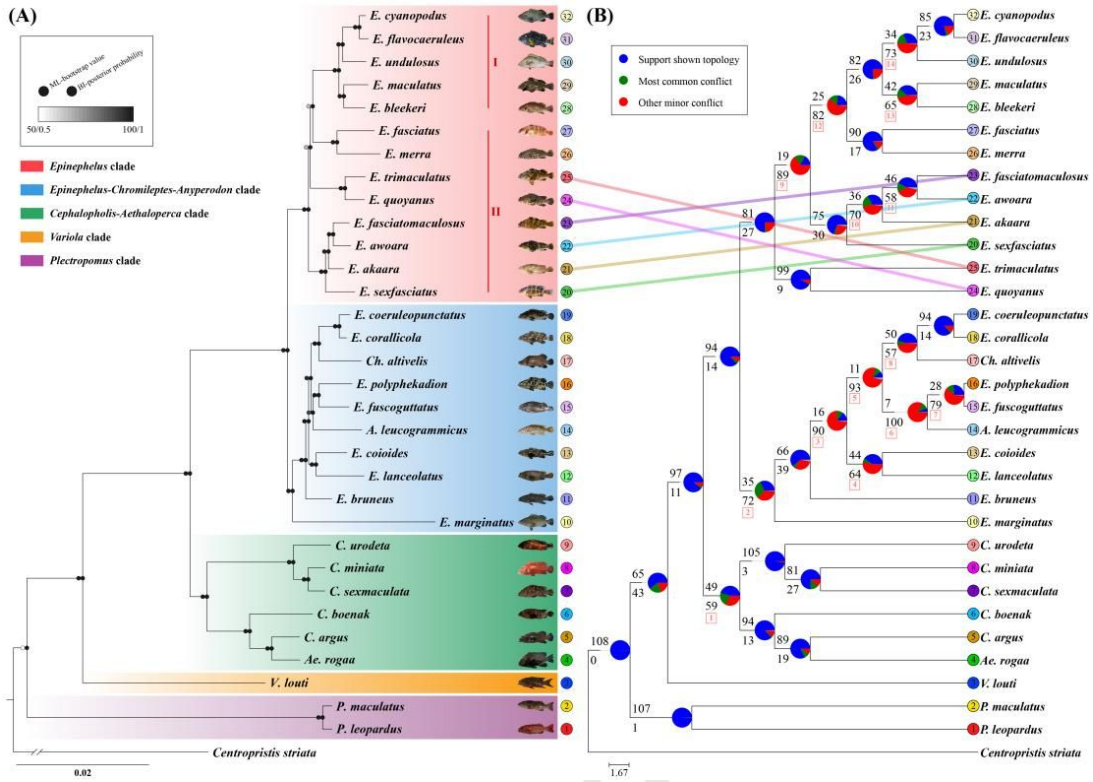


Figure 4 Phylogenetic comparison based on the nucleotide dataset using different gene integration strategies. (A) Concatenation-based tree yield the same topology using ML and BI methods, presented with the ML tree, and circles at nodes representing the bootstrap values and posterior probabilities, respectively. (B) Coalescent-based tree showing discordance among genes. At each node, the pie charts illustrate proportion of genes in concordant topology (blue), supporting the main alternative topology (green), and supporting other alternative topologies (red). Numbers above and below branches represent genes supporting or conflicting with the species tree, respectively. Nodes labeled 1–14 correspond to proportion of genes conflicting over 50%.

