Zhang et al. *Zool. Res.* 2024, 45(2): 314−328
https://doi.org/10.24272/j.issn.2095-8137.2023.222

ZR | Zoological Research

Article Open Access

Genetic mechanism of body size variation in groupers: Insights from phylotranscriptomics

Wei-Wei Zhang[1,#], Zhuo-Ying Weng[1,#], Xi Wang[3,#], Yang Yang[1], Duo Li[1], Le Wang[4], Xiao-Chun Liu[1,2], Zi-Ning Meng[1,2,*]

[1] *State Key Laboratory of Biocontrol*, *Institute of Aquatic Economic Animals and Guangdong Province Key Laboratory of Aquatic Economic Animals*, *School of Life Sciences*, *Sun Yat-sen University*, *Guangzhou*, *Guangdong* 510275, *China*

[2] *Southern Laboratory of Ocean Science and Engineering (Zhuhai)*, *Zhuhai*, *Guangdong* 519000, *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*

ABSTRACT

Animal body size variation is of particular interest in evolutionary biology, but the genetic basis remains largely unknown. Previous studies have shown the presence of two parallel evolutionary genetic clusters within the fish genus *Epinephelus* with evident divergence in body size, providing an excellent opportunity to investigate the genetic basis of body size variation in vertebrates. Herein, we performed phylotranscriptomic analysis and reconstructed the phylogeny of 13 epinephelids originating from the South China Sea. Two genetic clades with an estimated divergence time of approximately 15.4 million years ago were correlated with large and small body size, respectively. A total of 180 rapidly evolving genes and two positively selected genes were identified between the two groups. Functional enrichment analyses of these candidate genes revealed distinct enrichment categories between the two groups. These pathways and genes may play important roles in body size variation in groupers through complex regulatory networks. Based on our results, we speculate that the ancestors of the two divergent groups of groupers may have adapted to different environments through habitat selection, leading to genetic variations in metabolic patterns, organ development, and lifespan, resulting in body size divergence between the two locally adapted populations. These findings provide important insights into the genetic mechanisms underlying body size variation in groupers and species differentiation.

Keywords: Phylotranscriptomics; Grouper; Body size; Rapidly evolving genes (REGs); Positively selected genes (PSGs)

INTRODUCTION

Body size is an important trait among species and is associated with various physiological and ecological variables, such as metabolic rate (Martin & Palumbi, 1993), lifespan (Moss et al., 2016), generation time (Martin & Palumbi, 1993), and species distribution (Steele & López-Fernández, 2014). Among the differences observed between different species, body size is one of the most apparent. Numerous biologists have attempted to summarize the general evolutionary patterns of body size evolution at a macroevolutionary scale. For example, Bergmann's rule, which attempts to summarize the response of body size to climate change (Salewski & Watt, 2017; Teplitsky et al., 2008), states that: "Within species and amongst closely related species of homeothermic animals, a larger size is often achieved in colder climates than in warmer ones, which is linked to the temperature budget of these animals." (Salewski & Watt, 2017). Cope's rule of body size evolution asserts that lineage body size tends to increase over evolutionary time (Heim et al., 2015). Additionally, the Island rule states that special animal body size adaptations occur under island ecological constraints (Faurby & Svenning, 2016), while Rensch's rule describes the patterns of sexual dimorphism in body size among animals (Tubaro & Bertelli, 2003).

Fundamentally, animal body size variation is determined by cell number and size (Yang & Xu, 2011), but is also considered a complex quantitative trait controlled by multiple genes or single major effect genes regulated by intricate signaling mechanisms that maintain organ size stability. Multiple signaling pathways have been implicated in the regulation of cell growth and organ size, including the insulin signaling pathway (also known as the growth signaling pathway), target of rapamycin (TOR) signaling pathway, PI3K-

Received: 15 September 2023; Accepted: 05 December 2023; Online: 06 December 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (32273136, 31872572), Agriculture Research System of China (ARS-47), Science and Technology Planning Project of Guangdong Province (2023B1212060023), and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (SML2023SP201)

[#]Authors contributed equally to this work

*Corresponding author, E-mail: mengzn@mail.sysu.edu.cn

Akt pathway, transforming growth factor-β (TGF-β) signaling pathway, and Wnt signaling pathway (Texada et al., 2020). Various genes in the insulin signaling pathway are associated with animal body size (Lai et al., 2021). For instance, single nucleotide polymorphisms (SNPs) in the insulin-like growth factor I (*igf1*) gene are crucial determinants of small body size in domestic dogs (Bannasch et al., 2020), while the *igf2bp1* gene genotype explains variations in body size-related traits in Beijing ducks (Zhou et al., 2018). The TOR signaling pathway, essential for cell growth and metabolism regulation, primarily relies on mTORC1, which is influenced by the PI3K-Akt pathway and tuberous sclerosis complex (TSC), with mutations in *tsc1* or *tsc2* found to significantly increase cell and organ size (Richardson et al., 2004). The TGF-β signaling pathway, with 40 known members, features bone morphogenetic proteins (BMPs) associated with body size in multiple species, such as *bmp8a* in Siamese fighting fish (*Betta splendens*) (Wang et al., 2022b). The Wnt signaling pathway regulates organ size by inhibiting adipogenesis (Longo et al., 2004). Genes such as *lcorl* and *ncapg* in the SRC/STAT3 signaling pathway are closely related to body size in Chinese donkeys (Shen et al., 2021), Qinchuan cattle (Han et al., 2017), horses (Staiger et al., 2016), and domestic pigs (Rubin et al., 2012). The Hippo signaling pathway regulates growth, and genetic mutations in the related kinase cascade can result in tissue overgrowth (Hong et al., 2018). Moreover, mutations in the *mc4r* gene of the leptin-melanocortin pathway can lead to monogenic obesity in various vertebrates from fish to humans (Tao, 2022). In summary, animal body size regulation is a complex process, governed by interactions of multiple body size-related genes and signaling pathways. In the current study, we explored whether these body size-related genetic pathways and their interactions exhibit similar patterns among species.

Fish species account for over 50% of total extant vertebrates. Studies have shown that ecological niche and adaptive evolution are associated with body size variation in fish (Audzijonyte et al., 2020; Borstein et al., 2019; Rincon-Sandoval et al., 2020). Understanding body size variation among such species will help elucidate the relationship between body size-related regulatory pathways and evolution. Previous studies have identified genes that influence growth and development regulation in fish (Triantaphyllopoulos et al., 2020; Vélez et al., 2017), including growth axis-related genes (*gh*, *ghr*, *igf*, *ghrh*, and *ghih*) important for regulating fish metabolism and physiological processes, muscle growth-related genes (*mstn* and *MyoD*), appetite and energy metabolism-related genes (*mc4r*), and body size, muscle growth, and skeletal development-related BMPs. However, reports on the genetic mechanisms underlying body size variation among different fish species through genomics remain scarce.

The rapid development of sequencing technologies in recent years has provided valuable insights into the genetic mechanisms underlying fish body size evolution. Genome sequencing technology enables the reconstruction phylogenetic and evolutionary histories of species by analyzing numerous sites across the whole genome (Yang et al., 2020). Comparative genomics analysis can unveil the genetic factors contributing to significant body size differences among different species. For instance, multiple genes related to body size have been identified in Pacific ocean rockfish (Kolora et al., 2021). Genes under positive selection, such as *epcam* and *bmpr1b*, have been implicated in the morphological differentiation of African cichlid fish (Brawand et al., 2014; Fan et al., 2011). Additionally, *bmp8a* is considered a potential candidate gene in Siamese fighting fish (*B. splendens*) (Wang et al., 2022b). Comparative transcriptomic analysis is an advantageous tool for deciphering species evolution, as changes in coding regions can impact differential gene expression and population differentiation. Notably, changes in genetic regulation contribute to adaptations in species (Gilad et al., 2006; Romero et al., 2012). Based on comparative transcriptomic analysis, differentially expressed genes and positively selected genes (PSGs) have been analyzed in three closely related catfish species, with 20 candidate genes related to body size regulation identified in the Mekong giant catfish (*Bagarius yarrelli*), including *akt3*, *sh2b1*, and *pkm2a*, which are closely associated with growth regulation (Jiang et al., 2019). A study on the potential evolutionary differences in body size in four percid species revealed 78 rapidly evolving genes (REGs) and 41 PSGs in pike perch (*Sander lucioperca*) that may explain its large and slender body size characteristics (Xie et al., 2019). These findings provide strong support for a deeper comprehension of the molecular mechanisms underlying body size evolution in fish species.

Groupers (family Epinephelidae), representing a diverse group of coral reef fish, comprise more than 160 species in 16 genera (Craig et al., 2011; Ma et al., 2016; Qu et al., 2018). The genus *Epinephelus*, characterized by considerable body size variation among species, is an ideal taxonomic group for studying body size divergence and evolution. Notably, the giant grouper (*E. lanceolatus*), the largest species in the genus, can attain a length of 270 cm and body mass of over 400 kg, while the rock grouper (*E. fasciatomaculosus*) has a maximum recorded length of only 30 cm (Craig et al., 2011). Based on historical biogeographic analysis, Ma et al. (2016) reported that large- and small-bodied groupers are predominantly found in Clades E and F, respectively. Molecular phylogenetic studies have also identified two genetic lineages within *Epinephelus* (Ding et al., 2006; Durand et al., 2020; Liang et al., 2020; Zhuang et al., 2013), indicating distinct nodes in the evolutionary history of grouper body size divergence. Evidence has also suggested that the connection between speciation and morphological evolution has left a subtle molecular phylogenomic imprint in extant biota (Hamilton et al., 2022; Rabosky et al., 2013; Ronco et al., 2021). In addition, the effects of ancient geological events and climate change on the diversification of groupers at a broad spatial scale remains underexplored. Ding et al. (2006) noted that the species composition within the two parallel-evolving sister lineages did not correlate with geographic distribution. Ma et al. (2016) proposed that the diversification of groupers is more closely linked to global climate and environmental changes than to direct geological events. However, such studies have primarily been based on only a limited array of nuclear or mitochondrial genes.

Genome sequencing technology has enabled the reconstruction of phylogenetic and evolutionary histories of species using whole-genome data (Berthelot et al., 2014). While several reference genomes have been assembled and used to analyze phylogeny in *Epinephelus*, including the genomes of *E. lanceolatus* (Zhou et al., 2019a), *E. tukula* (Wang et al., 2022a), and *E. fuscoguttatus* (Yang et al., 2022), large-scale orthologous gene detection remains limited due to

challenges in obtaining annotated high-quality genomes. Phylotranscriptomics (transcriptome-based phylogenetic inference) is a cost-effective and powerful tool for identifying orthologous genes and reconstructing phylogenetic relationships among group species (Cheon et al., 2020; Morales-Briones et al., 2021). Given its near-equivalence in comprehensiveness and reliability to phylogenomics (Cheon et al., 2020), phylotranscriptomics has emerged as the preferred method for exploring evolutionary relationships among biological lineages (Lyra et al., 2021; Rancilhac et al., 2021; Washburn et al., 2017). Employing whole-genome data can generate more reliable gene trees to represent evolutionary relationships (Cheon et al., 2020), providing an important basis for exploring the drivers of early differentiation of groupers. Therefore, the establishment of reliable phylogeny is of great significance for elucidating the speciation, body differentiation, and evolutionary history of ancestral lineages in the above two genetic clades.

In this study, we employed phylotranscriptomics to reconstruct the phylogenetic relationships and estimate divergence times of 13 *Epinephelus* species in the South China Sea. We identified REGs and PSGs within the two major evolutionary branches of the genus *Epinephelus* and performed functional annotation analysis to explore the roles of these genes in body size evolution. This study aims to provide insights into the genetic mechanisms of body size evolution in groupers and, more broadly, in fish.

