

Article

Whole-Genome Sequencing Analyses Reveal the Whip-like Tail Formation, Innate Immune Evolution, and DNA Repair Mechanisms of *Eupleurogrammus muticus*

Fang-Yuan Han ¹, Ren-Xie Wu ^{1,*}, Ben-Ben Miao ², Su-Fang Niu ¹, Qing-Hua Wang ³ and Zhen-Bang Liang ¹

¹ College of Fisheries, Guangdong Ocean University, Zhanjiang 524088, China; hanfangyuan1999@163.com (F.-Y.H.); wolf0487@126.com (S.-F.N.); liangzhenbang0403@163.com (Z.-B.L.)

² State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China; miaobenben@stu.xmu.edu.cn

³ State Key Laboratory of Biocontrol, Institute of Aquatic Economic Animals and Guangdong Provincial Key Laboratory for Aquatic Economic Animals, Life Sciences School, Sun Yat-sen University, Guangzhou 510275, China; wqh31016877430@163.com

* Correspondence: wurenxie@163.com

Simple Summary: We constructed a high-quality genome assembly of *Eupleurogrammus muticus* at the chromosomal level using PacBio SMRT, Illumina Nova-Seq, and Hi-C technologies. By combining genomic annotation, comparative genomic analyses, and species attribute information, we identified many candidate genes related to the whip-like tail, innate immunity, and DNA repair of *E. muticus*, and determined the evolutionary relationship and divergence time between *E. muticus* and related species. These findings provide important genomic resources for exploring the genetic mechanisms underlying the unique characteristics of *E. muticus* and fishery resource conservation.

Abstract: Smallhead hairtail (*Eupleurogrammus muticus*) is an important marine economic fish distributed along the northern Indian Ocean and the northwest Pacific coast; however, little is known about the mechanism of its genetic evolution. This study generated the first genome assembly of *E. muticus* at the chromosomal level using a combination of PacBio SMRT, Illumina Nova-Seq, and Hi-C technologies. The final assembled genome size was 709.27 Mb, with a contig N50 of 25.07 Mb, GC content of 40.81%, heterozygosity rate of 1.18%, and repetitive sequence rate of 35.43%. *E. muticus* genome contained 21,949 protein-coding genes (97.92% of the genes were functionally annotated) and 24 chromosomes. There were 143 expansion gene families, 708 contraction gene families, and 4888 positively selected genes in the genome. Based on the comparative genomic analyses, we screened several candidate genes and pathways related to whip-like tail formation, innate immunity, and DNA repair in *E. muticus*. These findings preliminarily reveal some molecular evolutionary mechanisms of *E. muticus* at the genomic level and provide important reference genomic data for the genetic studies of other trichiurids.

Keywords: *Eupleurogrammus muticus*; genome sequencing; comparative genomics; positive selection



Citation: Han, F.-Y.; Wu, R.-X.; Miao, B.-B.; Niu, S.-F.; Wang, Q.-H.; Liang, Z.-B. Whole-Genome Sequencing Analyses Reveal the Whip-like Tail Formation, Innate Immune Evolution, and DNA Repair Mechanisms of *Eupleurogrammus muticus*. *Animals* **2024**, *14*, 434. <https://doi.org/10.3390/ani14030434>

Academic Editor: Fuhua Li

Received: 15 November 2023

Revised: 19 January 2024

Accepted: 25 January 2024

Published: 29 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fishes of the family Trichiuridae (Teleostei, Perciformes) are important marine fishery resources widely distributed in the tropical, subtropical, and temperate waters of the Pacific, Indian, and Atlantic Oceans [1]. There are 10 genera and approximately 45 species of trichiurids recorded worldwide [2], among which the economically capturable groups belong to the genera *Trichiurus*, *Lepturacanthus*, *Eupleurogrammus*, *Lepidopus*, and *Aphanopus* [3]. *Trichiurus*, *Lepturacanthus*, and *Eupleurogrammus* are harvested mainly from the Indo-West Pacific Oceans, while *Lepidopus* and *Aphanopus* are mainly produced from the local areas of the Atlantic Ocean [3]. There have been many reports on the fishery, resource,

biology, ecology, and genetics of the genera *Trichiurus* and *Lepturacanthus* [4–11], which are commercially important fishes in the Indo-West Pacific Oceans. However, there are fewer studies on *Eupleurogrammus*, limiting the conservation and management of its biodiversity and fishery resources.

Eupleurogrammus consists of the smallhead hairtail (*Eupleurogrammus muticus*) and longtooth hairtail (*E. glossodon*), and the former has a wider distribution range and higher fishing yield. *E. muticus* inhabits the benthopelagic zone of the continental shelf in the northern Indian Ocean and northwest Pacific coast (depths of 30–100 m), feeding on small crustaceans, mollusks, and other fishes [3,12]. It is distributed along the coast of China, with its largest yield occurring in the northern South China Sea [13]. However, with the continuous deterioration of the global marine environment, the fragmentation of biological habitats and the long-term overfishing by humans [14,15], the yield of *E. muticus* on the Chinese coast has been exhibiting a significant downward trend [16]. Furthermore, the total niche width of *E. muticus* is smaller than that of genera *Trichiurus* and *Lepturacanthus*, and its ability to utilize resources and adapt to the environment is also weaker than that of these two genera of fish [4]. However, *E. muticus* population resources in the southern Yellow Sea and the northern South China Sea are still assessable and contribute to the fishing yield of trichiurids [5,17]. According to the routine fishery resource survey in the past decade, the annual catch of *E. muticus* accounts for about 5–10% (i.e., 15,000–30,000 t) of the total annual catch (approximately 300,000 t, 2010–2022 years) of trichiurids in the northern South China Sea.