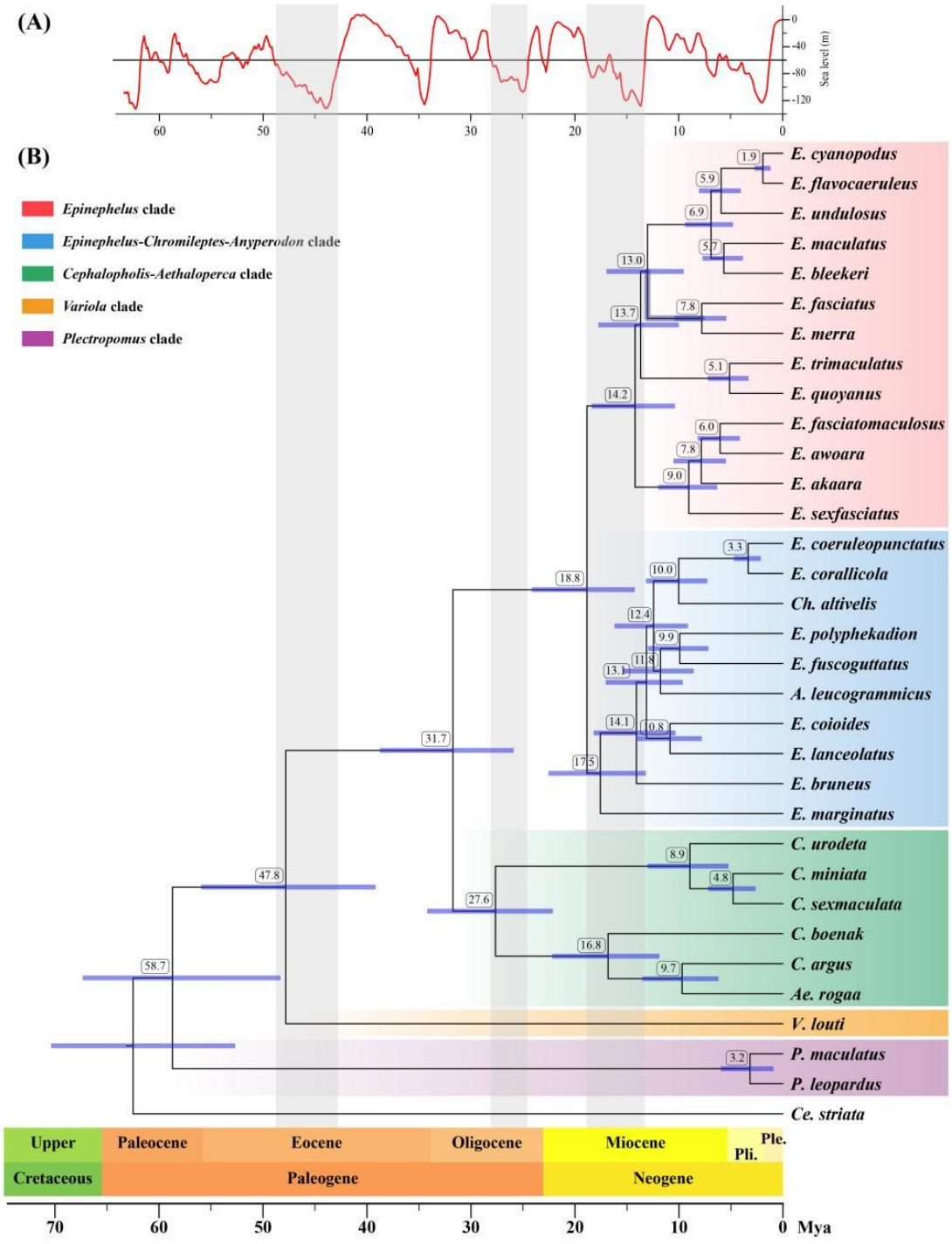


Figure 5 Divergence time estimation for Epinephelidae. (A) Sea level fluctuations across marine oxygen isotope stages (MIS), redrawn from Waelbroeck et al. (2002) and Miller et al. (2005). (B) Species divergence times estimated using MCMCTree. Mean divergence times are shown in the upper left. Bars at nodes indicate 95% posterior probability intervals. Pli, Pliocene; Ple, Pleistocene.