MATERIALS AND METHODS

Ethics statement

All experiments were performed according to 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 the School of Life Sciences, Sun Yat-sen University under approval No. SYSU-IACUS-2022-B0129.

Sample collection

Sampling included 15 epinephelids from the South China Sea, representing 13 *Epinephelus* species (*E. akaara*, *E. awoara*, *E. bleekeri*, *E. coioides*, *E. corallicola*, *E. fasciatomaculosus*, *E. fasciatus*, *E. fuscoguttatus*, *E. lanceolatus*, *E. merra*, *E. moara*, *E. polyphekadion*, and *E. quoyanus*) and two species from closely related genera (*Cephalopholis urodeta* and *Plectropomus maculatus*), which were used as the outgroup (Supplementary Table S1). Habitat and ecological data of groupers (including maximum size (total length), maximum weight (body mass), age, habitat depth range, habitat, and prey) were collected from the IUCN Red List of Threatened Species (https://www.iucnredlist.org), FishBase (https://www.fishbase.se/search.php), and Groupers of the World: A Field and Market Guide (Craig et al., 2011). Violin plots were constructed using GraphPad Prism v.8.0.0. The Mann-Whitney U test was performed using GraphPad Prism v.8.0.0 with a significance level of 0.05. In 2019, specimens were obtained from fish markets and live aquarium trades located in proximity to the South China Sea, specifically near Shenzhen, Huizhou, and Guangzhou in Guangdong Province, China. Brain, tail fin, heart, kidney, liver, muscle, spine, skin, and spleen tissues were collected from three individuals per species, which were preserved in RNAlater and stored at −80°C prior to RNA extraction.

RNA extraction, library construction, and RNA sequencing

Total RNA from each specimen was extracted using TRIzol reagent (Invitrogen, USA) following the manufacturer's protocols. RNA concentration and purity were measured using a NanoDrop 2000 (Thermo Fisher Scientific, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, USA). In total, 1 μg of RNA was generated by pooling RNA from nine tissues of each individual and used for RNA sample preparations. Sequencing libraries were generated using a NEBNext UltraTM RNA Library Prep Kit for Illumina (USA) following the manufacturer's instructions. In brief, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads and fragmented. First-strand cDNA was synthesized using random hexamer primer and M-MuLV reverse transcriptase, followed by second-strand cDNA synthesis and NEBNext adaptor ligation. The library fragments were purified using the AMPure XP system (Beckman Coulter, USA). Polymerase chain reaction (PCR) was performed using Phusion High-Fidelity DNA polymerase, with the PCR products then purified. Finally, paired-end sequencing was performed using the Illumina NovaSeq 6000 platform (Illumina, USA) by BioMarker (China).

Data filtering and *de novo* assembly

Raw data were quickly checked using FastQC v.0.12.1 (Schmieder & Edwards, 2011) and filtered using Trimmomatic v.0.32 (Bolger et al., 2014), with default parameters except for minimum reads length (MINLEN), which was set to 100. After filtering, the clean reads of each specimen were *de novo* assembled with Trinity v.2.8.5 (Grabherr et al., 2011) using default parameters, except for --CPU, which was set to 20, and --min_kmer_cov, which was set to 2. The three assemblies of each species were then merged into one transcriptome, and preliminary redundancy reduction was performed using CD-HIT-EST (Li & Godzik, 2006) with -c set to 0.95 and -n set to 10. For the transcriptome of each species, the longest transcripts were extracted with get_longest_isoform_seq_per_trinity_gene.pl in Trinity v.2.1.1 (Haas et al., 2013). Open reading frames (ORFs), coding sequences (CDS), and peptides were then predicted using TransDecoder v.5.0.0 (https://github.com/TransDecoder/TransDecoder/releases) assisted by BLASTP searches (E-value<10^{-5}) against the Manually Annotated Section of the UniProt Knowledge Base (UniProtKB/Swiss-Prot), NCBI Non-Redundant Protein Sequences (Nr) database, and Protein Family (Pfam) database. To further reduce redundancy, local BLASTX searches (E-value<10^{-5}) were conducted for the predicted CDS against the NCBI RefSeq (Pruitt et al., 2007) of all fish datasets (129 species) using DIAMOND (Buchfink et al., 2015). The best-hit sequences were selected as final unigenes, with functional analysis of the annotated genes then conducted based on the Nr, Swiss-Prot, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genome (KEGG) databases. Final transcriptome completeness was assessed using BUSCO v.5.5.0 (Manni et al., 2021) based on the actinopterygii_odb10 database (highly conserved fish orthologs, https://busco.ezlab.org).

Ortholog identification and sequence alignment

Orthogroups in the 15 epinephelid transcriptomes were identified using OrthoFinder v.2.5.2 (Emms & Kelly, 2015), employing an all-against-all BLAST (self and reciprocal

BLASTs simultaneously) algorithm to perform reciprocal best-hit BLASTs, with normalization for transcript length (Carruthers et al., 2018). Only the one-to-one single-copy orthologous genes shared by all 15 species were selected to perform downstream analysis, which were subsequently aligned by MAFFT v.7.453 (Katoh & Standley, 2013) with the L-INS-I algorithm and trimmed by trimAl v.1.4 (Capella-Gutiérrez et al., 2009) with the --automated1 parameter. Any alignments <100 amino acids were discarded after trimming, and the remaining orthologs were concatenated into a supermatrix. According to the amino acids of orthologs, the corresponding nucleotide CDS were aligned using PAL2NAL v.14.0 (Suyama et al., 2006).

Phylogenetic analysis and divergence time estimation

We reconstructed the phylogeny of the 15 epinephelids from the South China Sea using both concatenation- and coalescent-based methods. For the concatenated supermatrix, maximum-likelihood (ML) and Bayesian inference (BI) trees were inferred. In ML analysis, the best-fit model was determined by ModelFinder (Kalyaanamoorthy et al., 2017), and tree reconstruction was conducted using IQ-TREE v.2.1.2 (Nguyen et al., 2014), with 1 000 replicates of ultrafast bootstrap approximation (UFboot) and SH-like approximate likelihood ratio test (SH-aLRT) for branch support evaluation (Guindon et al., 2010). In BI analysis, the best-fit model was inferred using PartitionFinder v.2.1.1 (Lanfear et al., 2016) based on AICc scores, and tree reconstruction was conducted using MrBayes v.3.2.7 (Ronquist et al., 2012) with Markov Chain Monte Carlo (MCMC) chains (one cold and three heated chains in each run separately) running for 1 000 000 generations and sampled every 1 000 generations. The first 20% of sampled generations were discarded as burn-in, and Tracer v. 1.5 (http://tree.bio.ed.ac.uk/software/tracer/) was used to check whether the runs had reached effective sampling size (ESS)>200 in all model parameters. For coalescent-based phylogeny reconstruction, single trees were generated for each gene using IQ-TREE with the same parameters as above. Subsequently, ASTRAL-III analysis (Zhang et al., 2018) was performed, which reconstructs a species tree that maximizes the number of shared induced quartet trees with the input gene trees, while also accounting for gene tree discordance (Mirarab et al., 2014; Zhang et al., 2018). Trees were rerooted with *P. maculatus* and branches with UFboot<10 were collapsed using the pxrr and pxcol tools in phyx, respectively (Brown et al., 2017). An approximately unbiased (AU) test (Shimodaira, 2002) was performed to check the alternative topologies of the trees based on log-likelihood score comparison with IQ-TREE. Additionally, conflicts and concordance of single gene trees were checked using PhyParts (Smith et al., 2015). The results were then visualized with phypartspiecharts.py script (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts).

Divergence time was estimated using MCMCTREE in PAML v.4.9 (Yang, 2007) based on the approximate likelihood method, with the concatenation-based ML tree specified as the reference tree. As fossil records of Epinephelidae are uncommon, fossil calibration was based on two known fossils of *Epinephelus*, including *E. cassorti* (dated to the Badenian) (Schultz, 2000) calibrated as the lower boundary of the common ancestor of *Epinephelus*, and *E. itajara* (dated to the late Miocene) (Aguilera & de Aguilera, 2004) calibrated as the lower boundary of the common ancestor of *E. lanceolatus* & *E. coioides*. Time was also calibrated according to the recorded divergence time range in TimeTree (http://timetree.org/) of four species pairs: *E. fuscoguttatus* & *E. lanceolatus* (10.49–22.50 Ma), *E. awoara* & *E. merra* (12.45–28.51 Ma), *E. fasciatus* & *E. merra* (7.25–17.95 Ma), and *C. urodeta* & *E. lanceolatus* (>16 Ma). The root was maximally bound at 48 million years ago (Ma) according to the TimeTree estimation. Bayesian analysis was run for 10 000 000 generations and sampled every 100 generations, with the first 20% of sampled generations discarded as burn-in and ESS checked by Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/).

Positive selection analysis

Comparison of nonsynonymous/synonymous substitution ratios (d_N/d_S, ω) in protein-coding genes is a powerful method for detecting adaptive molecular evolution (Yang & Bielawski, 2000). Notably, ω>1 suggests positive selection, while ω<1 or ω=1 indicates purifying and neutral evolution, respectively (Sun et al., 2019). To explore whether genes under different selection pressures have contributed to contrasts in body size between the two diverged *Epinephelus* clades, branch models were used to detect lineage-specific selection pressure by CodeML in PAML, with two three-ratio models: (1) large body-sized clade group, small body-sized clade group, and outgroup were assumed to have independent ω; (2) large body-sized group, small body-sized group, and outgroup were assumed to have independent ω. The above three-ratio models were compared with the one-ratio model that assumed all branches of the phylogenetic tree shared the same ω based on likelihood ratio tests (LRTs), with false discovery rate (FDR) used for multiple testing correction. Genes with an extreme ω value (ω≤0.0001 or ≥999) were filtered, and an adjusted P<0.05 was considered to identify REGs (Yang & Dos Reis, 2011; Yin et al., 2021).

To explore whether genes have experienced positive selection in specific sites along particular lineages, we also used branch-site models to test the branches of interest (foreground branch) separately, including: (1) common ancestor of large body-sized clade group, (2) common ancestor of small body-sized clade group, or (3) each terminal node. We first calculated and compared site-model pairs of M7 & M8 (by LRTs) to search for genes with varied ω among sites. For site-selected genes, we conducted branch-site model analysis for each foreground branch, followed by filtering based on LRTs, FDR correction (adjusted P≤0.05), and Bayes Empirical Bayes (BEB) estimation>95%. The remaining genes with 1<ω<999 were considered to be PSGs (Yang & Dos Reis, 2011; Yin et al., 2021).

Functional enrichment analysis

All obtained candidate REGs and PSGs were annotated using the GO database for functional prediction. GO functional enrichment analysis (P<0.05) was implemented using the clusterProfiler v.4.0 package (Wu et al., 2021). GO and KEGG functional enrichment analyses were performed using the Metascape database (https://metascape.org/gp/index.html#/main/step1), with zebrafish (*Danio rerio*) as a reference species, to explore the functional relevance of these genes (Zhou et al., 2019b). Min overlap≥3 and P≤0.01 were considered statistically significant. To further explore the potential functional changes caused by mutations in positive selection sites, the functional domains of PSGs were predicted using the PROSITE database (https://prosite.expasy.org). Pathway enrichment results were generated with the

Omicshare online website (http://www.omicshare.com/tools/Home/Index).