Similar to other caudal-finless trichiurids, the body of *E. muticus* is elongated with no scales, laterally compressed, and ribbon-like-shaped, with well-developed sharp teeth and a tapering whip-like tail [18]. Notably, *E. muticus* has three unique features: (1) the adult *E. muticus* is relatively small, with a body length measuring 20–50 cm; (2) the body color of *E. muticus* is more silvery-white, with white dorsal and pectoral fins; and (3) the lateral line is not curved above the pectoral fin, showing a relatively straight line backward along the ventral margin to the caudal end [3]. Thus, the widespread distribution, resource stability, and biological traits of *E. muticus* suggest that it may have a unique genomic characterization. However, only a few studies have investigated the biology of *E. muticus* in the northern Indian Ocean [12,19–21] and its genetic variation in the Yellow Sea of China [16,22]. There is also a lack of research to reveal the genetic and evolutionary mechanisms of the species at the genomic level. Although the genome of *Lepturacanthus savala*, belonging to the family Trichiuridae, and several genes and pathways related to its specific morphological and behavioral characteristics have been established [23], the exact molecular mechanisms associated with these features are not yet clear. Therefore, there is an urgent need to further explore the genomic information of other trichiurids with unique species attributes, such as *E. muticus*, to gain a comprehensive and in-depth understanding of the molecular mechanisms underlying the formation of this particular taxon of trichiurids.

In this study, we combined the PacBio SMRT, Illumina Nova-Seq, and Hi-C technologies to obtain high-quality chromosome-level genomic data for *E. muticus*. The molecular mechanisms associated with the whip-like tail, innate immunity, and DNA repair of *E. muticus* were further explored based on genome annotation, gene family contraction and expansion analyses, positive selection analysis, and species attributes. The phylogenetic relationship and divergence time between *E. muticus* and other fishes were also evaluated using the constructed genome phylogenetic tree. The results of this study provide a solid foundation for understanding the genetic composition, evolutionary history, and ecological adaptations of *E. muticus* and accurate reference genomic data for the genetic resources of other trichiurids.

2. Materials and Methods

2.1. Fish Capture and Sampling

Two live female *E. muticus* (specimen numbers 20210826015 and 20210826016) were caught on 30 September 2021 during a bottom trawl survey of inshore fishery resources in Wuchuan City, Guangdong Province, China. The full body length, preanal length, and body weight of specimen 20210826015 were 38.5 cm, 9.8 cm, and 39.5 g, respectively, while those of specimen 20210826016 were 38.2 cm, 9.8 cm, and 42.3 g, respectively. These two fish were anesthetized with MS-222 (ethyl 3-aminobenzoate methanesulfonate, Sigma-Aldrich, Shanghai, China) at a concentration of 200 mg/L and immediately dissected for sterile anatomical sampling. The muscle, liver, brain, and heart tissues of each fish were collected separately in 1.5 mL sterile tubes and quickly frozen in liquid nitrogen. The frozen samples and fish bodies were transferred to a -80°C refrigerator at the Guangdong Ocean University for storage.

2.2. DNA and RNA Extraction and Sequencing

After extracting DNA from all tissues of each fish and conducting the quality inspection, one *E. muticus* (20210826015, Figure 1) with good DNA quality tissue samples was selected for subsequent genome sequencing. The genomic DNA from muscle tissue was extracted using the standard phenol/chloroform extraction protocol [24], and the DNA concentration, purity, and integrity were detected using Qubit 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and Agilent 4200 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), respectively. Subsequently, a paired-end sequencing library with an insertion length of 350 bp was constructed using the Illumina TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina NovaSeq-6000 platform (Illumina, San Diego, CA, USA). Raw reads from the Illumina sequencing were quality-filtered by FASTQ v0.23.2 [25], according to the setting criteria. Furthermore, SMRTbell template preparation kit 1.0 (Pacific Biosciences, Menlo Park, CA, USA) was used to construct the SMRTbell library with a fragment size of 20 kb from the same genomic DNA used for Illumina sequencing, following the manufacturer's protocol. After library construction, the accurate quantification and size of the SMRTbell library were detected using the Qubit 3.0 Fluorometer and Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), respectively. The library was then sequenced on the PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA, USA).