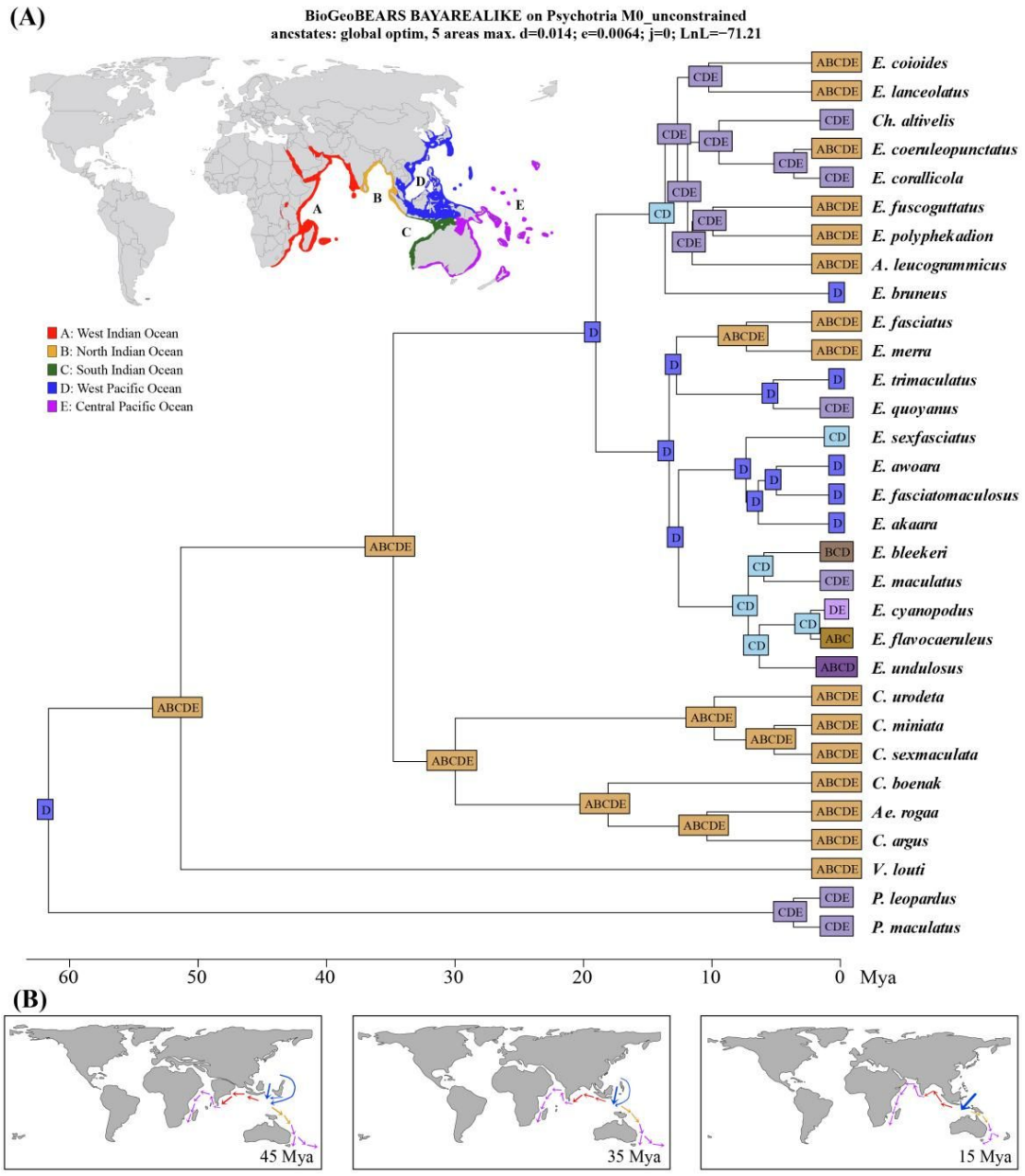


Figure 6 Ancestral area reconstruction of Epinephelidae. (A) Analysis excluding *E. marginatus*, performed using BioGeoBEARS under the BAYAREALIKE model. Grouper distributions are based on Craig et al. (2011), Groupers of the World. (B) A putative origin and migration of groupers currently distributed across the Indian and Pacific Oceans. Ocean currents were redrawn by Gordon et al. (2012), Cetina-Heredia et al. (2014), Bull et al. (2017), and Pang et al. (2021). Blue arrows represent Indonesian Throughflow, red arrows indicate South Equatorial Current, orange arrows illustrate New Guinea Coastal Currents, and purple arrows show potential ocean currents.