RESULTS

Transcriptome *de novo* assembly and annotation

After sequencing and initial filtering, a total of 112 million paired-end clean reads and 139.00 Gb of clean bases were generated from the 15 epinephelids. The *de novo* transcriptome generated a total of 6 133 286 (110 316 to 210 233) transcripts with average unigenes, contig length, and N50 size per specimen of 114,967 (97 228 to 175 395), 760 bp (650 bp to 849 bp), and 1 383 bp (1 066 bp to 1 578 bp), respectively (Supplementary Table S2). The *de novo* assemblies were then filtered for redundancy reduction, resulting in an average of 24 987 unigenes per species (23 753 to 27 217) with a mean length of 1 279 bp (Supplementary Table S3). The filtered transcriptomes were then annotated against currently known protein databases, with the most matches in Nr (99.70% on average), followed by Swiss-Prot (84.12% on average), GO (82.34% on average), and KEGG (56.23% on average) (Supplementary Table S3). The percentage of complete and single-copy BUSCOs identified in the 15 transcriptomes ranged from 64.3% to 72.0%, and the percentage of missing BUSCOs was lower than 30% in all species (Supplementary Figure S1), indicating that the final transcriptomes after redundancy reduction were highly consistent and relatively complete. After filtering, 2 042 one-to-one single-copy orthologous genes were obtained for downstream phylogenetic and positive selection analyses.

Phylogenomic analyses and divergence time estimation

The reconstructed phylogeny of 13 epinephelids was well-resolved and highly supported. The concatenation-based BI and ML trees showed topological consensus, with robust posterior probability (PP>0.95) support for all nodes and SH-aLRT (≥80)/UFboot (≥95) support for most nodes, respectively (Supplementary Figures S2–S3). This phylogeny demonstrated that the 13 *Epinephelus* species formed two major clades correlated with body size. With the reconstructed phylogeny of the 13 *Epinephelus* species, two major clades were identified, and divergence time was estimated to be approximately 15.4 Ma (8.4–23.3 Ma; Figure 1A). The first clade was comprised of the large groupers (large-bodied clade, LBC), with an average maximum body length of 123.7 cm, including *E. polyphekadion*, *E. fuscoguttatus*, *E. corallicola*, *E. lanceolatus*, *E. coioides*, and *E. moara*. The second clade mainly included small groupers (small-bodied clade, SBC), with an average maximum body length of 41.5 cm, including *E. merra*, *E. fasciatus*, *E. bleekeri*, *E. quoyanus*, *E. fasciatomaculosus*, *E. awoara*, and *E. akaara* (Figure 1A, C). The Mann-Whitney U test indicated significant differences in maximum recorded total length (P=0.0076<0.05), maximum recorded body mass (P=0.0076<0.05), and maximum recorded age (P=0.0012<0.05) between the LBC and SBC groups (Figure 2A–C). Although no significant difference in habitat depth was found between the LBC and SBC groups (P=0.3800>0.05), average habitat depth of LBC was deeper than that of SBC (Figure 2D).

Positive selection analysis of body size variation in groupers

Branch model: We performed LRTs using a three-ratio branch model to explore genes under different selection pressure during the diversification of grouper species with different body sizes. The null hypothesis of the LRT one-ratio model showed that the average ω ratio of 1 832 one-to-one orthologous genes was 0.274, with most (98.96%) being less

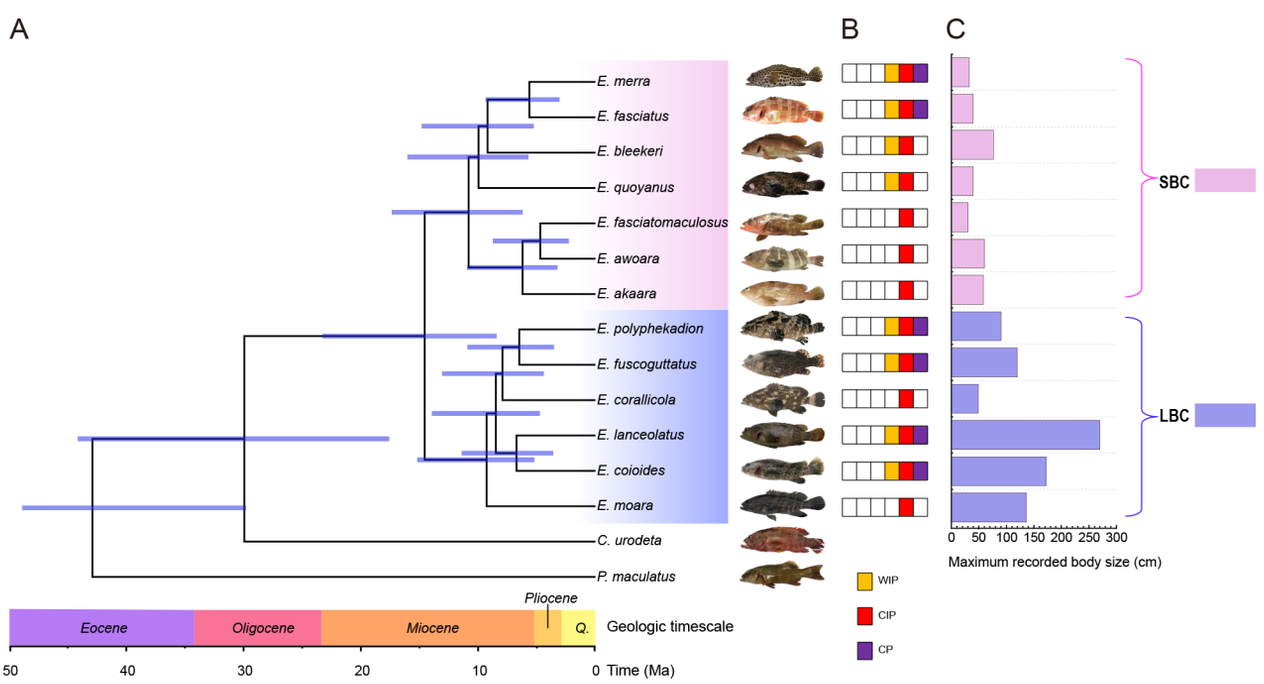

Figure 1 Phylogenetic relationships and body size variations in 13 epinephelids

A: Phylogenetic relationships reconstructed based on concatenation of 13 epinephelids with divergence time estimation. The 95% confidence intervals (CI) of estimated divergence times are shown in blue bars at each node. Species could be differentiated into a small body-sized clade (SBC) (pink) and large body-sized clade (LBC) (blue). Ma: Million years ago; Q: Quaternary. B: Extant ranges are indicated for 13 epinephelids, modified after Ma et al. (2016). See figure key for interpretation of symbols and color fills. WIP: West Indo-Pacific, CIP: Central Indo-Pacific, CP: Central Pacific. C: Maximum recorded body size of 13 epinephelids.

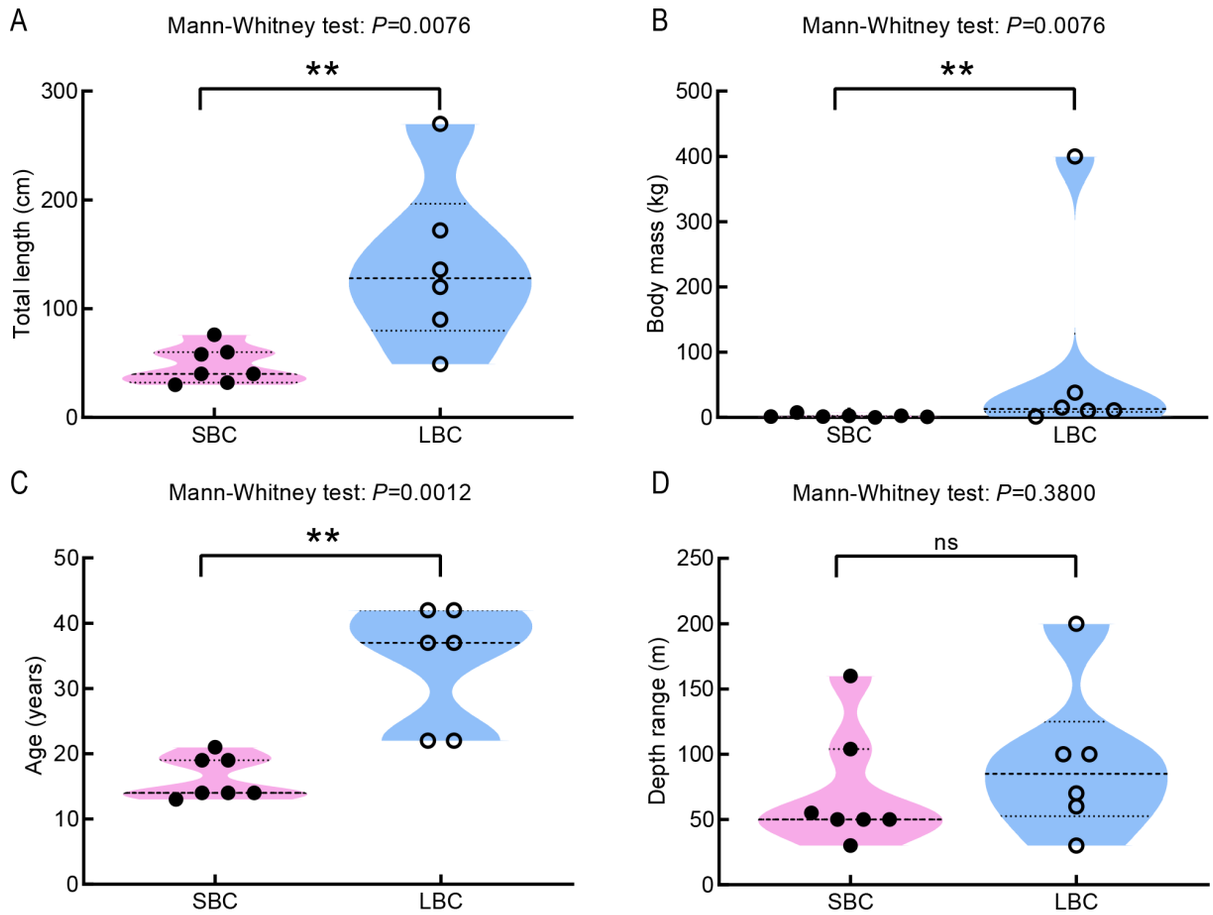

Figure 2 Violin plot comparing body size variation, age, and depth range in SBC and LBC groups

A–D: Violin plot showing maximum recorded total length (A), maximum recorded body mass (B), maximum recorded age (C), and maximum recorded depth range (D) of SBC and LBC. Datapoints represent each individual core and boxplot to visualize body size variation in groupers. Dotted lines are 75th and 25th percentiles, and dashed line are values within the median. ns: Not significant (P>0.05); **: P<0.01.

than 1.0 (Figure 3A, B), indicating widespread purifying selection on these orthologous genes. A total of 189 REGs were identified through branch-model testing. Regarding LBC and SBC divergence, 180 REGs were detected after stringent filtering (Supplementary Table S4), with 72 exhibiting significantly higher evolutionary rates in LBC and 108 REGs exhibiting significantly higher evolutionary rates in SBC.

Branch-site model: Following initial screening using the site models M7 and M8, a total of 312 genes containing positively selected sites were retained for branch-site model testing. One PSG, *ndrg1a*, was detected in the foreground branch with the common ancestor of the LBC, while another PSG, *jup*, was detected in the foreground branch with the common ancestor of the SBC (Table 1; Supplementary Figure S4–S5). Within these two PSGs, a total of seven divergent sites were identified, with two significant divergent sites found in *ndrg1a* and one significant divergent site found in *jup* (Table 1; Supplementary Figures S4, S5). According to PROSITE database predictions, three of the significantly mutated sites were not located within predicted functional domains. However, the Asp215Thr and Ser344Ala sites in *ndrg1a* underwent mutations from aspartic acid to threonine and from serine to alanine, respectively, which may impact protein phosphorylation (Figure 3C; Table 1). Regarding the remaining four sites, functional domain predictions revealed that the Thr136Ile divergent site in *ndrg1a* was situated in the phosphopantetheine attachment site (Figure 3C; Table 1), the Val160Ala divergent site in *jup* was located in the armadillo/plakoglobin (ARM) repeat region (Figure 3D; Table 1), and the Ser2Gln site in *jup* underwent an amino acid change from serine to glutamine (Figure 3D; Table 1), resulting in alterations in polarity or acidity/basicity, which may also affect protein function.

Functional enrichment analysis of REGs and PSGs

The top 50 enriched GO terms for LBC and SBC, listed in Supplementary Table S5, included genes involved in biological processes such as metabolic process (GO:0008152), development process (GO:0032502), response to stimulus (GO:0050896), and growth (GO:0040007) (Figure 4; Supplementary Table S5). Subsequently, Metascape was used to identify major enriched pathways and related genes. Results showed that the REGs of LBC were enriched in 11 functional annotation clusters, while the REGs of SBC were enriched in 16 functional annotation clusters (Table 2).

GO enrichment analysis of REGs in LBC and SBC exhibited differences in genes and pathways. The insulin/IGF and mTORC1 signaling pathways are well-known pathways involved in the regulation of cell growth and body size. Based on literature and database searches to assess gene functions, *igfbp1a*, *kl*, *insra*, and *itgav*, associated with the insulin/IGF signaling pathway, underwent rapid evolution in SBC, while *tbc1d7*, *cul4b*, *npc1*, and *parp1*, associated with the mTORC1 signaling pathway, experienced rapid evolution in LBC.

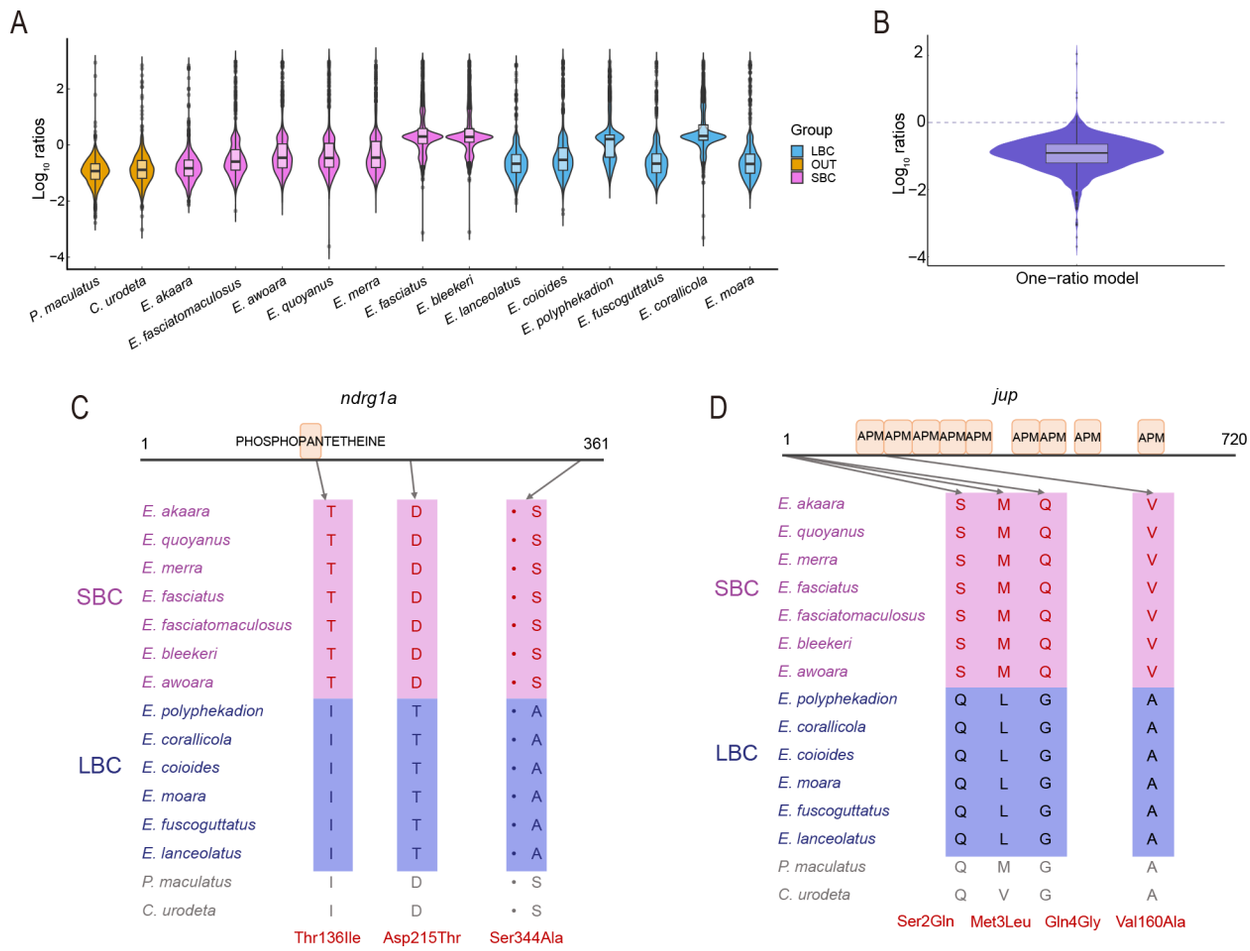

Figure 3 Positive selection analysis of body size variation in groupers

A: Violin plot of dN/dS ratios (ω) of 15 epinephelids for single-copy orthologous genes based on one-ratio models. B: Violin plot of mean ω for all single-copy orthologous genes based on one-ratio models. Ratios were log-transformed. C, D: Amino acid positions (numbers) and changes at each position in *ndrg1a* (C) and *jup* (D) of PSGs between LBC and SBC groups.

Table 1 Positive selected sites under branch-site for foreground lineages Prob (ω>1)

Gene	Background: with outgroup				Background: without outgroup			
	Position	Variance	ω	BEB	Position	Variance	ω	BEB
ndrg1a	215	T	409.5	0.997**	136	I	143.8	0.863
	344	A		0.997**	215	T		0.992**
					344	A		0.992**
jup	2	Q	333.5	0.931	2	Q	195.5	0.827
	4	G		0.979*	3	L		0.534
	160	A		0.672	4	G		0.913
					160	A		0.510

A, Q, G, I, and L represent amino acids. A: Alanine; Q: Glutamine; G: Glycine; I: Isoleucine; L: Leucine. ω: Comparison of nonsynonymous/synonymous substitution ratios (d_N/d_S). BEB: Bayes Empirical Bayes. *: Posterior probability >95%; **: Posterior probability >99%.

Additionally, several genes (*tbc1d7*, *arhgef15b*, *arhgap29a*, *bcl2l1*, and *pcsk1nl*) related to GTPase activity activation and mTORC1 function were also enriched in LBC (Supplementary Table S5). In terms of catabolic and metabolic processes, LBC was primarily enriched in protein metabolic processes (GO:0043161 proteasome-mediated ubiquitin-dependent protein catabolic process and GO:0010498 proteasomal protein catabolic process), while SBC was associated with carbohydrate metabolism (GO:0005975) and lipid metabolism (GO:0016042) (Supplementary Table S5). In organ system development, LBC was enriched in skeletal system development and tissue morphogenesis involving multiple organs, including kidney development, renal system development, and urogenital system development (Table 2; Supplementary Table S5). Gene intersections were also present among these processes. SBC exhibited enrichment in genes related to cranial skeletal system development and blood vessel morphogenesis, which share genes related to vascular and heart development. Notably, two important clotting factors, *f7i* and *pros1*, involved in regulation of stress response, were identified in SBC (Table 2; Supplementary Table S5). Conversely, LBC was enriched in pathways and genes associated with aging and lifespan, including ribosomal precursor (*pwp2h*, *nol6*, and *riok3*), DNA repair (*topbp1*, *cul4b*,

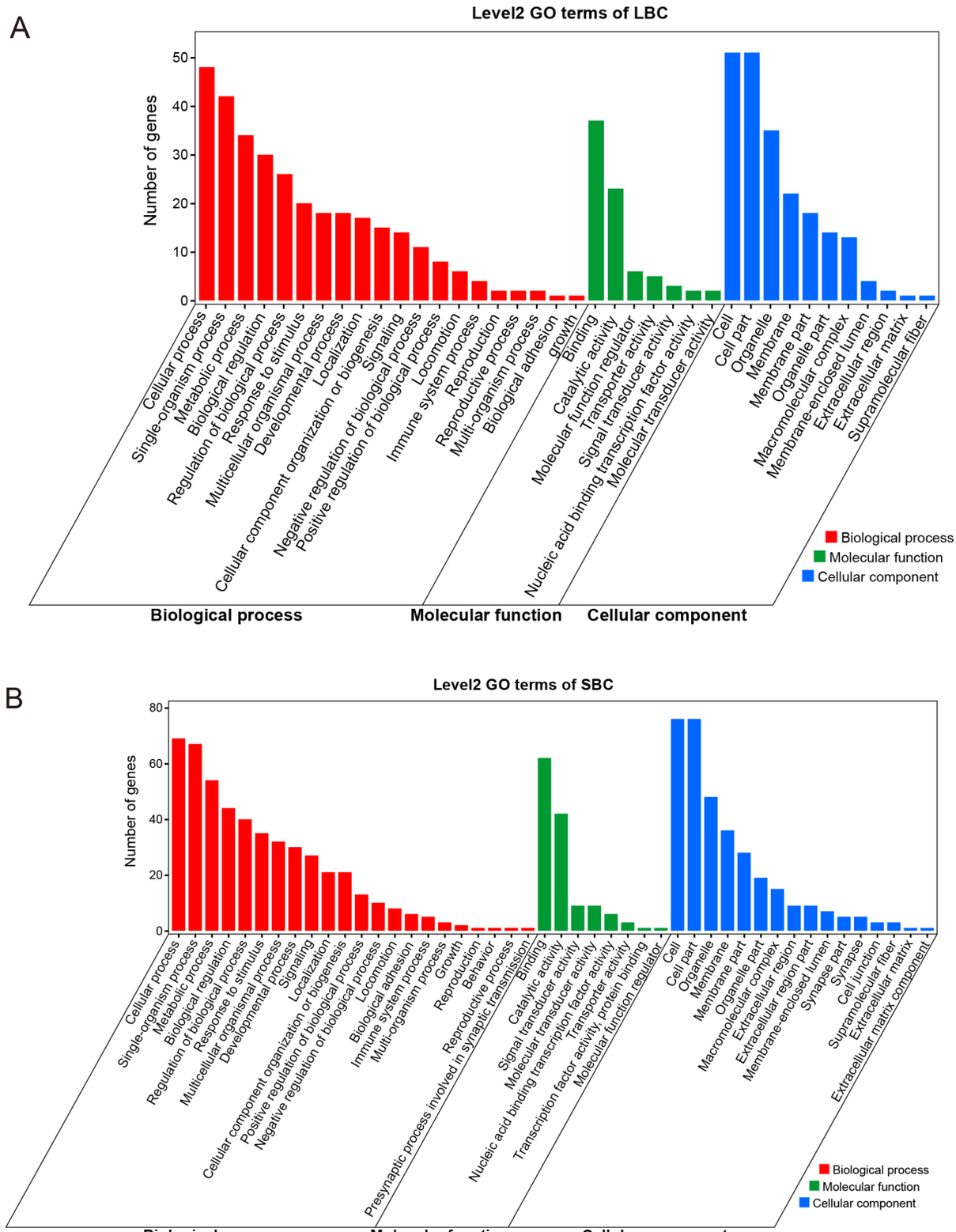

Figure 4 Functional enrichment analysis of REGs and PSGs

A, B: Histograms of GO term annotations for REGs of LBC (A) and SBC (B). Abscissa is secondary GO terms, and ordinate is number of differentially expressed genes enriched in that term.

and *parp1*), cellular response to DNA damage stimulus (*bcl2l1*, t*opbp1*, *cul4b*, *parp1*, and *bag6l*), and negative regulation of proteolysis (*bcl2l1*, *pcsk1nl*, *bag6l*, *topbp1*, and *cul4b*), which were not enriched in SBC (Table 2).

Regarding PSG functions, our literature review indicated that *ndrg1a* predominantly functions as a cytoplasmic protein involved in various biological processes, encompassing cellular stress responses, cell differentiation, proliferation, growth arrest, tumor formation, hypoxia response, and DNA damage response, while *jup* functions in heart development and negative regulation of the Wnt signaling pathway (Table 2).

Table 2 Functions of REGs with the most significant correlations (*P*<0.01)

Model	Group	Term	Description	Gene symbol
Branch models	LBC	GO:0090630	Activation of GTPase activity	*tbc1d7*, *arhgef15b*, *arhgap29a*, *bcl2l1*, *pcsk1nl*
		GO:0030684	Pre-ribosome	*pwp2h*, *nol6*, *riok3*
		GO:0045861	Negative regulation of proteolysis	*bcl2l1*, *pcsk1nl*, *bag6l*, *topbp1*, *cul4b*
		GO:0048592	Tissue morphogenesis	*tbx2b*, *aldh1a2*, *top2b*, *npc1*, *ezrb*, *pfn1*, *bag6l*
		R-DRE-73894	DNA Repair	*topbp1*, *cul4b*, *parp1*
		GO:0005319	Lipid transporter activity	*osbpl2b*, *npc1*, *slco2b1*
		R-DRE-2990846	SUMOylation	*uba2*, *parp1*, *top2b*
		GO:0080135	Regulation of cellular response to stress	*topbp1*, *otub1b*, *il6st*
		GO:0008514	Organic anion transmembrane transporter activity	*slc35b4*, *slc38a2*, *slco2b1*
		GO:0043632	Modification-dependent macromolecule catabolic process	*fbxl5*, *cul4b*, *bag6l*, *sgpl1*, *ufc1*, *cand2*, *uba2*, *otub1b*, *usp3*, *smg1*
		GO:0001501	Skeletal system development	*aldh1a2*, *b4galt7*, *npc1*, *crispld2*
	SBC	GO:0005975	Carbohydrate metabolic process	*insra*, *amdhd2*, *hexa*, *neu1*, *kl*, *galk2*, *pargl*, *dpydb*, *lipeb*, *pafah1b2*
		dre04512	ECM-receptor interaction	*thbs1b*, *hspg2*, *col6a2*, *itgav*, *dync1h1*, *tfr1b*, *stx12l*, *rock2a*, *kdrl*
		GO:1901565	Organonitrogen compound catabolic process	*rock2a*, *lnpep*, *amdhd2*, *dpydb*, *uba1*, *hexa*, *ube3b*, *neu1*, *uba7*, *usp18*
		GO:1901699	Cellular response to nitrogen compound	*nr4a1*, *sesn1*, *chrne*, *kl*, *esr2b*, *cnn2*
		GO:0106310	Protein serine kinase activity	*rock2a*, *ripk2*, *sik1*, *dapk1*, *insra*, *nrbp1*, *pkn2a*, *stk17al*, *kdrl*
		GO:0019838	Growth factor binding	*igfbp1a*, *kl*, *kdrl*
		GO:0008238	Exopeptidase activity	*lnpep*, *scpep1*, *prss16*, *f7i*, *cfi*, *pgpep1*, *usp18*
		GO:0080134	Regulation of response to stress	*f7i*, *uba1*, *sesn1*, *pros1*, *pargl*, *gpr183a*
		GO:0098802	Plasma membrane signaling receptor complex	*insra*, *chrne*, *itgav*, *kdrl*
		WP467	mRNA processing	*hnrnpaba*, *xrn2*, *snrpa1*
		GO:0000904	Cell morphogenesis involved in differentiation	*unk*, *uba1*, *pls3*, *gpr183a*, *als2a*, *col6a2*
		GO:0005769	Early endosome	*tfa*, *tmem184a*, *als2a*
		GO:1904888	Cranial skeletal system development	*pls3*, *hspg2*, *pkn2a*, *usp18*
		GO:0048514	Blood vessel morphogenesis	*cfi*, *hspg2*, *kdrl*, *itgav*, *tgfbr3*
		GO:0005654	Nucleoplasm	*hnrnpaba*, *chtopa*, *rcor3*, *pnisr*, *ibtk*, *krt8*
Branch-site models	LBC	GO:0045576	Mast cell activation	*ndrg1a*
	SBC	GO:0003308	Negative regulation of Wnt signaling pathway involved in heart development	*jup*

DISCUSSION

Water habitat facilitated ancestral body size differentiation in groupers

Species diversity is shaped by a variety of abiotic and biotic factors, including geological and climatic events, which markedly influence biological diversity (Cowman, 2014; Kamikuri & Moore, 2017). Determining the effects of ancient geological events and climate change on species diversification, such as that of groupers on a large spatial scale, remains a complex task. Establishing phylogenetic relationships among species is fundamental for understanding body size evolution. In this study, 2 042 orthologous genes obtained from transcriptomic data were used to reconstruct the phylogenetic tree of 13 epinephelids distributed in the South China Sea, confirming the division of these species into two major topological clades distinguished by body size. These topological branches diverged approximately 15.4 Ma (95% CI: 8.4–23.3 Ma) during the Miocene epoch, a period lasting 17.7 million years (23.03–5.33 Ma) and marked by a series of geological events that impacted biogeographic and climatic changes, as documented in the comprehensive review “The Miocene: The Future of the Past” (Steinthorsdottir et al., 2021). Notably, several significant geological events occurred in the ancient Indo-West Pacific region during the Miocene period, including the fusion of the Indo-Australian Archipelago (IAA) (23–5 Ma) (Cowman & Bellwood, 2011; Lohman et al., 2011), closure of the Tethys Seaway (Terminal Tethyan Event, TTE) (12.8 Ma) (Harzhauser et al., 2007; Sun et al., 2021), and proliferation of coral reefs (Montaggioni & Braithwaite, 2009), all of which influenced the distribution and evolution of coral reef fish (Cowman, 2014; Cowman & Bellwood, 2011). The subsequent Mid-Miocene Climate Transition (MMCT, 14.2–13.8 Ma) brought global cooling, East Antarctic Ice Cap expansion, and reduced global sea levels, resulting in the isolation of coral reef habitats into shallow waters (Shevenell et al., 2004). The origin of the 13 studied epinephelids was traced to the Proto-Indo-West Pacific Region (Figure 1B). Analyzing the differences in maximum depth range of existing grouper species, our results showed that most species in the small-bodied branch primarily inhabit shallow waters, while the large-bodied groupers mainly occupy deeper habitats (Figure 2C). Integrated whole-genome systematics and modern taxonomy suggested that grouper differentiation may not be directly linked to events such as bony coral reef expansion, IIA uplift, or TTE, but the shallow water environments resulting from global climate cooling significantly impacted grouper distribution and evolution, aligning with the findings of Ma et al. (2016). We speculate that the common ancestor of the 13 epinephelids adapted to

habitats with different water depths, leading to the selection of different body sizes in response to different water layer environments, which, in turn, promoted grouper radiation and speciation.

Population isolation can also contribute to increased genomic variation in coral reef fish, allowing them to adapt and diverge along ecological axes (Bromham, 2011; Drury et al., 2016; Schluter & Conte, 2009). The variation in body size of grouper ancestors may support the hypothesis that ecological opportunities promote phenotypic variation under abiotic factors, such as geological activities, climate events, and resource diversity. However, further exploration is needed to investigate the role of biological factors (genetic differences) in this variation.

Genetic mechanism of body size variation in common ancestor of groupers

(1) Metabolism

Body size is associated with energy requirements and food abundance (White et al., 2007). Our study revealed distinct gene profiles in catabolic and metabolic processes between the LBC and SBC groups. Specifically, genes related to protein metabolism were predominantly associated with LBC, while genes related to glucose and lipid metabolism were predominantly associated with SBC. Furthermore, the well-recognized insulin/mTOR pathway, crucial for regulating cell growth and body size and for maintaining normal energy balance, body weight, and nutrition (Saxton & Sabatini, 2017), was observed to be under selection within these clades, suggesting a potential role of genes in the mTOR pathway in driving body size diversification within this group.

The insulin signaling pathway plays key roles in systemic growth. In this study, five genes (*tbc1d7*, *igfbp1a*, *kl*, *insra*, and *itgav*) involved in this signaling pathway were identified. *Igfbp1* serves as a crucial inhibitor of IGF activity, while *igfbp1a* mediates hypoxia-induced embryonic growth and developmental retardation in zebrafish (Kajimura et al., 2005). *kl* is an aging-related gene involved in the complex regulation of the GH-IGF1 axis (Kenyon, 2011), and is a PSG related to body size in ruminants (Chen et al., 2019). The insulin receptor *insra* mediates the metabolic effects of insulin and plays essential roles in vertebrate and growth (Das et al., 2016; Toyoshima et al., 2008). In addition, the *itgav* gene directly interacts with *igf1* and regulates the *igf1* signaling pathway, thus influencing cell fate (Khurana et al., 2016). Within the SBC group, the identification of PSGs associated with the insulin signaling pathway (*igfbp1a*, *kl*, *insra*, and *itgav*) suggests that insulin regulation via the IGF axis and subsequent constraints on energy metabolism may play a major role in SBC.

Moreover, *mTORC1* plays a crucial regulatory role in balancing catabolic and anabolic processes, and its activation is strictly controlled by numerous upstream signaling pathways. Here, four genes (*tbc1d7*, *cul4b*, *npc1l*, and *parp1*) were found to be associated with the mTOR signaling pathway. Mutations in *tbc1d7*, the third subunit of the tuberous sclerosis complex (TSC), lead to mutant flies exhibiting a larger body size than wild-type *Drosophila* counterparts (Dibble et al., 2012; Ren et al., 2018). The *cul4b* gene is thought to influence mTOR activity through interactions of the TSC1-TSC2 complex with inhibitory signals of mTOR (Hu et al., 2008). Cholesterol binds to and activates mTORC1 to promote cell growth (Davis et al., 2021), whereas *npc1* prevents sustained activation of mTORC1 signaling via cholesterol-dependent physical interactions (Castellano et al., 2017). *Parp1* activation and PAR synthesis affect cellular energy status, inhibit the mTORC1 signaling pathway, and influence cell fate (Éthier et al., 2012). Thus, these genes are potential key candidate genes contributing to body size variation in groupers. Additionally, the TSC complex, through its GTPase-activating protein (GAP) activity towards Rheb, inhibits mTORC1, a crucial promoter of cell growth (Dibble et al., 2012). In the LBC group, several genes related to GTPase activity activation were enriched. Thus, genes related to the mTOR signaling pathway (*tbc1d7*, *cul4b*, *npc1l*, and *parp1*) and GTPase activity activation (*tbc1d7*, *arhgef15b*, *arhgap29a*, *bcl2l1*, and *pcsk1nl*) were positively selected in LBC, suggesting they may be targets of natural selection and strongly influence size-related energy metabolism processes, potentially explaining the evolution of body size in LBC.

The differences in the primary metabolic genes between the small and large groupers may partly explain their adaptation to different food resources and habitats. Existing large and small body-sized groupers exhibit distinct ranges and foraging strategies: small groupers are active predators that inhabit shallow areas, such as coral and rocky reefs, and sometimes estuaries, while large groupers are ambush predators that inhabit deeper areas of coral reefs ((Linde et al., 2004; Ma et al., 2016), Supplementary Table S1). Transparency decreases with increasing ocean depth, and maximum transparency can reach 60 m (He et al., 2017), similar to the maximum depth of most small groupers. Previous studies have shown that the diet of groupers shifts with size. Notably, smaller individuals (total length<35 cm) tend to feed on crustaceans (such as crabs), with an increased preference for cephalopods as they grow, while larger individuals (total length>65 cm) tend to consume other fish and mollusks (Condini et al., 2015; Reñones et al., 2002). We speculate that the difference in water depth (transparency) plays a crucial role, with higher transparency in shallow areas facilitating active detection and selection of prey (crustaceans) by small groupers, but darker and deeper environments favoring ambush predation of cephalopods and fish by large groupers. Overall, ancestral groupers likely occupied different ecological niches and underwent population segregation, adaptation, and evolution based on different habitats and energy requirements, a pattern supported by evidence from other coral reef fish studies (Goatley & Bellwood, 2016; Mihalitsis et al., 2022; Munday & Jones, 1998).

(2) Organ development

Organ development involves interactions among genes that regulate cell growth, proliferation, differentiation, epithelial-to-mesenchymal transition, and apoptosis (Şahin Uysal et al., 2023). For example, the skeletal system provides a structural framework for muscle attachment and is crucial for supporting movement, protecting organs, and maintaining vascular homeostasis. In LBC, we identified several REGs related to skeletal differentiation, growth, and development (*aldh1a2*, *b4galt7*, *npc1*, *crispld2*), which may play critical roles in maintaining body size in groupers. Our results also identified several pathways related to organ morphogenesis and development, such as tissue morphogenesis, kidney development, renal system development, urogenital system development, morphogenesis of an epithelial sheet, gastrulation, morphogenesis of an epithelium, camera-type eye morphogenesis, embryonic morphogenesis, eye morphogenesis, epithelial cell differentiation, and gland development, which were enriched in genes overlapping with

the skeletal system development (Table 2; Supplementary Table S5). These genes, including *tbx2b*, *aldh1a2*, *npc1*, *ezrb*, *pfn1*, *crispld2*, and *prkd2*, may play important roles in coordinating tissue and organ development in large groupers to better adapt to deeper water environments. In summary, large groupers may have achieved adaptation to deeper water habitats by regulating skeletal system development and various organ morphogenesis and development processes.

In the present study, only one PSG (*jup*) was identified in the branch-site model for SBC. The *jup* gene is a negative regulator of the Wnt signaling pathway and is implicated in embryonic heart development in zebrafish (Martin et al., 2009), suggesting a relationship between heart development and evolution of small groupers. In SBC, several important genes related to blood vessel morphogenesis were identified, including *hspg2*, *kdrl*, *tgfbr3*, *cfi*, and *itgav*. Among them, *hspg2* plays a significant role as a cell adhesion substrate in zebrafish blood vessel formation and heart development (Zoeller et al., 2009), *kdrl* is a vascular endothelial growth factor receptor involved in regulating vascular development and embryonic heart development (Lee et al., 2006), and *tgfbr3* plays an inhibitory role in the TGF-β signaling pathway, which is important in heart development (Compton et al., 2007; Tazat et al., 2015). The SBC group was also enriched in genes related to the regulation of stress response, including *f7i*, *sesn1*, *pros1*, and *pargl*. Both *f7i* and *pros1* are important clotting factors (Wang et al., 2023; Ye et al., 2023). Furthermore, genes involved in cranial skeletal system development, such as *hspg2*, *pkn2a*, *pls3*, and *usp18*, were also identified. Among them, *hspg2* and *pkn2a* are associated with heart development. Fish living in shallow water habitats face higher survival pressures, including competition, predation, and pathogens (Smallhorn-West et al., 2017). Small groupers may limit the development of organs, such as the cranial skeletal system, by regulating blood vessel formation and heart development, thereby maintaining a small body size and adapting to the inherent survival pressures of shallow water habitats.

(3) Lifespan

Animal lifespan varies greatly within and among species, although larger animals are known to have longer lifespans (Speakman, 2005). Analysis of the 13 species of groupers (Supplementary Table S1) indicated that the lifespans of larger groupers are significantly longer than those of smaller groupers (Figure 2). Analyzing the molecular mechanisms underlying size variation in groupers is crucial for understanding the causal relationship between longevity and body size, and it is essential for deciphering the driving factors of lifespan regulation. Inflammation, ribosomes, DNA damage, and protein degradation are associated with lifespan (Sun et al., 2022; Tyshkovskiy et al., 2023). Inflammation is considered a crucial hallmark of aging, and evolutionary strategies that minimize inflammation are beneficial for longevity (Tyshkovskiy et al., 2023). Species with longer lifespans tend to contain a greater number of immune-regulatory genes that down-regulate inflammation, leading to increased longevity (Tyshkovskiy et al., 2023). In our study, *ndrg1a* was the only gene under positive selection in the branch-site model. This gene functions as a negative regulator of interferon (IFN), playing a crucial role in the induction of IFN and the antiviral innate immune response in zebrafish (Lu et al., 2019). This result suggests that the anti-inflammatory capacity of larger body-sized groupers may be associated with extended lifespan. Up-regulation of ribosomal protein genes is an important functional feature in long-lived animals (Jiang et al., 2019). In the LBC group, results showed a significant enrichment in ribosomal precursor-related genes (*pwp2h*, *nol6*, and *riok3*), which may drive the ability of larger groupers to resist aging and maintain higher levels of cell growth and proliferation. Inflammation, a significant result of DNA damage, can be mitigated by DNA repair mechanisms that facilitate the replacement of cells harboring extensive DNA damage through clonal growth, effectively preventing the accumulation of random DNA damage (Sun et al., 2019; Tyshkovskiy et al., 2023; Zhao et al., 2023). Our results showed marked enrichment of DNA damage repair factors in the LBC group, but no enrichment in such factors in smaller groupers. This finding is similar to that reported in Pacific Ocean rockfish showing enrichment in DNA damage repair factors in larges individuals (Kolora et al., 2021). Recent research has also indicated that protein degradation plays an important role in long-lived species (Tyshkovskiy et al., 2023). In this study, protein degradation-related pathways were enriched in the LBC group. Thus, we speculate that these immune factors, DNA repair mechanisms, and protein degradation-related genes were under selection, possibly conferring larger groupers with an enhanced ability to reduce inflammation, counteract the inflammatory response associated with longevity and larger body size, and subsequently promote development into a larger size.

Hypothesis on body size diversification of common ancestor of groupers

Molecular-level adaptations can be achieved through adaptive mutations in key genes over long evolutionary time scales. To investigate the mechanisms underlying such adaptations, this study employed phylotranscriptomic analysis and phylogenetic reconstruction to explore the correlation between body size and genotype in 13 epinephelids from the South China Sea. Based on the geological events and climate changes associated with the habitat depth of existing grouper species and their divergence times, we hypothesize that the increase in shallow-water environments during the middle Miocene period was the primary driving force behind the evolution of two size-related branches of groupers from a common ancestor, which adapted to different ecological niches. In the current study, we identified genetic differences between larger and smaller groupers in terms of metabolism, organ development, and lifespan (Figure 5), leading to divergence of body size between the two locally adapted populations. Notably, metabolic differences were observed between the large-size and small-size groupers, including differences in protein, lipid, and carbohydrate metabolism processes. Regarding differences in organ development, LBC exhibited enrichment in the skeletal development system, including genes related to multiple organ morphologies and development, while SBC showed enrichment in blood vessel formation and heart development, including several genes shared with cranial skeletal system development. Furthermore, the LBC groupers exhibited significantly longer lifespans compared to the SBC groupers. Pathways associated with aging, including negative regulation of immune response, ribosomal precursor, DNA repair, and protein hydrolysis, were enriched in the LBC group but not in the SBC group. Thus, through the combined effects of long-term generational selection, environmental change, and phenotypic selection, genetic differentiation in body size

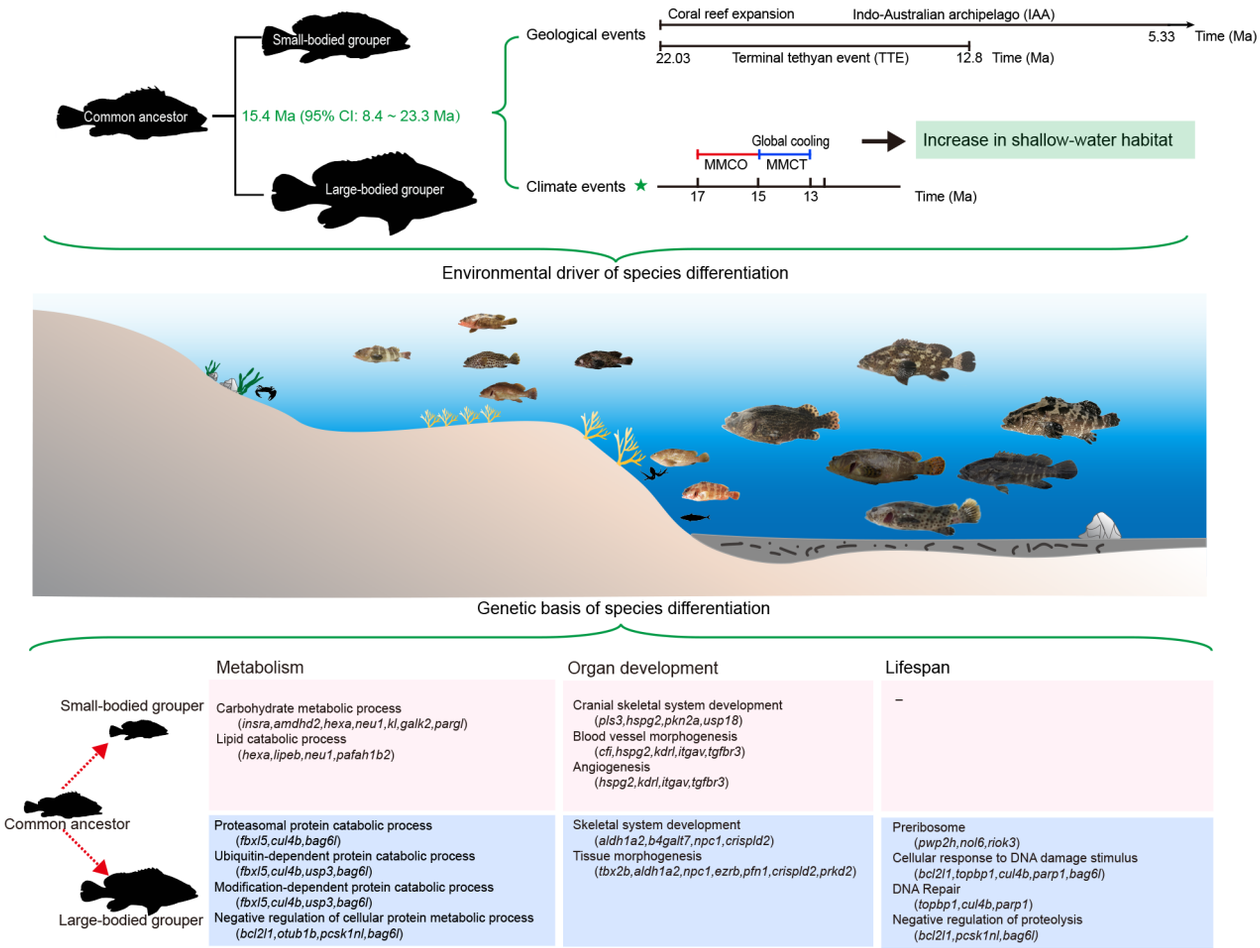

Figure 5 Hypothesis on body size diversification of common ancestor of groupers

Schematic of the main genetic clues revealed in this study, showing that the increase in shallow-water environments due to climate change may explain the differentiation of the common ancestor of groupers, resulting in genetic differences between large and small bodies, such as metabolism, organ development, and lifespan, thus forming the genetic basis for species differentiation of groupers. –: Not available.

between larger and smaller groupers led to the formation of species with diverse body size.

DATA AVAILABILITY

All raw data were deposited in the NCBI Sequence Read Archive (SRA) with accession number SRP483481, Genome Sequence Archive (GSA) in the National Genomics Data Center (NGDC) under accession code CRA011857, and Science Data Bank (https://www.scidb.cn/) under DOI: https://doi.org/10.57760/sciencedb.j00139.00099.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Z.N.M. initiated, designed, and supervised the study. Z.Y.W., X.W., and Y.Y. prepared the samples and sequenced the transcriptome. W.W.Z., Z.Y.W., and X.W. analyzed the data, prepared the figures, and drafted the manuscript. Z.N.M., X.C.L., L.W., Y.Y., and D.L. revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the reviewers for their important suggestions to improve the paper. We thank Li-Na Wu, Tong Wang, and Wen-Hua Huang for kind help in collecting samples.

REFERENCES

Aguilera O, de Aguilera DR. 2004. Goliath grouper (Pisces: Serranidae) from the Upper Miocene Urumaco Formation, Venezuela. *Journal of Paleontology*, **78**(6): 1202−1206.

Audzijonyte A, Richards SA, Stuart-Smith RD, et al. 2020. Fish body sizes change with temperature but not all species shrink with warming. *Nature Ecology & Evolution*, **4**(6): 809−814.

Bannasch DL, Baes CF, Leeb T. 2020. Genetic variants affecting skeletal morphology in domestic dogs. *Trends in Genetics*, **36**(8): 598−609.

Berthelot C, Brunet F, Chalopin D, et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, **5**: 3657.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15): 2114−2120.

Borstein SR, Fordyce JA, O'Meara BC, et al. 2019. Reef fish functional traits evolve fastest at trophic extremes. *Nature Ecology & Evolution*, **3**(2): 191−199.

Brawand D, Wagner CE, Li YI, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**(7518): 375−381.

Bromham L. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **366**(1577): 2503−2513.

Brown JW, Walker JF, Smith SA. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics*, **33**(12): 1886−1888.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment

using DIAMOND. *Nature Methods*, **12**(1): 59−60.
Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15): 1972−1973.
Carruthers M, Yurchenko AA, Augley JJ, et al. 2018. *De novo* transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics*, **19**(1): 32.
Castellano BM, Thelen AM, Moldavski O, et al. 2017. Lysosomal cholesterol activates mTORC1 via an SLC38A9–Niemann-Pick C1 signaling complex. *Science*, **355**(6331): 1306−1311.
Chen L, Qiu Q, Jiang Y, et al. 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, **364**(6446): eaav6202.
Cheon S, Zhang JZ, Park C. 2020. Is phylotranscriptomics as reliable as phylogenomics?. *Molecular Biology and Evolution*, **37**(12): 3672−3683.
Compton LA, Potash DA, Brown CB, et al. 2007. Coronary vessel development is dependent on the type III transforming growth factor β receptor. *Circulation Research*, **101**(8): 784−791.
Condini MV, Hoeinghaus DJ, Garcia AM. 2015. Trophic ecology of dusky grouper *Epinephelus marginatus* (Actinopterygii, Epinephelidae) in littoral and neritic habitats of southern Brazil as elucidated by stomach contents and stable isotope analyses. *Hydrobiologia*, **743**(1): 109−125.
Cowman PF. 2014. Historical factors that have shaped the evolution of tropical reef fishes: a review of phylogenies, biogeography, and remaining questions. *Frontiers in Genetics*, **5**: 394.
Cowman PF, Bellwood DR. 2011. Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *Journal of Evolutionary Biology*, **24**(12): 2543−2562.
Craig MT, Sadovy De Mitcheson Y, Heemstra PC. 2011. Groupers of the World: A Field and Market Guide. Grahamstown: NISC (Pty) Ltd.
Das D, Nath P, Pal S, et al. 2016. Expression of two insulin receptor subtypes, *insra* and *insrb*, in zebrafish (*Danio rerio*) ovary and involvement of insulin action in ovarian function. *General and Comparative Endocrinology*, **239**: 21−31.
Davis OB, Shin HR, Lim CY, et al. 2021. NPC1-mTORC1 signaling couples cholesterol sensing to organelle homeostasis and is a targetable pathway in Niemann-Pick type C. *Developmental Cell*, **56**(3): 260−276.e7.
Dibble C, Elis W, Menon S, et al. 2012. TBC1D7 is a third subunit of the TSC1-TSC2 complex upstream of mTORC1. *Molecular Cell*, **47**(4): 535−546.
Ding SX, Zhuang X, Guo F, et al. 2006. Molecular phylogenetic relationships of China Seas groupers based on cytochrome *b* gene fragment sequences. *Science in China Series C*, **49**(3): 235−242. (in Chinese)
Drury C, Dale KE, Panlilio JM, et al. 2016. Genomic variation among populations of threatened coral: acropora cervicornis. *BMC Genomics*, **17**(1): 286.
Durand JD, Pham MH, Tran TTV, et al. 2020. Sorting the wheat from the chaff: a review of BINs associated with groupers of Vietnam and the implications for species identification from DNA barcoding. *Marine Biodiversity*, **50**(3): 38.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, **16**(1): 157.
Éthier C, Tardif M, Arul L, et al. 2012. PARP-1 modulation of mTOR signaling in response to a DNA alkylating agent. *PLoS One*, **7**(10): e47978.
Fan SH, Elmer KR, Meyer A. 2011. Positive Darwinian selection drives the evolution of the morphology-related gene, EPCAM, in particularly species-rich lineages of African cichlid fishes. *Journal of Molecular Evolution*, **73**(1-2): 1−9.
Faurby S, Svenning JC. 2016. Resurrection of the Island Rule: human-driven extinctions have obscured a basic evolutionary pattern. *The American Naturalist*, **187**(6): 812−820.
Gilad Y, Oshlack A, Rifkin SA. 2006. Natural selection on gene expression. *Trends in Genetics*, **22**(8): 456−461.
Goatley CHR, Bellwood DR. 2016. Body size and mortality rates in coral reef fishes: a three-phase relationship. *Proceedings of the Royal Society B:Biological Sciences*, **283**(1841): 20161858.
Grabherr MG, Haas BJ, Yassour M, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, **29**(7): 644−652.
Guindon S, Dufayard JF, Lefort V, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**(3): 307−321.
Haas BJ, Papanicolaou A, Yassour M, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols*, **8**(8): 1494−1512.
Hamilton CA, Winiger N, Rubin JJ, et al. 2022. Hidden phylogenomic signal helps elucidate arsenurine silkmoth phylogeny and the evolution of body size and wing shape trade-offs. *Systematic Biology*, **71**(4): 859−874.
Han YJ, Chen Y, Liu Y, et al. 2017. Sequence variants of the *LCORL* gene and its association with growth and carcass traits in Qinchuan cattle in China. *Journal of Genetics*, **96**(1): 9−17.
Harzhauser M, Kroh A, Mandic O, et al. 2007. Biogeographic responses to geodynamics: a key study all around the Oligo–Miocene Tethyan Seaway. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, **246**(4): 241–256.
He XQ, Pan DL, Bai Y, et al. 2017. Recent changes of global ocean transparency observed by SeaWiFS. *Continental Shelf Research*, **143**: 159−166.
Heim NA, Knope ML, Schaal EK, et al. 2015. Cope's rule in the evolution of marine animals. *Science*, **347**(6224): 867−870.
Hong LX, Li X, Zhou DW, et al. 2018. Role of hippo signaling in regulating immunity. *Cellular & Molecular Immunology*, **15**(12): 1003−1009.
Hu J, Zacharek SM, He YJ, et al. 2008. WD40 protein FBW5 promotes ubiquitination of tumor suppressor TSC2 by DDB1-CUL4-ROC1 ligase. *Genes & Development*, **22**(7): 866−871.
Jiang WS, Guo YC, Yang KF, et al. 2019. Insights into body size evolution: a comparative transcriptome study on three species of Asian sisoridae catfish. *International Journal of Molecular Sciences*, **20**(4): 944.
Kajimura S, Aida K, Duan CM. 2005. Insulin-like growth factor-binding protein-1 (IGFBP-1) mediates hypoxia-induced embryonic growth and developmental retardation. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(4): 1240−1245.
Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**(6): 587−589.
Kamikuri SI, 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.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**(4): 772−780.
Kenyon C. 2011. The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **366**(1561): 9−16.
Khurana S, Schouteden S, Manesia JK, et al. 2016. Outside-in integrin signalling regulates haematopoietic stem cell function via Periostin-Itgav axis. *Nature Communications*, **7**: 13500.
Kolora SRR, Owens GL, Vazquez JM, et al. 2021. Origins and evolution of extreme life span in Pacific Ocean rockfishes. *Science*, **374**(6569): 842−847.
Lai ZY, Wu F, Li M, et al. 2021. Tissue expression profile, polymorphism of

IGF1 gene and its effect on body size traits of Dezhou donkey. *Gene*, **766**: 145118.

Lanfear R, Frandsen PB, Wright AM, et al. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**(3): 772−773.

Lee YM, Cope JJ, Ackermann GE, et al. 2006. Vascular endothelial growth factor receptor signaling is required for cardiac valve formation in zebrafish. *Developmental Dynamics*, **235**(1): 29−37.

Li WZ, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**(13): 1658−1659.

Liang RS, Chen M, Liao GW, 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**(6): 9−19. (in Chinese)

Linde M, Grau A, Riera F, et al. 2004. Analysis of trophic ontogeny in *Epinephelus marginatus* (Serranidae). *Cybium:International Journal of Ichthyology*, **28**(1): 10086.

Lohman DJ, de Bruyn M, Page T, et al. 2011. Biogeography of the Indo-Australian archipelago. *Annual Review of Ecology, Evolution, and Systematics*, **42**: 205−226.

Longo KA, Wright WS, Kang S, et al. 2004. Wnt10b inhibits development of white and brown adipose tissues. *Journal of Biological Chemistry*, **279**(34): 35503−35509.

Lu LF, Li S, Wang ZX, et al. 2019. Zebrafish NDRG1a negatively regulates IFN induction by promoting the degradation of IRF7. *The Journal of Immunology*, **202**(1): 119−130.

Lyra ML, Monteiro JPC, Rancilhac L, et al. 2021. Initial phylotranscriptomic confirmation of homoplastic evolution of the conspicuous coloration and bufoniform morphology of pumpkin-toadlets in the genus *Brachycephalus*. *Toxins*, **13**(11): 816.

Ma KY, Craig MT, Choat JH, et al. 2016. The historical biogeography of groupers: clade diversification patterns and processes. *Molecular Phylogenetics and Evolution*, **100**: 21−30.

Manni M, Berkeley MR, Seppey M, et al. 2021. BUSCO: assessing genomic data quality and beyond. *Current Protocols*, **1**(12): e323.

Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **90**(9): 4087−4091.

Martin ED, Moriarty MA, Byrnes L, et al. 2009. Plakoglobin has both structural and signalling roles in zebrafish development. *Developmental Biology*, **327**(1): 83−96.

Mihalitsis M, Morais RA, Bellwood DR. 2022. Small predators dominate fish predation in coral reef communities. *PLoS Biology*, **20**(11): e3001898.

Mirarab S, Reaz R, Bayzid MS, et al. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**(17): i541−i548.

Montaggioni LF, Braithwaite CJR. 2009. Quaternary coral reef systems: history, development processes and controlling factors. *Developments in Marine Geology*, **5**: iii.

Morales-Briones DF, Kadereit G, Tefarikis DT, et al. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in amaranthaceae s. l. *Systematic Biology*, **70**(2): 219−235.

Moss DK, Ivany LC, Judd EJ, et al. 2016. Lifespan, growth rate, and body size across latitude in marine Bivalvia, with implications for Phanerozoic evolution. *Proceedings of the Royal Society B:Biological Sciences*, **283**(1836): 20161364.

Munday PL, Jones GP. 1998. The ecological implications of small body size among coral-reef fishes. *In*: Ansell A, Barnes M, Gibson RN, et al. Oceanography and Marine Biology. London: CRC Press, 373–411.

Nguyen LT, Schmidt HA, von Haeseler A, et al. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**(1): 268−274.

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**(S1): D61−D65.

Qu M, Tang W, Liu QH, et al. 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.

Rabosky DL, Santini F, Eastman J, et al. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, **4**: 1958.

Rancilhac L, Irisarri I, Angelini C, et al. 2021. Phylotranscriptomic evidence for pervasive ancient hybridization among old world salamanders. *Molecular Phylogenetics and Evolution*, **155**: 106967.

Ren SX, Huang ZY, Jiang YQ, et al. 2018. dTBC1D7 regulates systemic growth independently of TSC through insulin signaling. *Journal of Cell Biology*, **217**(2): 517−526.

Reñones O, Polunin NVC, Goni R. 2002. Size related dietary shifts of *Epinephelus marginatus* in a western Mediterranean littoral ecosystem: an isotope and stomach content analysis. *Journal of Fish Biology*, **61**(1): 122−137.

Richardson CJ, Schalm SS, Blenis J. 2004. PI3-kinase and TOR: PIKTORing cell growth. *Seminars in Cell & Developmental Biology*, **15**(2): 147−159.

Rincon-Sandoval M, Duarte-Ribeiro E, Davis AM, et al. 2020. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(52): 33396−33403.

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, **13**(7): 505−516.

Ronco F, Matschiner M, Böhne A, et al. 2021. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*, **589**(7840): 76−81.

Ronquist F, Teslenko M, van der Mark P, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**(3): 539−542.

Rubin CJ, Megens HJ, Barrio AM, et al. 2012. Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(48): 19529−19536.

Şahin Uysal Nİ, Şahin F, Terzi YK. 2023. The impact of developmental genes in non-syndromic cleft lip and/or palate. *Journal of the Turkish-German Gynecological Association*, **24**(1): 57−64.

Salewski V, Watt C. 2017. Bergmann's rule: a biophysiological rule examined in birds. *Oikos*, **126**(2): 161−172.

Saxton RA, Sabatini DM. 2017. mTOR signaling in growth, metabolism, and disease. *Cell*, **168**(6): 960−976.

Schluter D, Conte GL. 2009. Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(S1): 9955−9962.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**(6): 863−864.

Schultz VO. 2000. Ein zackenbarsch (Epinephelus, Serranidae, Pisces) aus dem Mittel-Miozän von Retznei, Steiermark. *Joannea Geologie und Palä ontologie*, **2**: 5−56.

Shen JF, Yu J, Dai XL, et al. 2021. Genomic analyses reveal distinct genetic architectures and selective pressures in Chinese donkeys. *Journal of Genetics and Genomics*, **48**(8): 737−745.

Shevenell AE, Kennett JP, Lea DW. 2004. Middle Miocene southern ocean cooling and Antarctic cryosphere expansion. *Science*, **305**(5691): 1766−1770.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree

selection. *Systematic Biology*, **51**(3): 492−508.
Smallhorn-West PF, Bridge TCL, Munday PL, et al. 2017. Depth distribution and abundance of a coral-associated reef fish: roles of recruitment and post-recruitment processes. *Coral Reefs*, **36**(1): 157−166.
Smith SA, Moore MJ, Brown JW, et al. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, **15**: 150.
Speakman JR. 2005. Body size, energy metabolism and lifespan. *Journal of Experimental Biology*, **208**(9): 1717–1730.
Staiger EA, Al Abri MA, Pflug KM, et al. 2016. Skeletal variation in Tennessee walking horses maps to the *LCORL/NCAPG* gene region. *Physiological Genomics*, **48**(5): 325−335.
Steele SE, López-Fernández H. 2014. Body size diversity and frequency distributions of Neotropical cichlid fishes (Cichliformes: Cichlidae: Cichlinae). *PLoS One*, **9**(9): e106336.
Steinthorsdottir M, Coxall HK, De Boer AM, et al. 2021. The Miocene: the future of the past. *Paleoceanography and Paleoclimatology*, **36**(4): e2020PA004037.
Sun D, Chai SM, Huang X, et al. 2022. Novel genomic insights into body size evolution in cetaceans and a resolution of Peto's paradox. *The American Naturalist*, **199**(2): E28−E42.
Sun JM, Sheykh M, Ahmadi N, et al. 2021. Permanent closure of the Tethyan Seaway in the northwestern Iranian Plateau driven by cyclic sea-level fluctuations in the late Middle Miocene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **564**: 110172.
Sun YY, Liu YZ, Sun XH, et al. 2019. Insights into body size variation in cetaceans from the evolution of body-size-related genes. *BMC Evolutionary Biology*, **19**(1): 157.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**(S2): W609−W612.
Tao YX. 2022. Chapter Seven - Mutations in melanocortin-4 receptor: from fish to men. *Progress in Molecular Biology and Translational Science*, **189**(1): 215−257.
Tazat K, Hector-Greene M, Blobe GC, et al. 2015. TβRIII independently binds type I and type II TGF-β receptors to inhibit TGF-β signaling. *Molecular Biology of the Cell*, **26**(19): 3535−3545.
Teplitsky C, Mills JA, Alho JS, et al. 2008. Bergmann's rule and climate change revisited: disentangling environmental and genetic responses in a wild bird population. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(36): 13492−13496.
Texada MJ, Koyama T, Rewitz K. 2020. Regulation of body size and growth control. *Genetics*, **216**(2): 269−313.
Toyoshima Y, Monson C, Duan CM, et al. 2008. The role of insulin receptor signaling in zebrafish embryogenesis. *Endocrinology*, **149**(12): 5996−6005.
Triantaphyllopoulos KA, Cartas D, Miliou H. 2020. Factors influencing *GH* and *IGF-I* gene expression on growth in teleost fish: how can aquaculture industry benefit?. *Reviews in Aquaculture*, **12**(3): 1637−1662.
Tubaro PL, Bertelli S. 2003. Female-biased sexual size dimorphism in tinamous: a comparative test fails to support Rensch's rule. *Biological Journal of the Linnean Society*, **80**(3): 519−527.
Tyshkovskiy A, Ma SM, Shindyapina AV, et al. 2023. Distinct longevity mechanisms across and within species and their association with aging. *Cell*, **186**(13): 2929−2949.e20.
Vélez EJ, Lutfi E, Azizi S, et al. 2017. Understanding fish muscle growth regulation to optimize aquaculture production. *Aquaculture*, **467**: 28−40.
Wang JX, Wu NS, Feng XW, et al. 2023. PROS1 shapes the immune-suppressive tumor microenvironment and predicts poor prognosis in glioma. *Frontiers in Immunology*, **13**: 1052692.
Wang L, Sun F, Lee M, et al. 2022b. Whole-genome resequencing infers genomic basis of giant phenotype in Siamese fighting fish (*Betta splendens*). *Zoological Research*, **43**(1): 78−80.
Wang LN, Li ZT, Liu Y, et al. 2022a. A chromosome-level genome assembly of the potato grouper (*Epinephelus tukula*). *Genomics*, **114**(5): 110473.
Washburn JD, Schnable JC, Conant GC, et al. 2017. Genome-guided phylo-transcriptomic methods and the nuclear phylogenetic tree of the paniceae grasses. *Scientific Reports*, **7**(1): 13528.
White EP, Ernest SKM, Kerkhoff AJ, et al. 2007. Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, **22**(6): 323−330.
Wu TZ, Hu EQ, Xu SB, et al. 2021. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, **2**(3): 100141.
Xie P, Yi SK, Yao H, et al. 2019. Comparative transcriptome analysis reveals potential evolutionary differences in adaptation of temperature and body shape among four Percidae species. *PLoS One*, **14**(5): e0215933.
Yang XL, Xu T. 2011. Molecular mechanism of size control in development and human diseases. *Cell Research*, **21**(5): 715−729.
Yang Y, Wang T, Chen JF, et al. 2022. Whole-genome sequencing of brown-marbled grouper (*Epinephelus fuscoguttatus*) provides insights into adaptive evolution and growth differences. *Molecular Ecology Resources*, **22**(2): 711−723.
Yang Y, Wu LN, Chen JF, 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**(3): 328−340.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586−1591.
Yang Z, Dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution*, **28**(3): 1217−1228.
Yang ZH, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, **15**(12): 496−503.
Ye XY, Mi XX, Sun JW, et al. 2023. *PROS1* variant c. 1574C>T p. Ala525Val causes portal vein thrombosis with protein S deficiency. *Clinics and Research in Hepatology and Gastroenterology*, **47**(6): 102141.
Yin Y, Fan H, Zhou B, et al. 2021. Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer. *Nature Communications*, **12**(1): 6858.
Zhang C, Rabiee M, Sayyari E, et al. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, **19**(S6): 153.
Zhao Y, Simon M, Seluanov A, et al. 2023. DNA damage and repair in age-related inflammation. *Nature Reviews Immunology*, **23**(2): 75−89.
Zhou Q, Gao HY, Zhang Y, et al. 2019a. A chromosome-level genome assembly of the giant grouper (*Epinephelus lanceolatus*) provides insights into its innate immunity and rapid growth. *Molecular Ecology Resources*, **19**(5): 1322−1332.
Zhou YY, Zhou B, Pache L, et al. 2019b. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, **10**(1): 1523.
Zhou ZK, Li M, Cheng H, et al. 2018. An intercross population study reveals genes associated with body size and plumage color in ducks. *Nature Communications*, **9**(1): 2648.
Zhuang X, Qu M, Zhang X, et al. 2013. A comprehensive description and evolutionary analysis of 22 grouper (Perciformes, Epinephelidae) mitochondrial genomes with emphasis on two novel genome organizations. *PLoS One*, **8**(8): e73561.
Zoeller JJ, Whitelock JM, Iozzo RV. 2009. Perlecan regulates developmental angiogenesis by modulating the VEGF-VEGFR2 axis. *Matrix Biology*, **28**(5): 284−291.