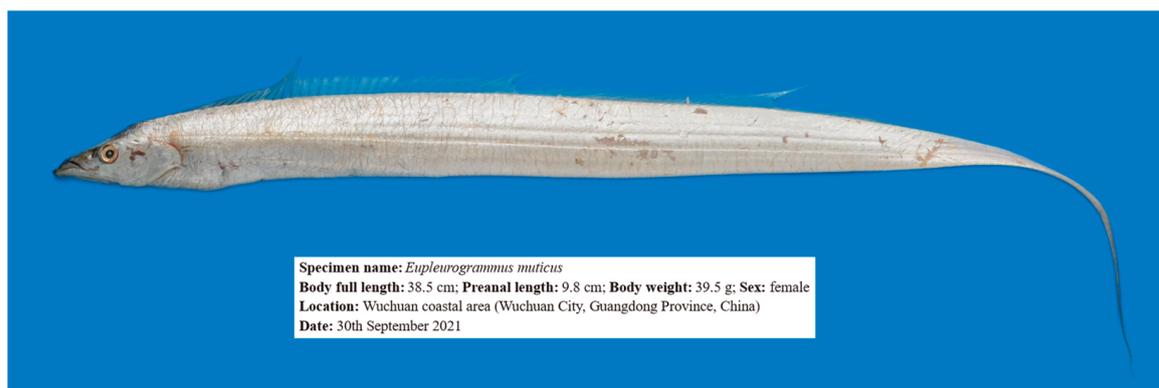


Figure 1. *E. muticus* used for sequencing.

The total RNA from the muscle, liver, brain, and heart tissues was sequenced to assist in the annotation of the *E. muticus* genome. RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), and its concentration and integrity were checked by Agilent 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA) and Agilent RNA ScreenTape Assay (Agilent Technologies, Palo Alto, CA, USA), respectively. Thereafter, the

RNA samples from the four tissues were evenly mixed for RNA library construction and sequencing. The mRNA was then purified from the mixed RNA using magnetic beads with Oligo (dT). The purified mRNA was reverse transcribed using Reagent TURScript First Stand cDNA Synthesis Kit (AidLab, Beijing, China) to synthesize a double-stranded cDNA library, and the Illumina pair-end sequencing library with an insert of about 350 bp was constructed. PCR Barcoding Kit (SQK-PBK004, Oxford Nanopore Technologies, Oxford, UK) and PCR-cDNA Sequencing Kit (SQK-PCS109, Oxford Nanopore Technologies, Oxford, UK) were used to construct a nanopore full-length transcriptome library. The Illumina and Nanopore libraries were then sequenced on the Illumina NovaSeq-6000 and PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK) platforms, respectively.

2.3. Evaluation of Genome Contamination, Size, Heterozygosity, and Repeat Sequence Rate

The Illumina sequencing raw reads of muscle DNA were filtered using FASTQ v0.23.2 to obtain clean reads for the preliminary estimation of genomic features. Blastn v2.11.0+ [26] was used to map 10,000 randomly selected clean reads (5000 for Read1 and 5000 for Read2) to the NCBI nucleotide database and rank the mapping times in descending order to show the top 80% of the mapped species. If all the mapped results were homologous, it was considered that there was no exogenous pollution in the sample. Genome size and heterozygosity were estimated using the GCE v1.0.0 [27] based on the K-mer frequency distribution method. Genome size (unit: Megabits) was calculated as the number of K-mer/depth of K-mer (K-mer = 17). The depth of K-mer was the expected value corresponding to the Poisson distribution. The corrected genome size was obtained after eliminating the error effect caused by the wrong K-mer. The heterozygosity rate is the estimated proportion of heterozygous sites in the sequence. The repeat sequence rate was calculated based on the area difference between the standard Poisson distribution and the actual data curve after the peak.

2.4. Genome Assembly and Assessment of Genomic Integrity and Consistency

The raw sequencing data obtained via PacBio SMRT technology contained a dumbbell-shaped structural sequence of two adaptors, called polymerase reads. Subreads were obtained after the adaptor sequences were interrupted and removed, and then the high-precision HiFi reads were generated via the Circular Consensus Sequencing (CCS) mode using SMRT-Link v10.2 [28]. The HiFi reads were then assembled using Hifiasm v0.16.1 [29], and the assembled genome was de-redundantly processed by Purge Haplotigs v1.0.4 [30].

Several methods were applied to evaluate the integrity and consistency of genome assemblies. First, the base composition and content of the genome sequence were statistically analyzed to preliminarily assess the assembly results. Secondly, the genomic sequences were interrupted in steps of 1000 bp and then compared against the NCBI nucleotide database using Blastn v2.11.0+, revealing the top five genera of the comparison results to confirm whether the assembled genome belongs to the target species. Thirdly, Illumina clean reads and PacBio HiFi reads were mapped to the genomic sequences using BWA v0.7.12 [31] and Minimap2 v2.22 [32], respectively. The integrity of the genome assembly and the uniformity of sequencing were assessed based on the mapping rate, coverage rate, and sequencing depth. Fourthly, mutations in the assembled genome were identified using Samtools v1.9 [33], Picard v1.124, and GATK v4.2.0.0, and the homozygous and heterozygous rates of single nucleotide polymorphisms (SNPs) and insertion-deletion (InDel) were counted, respectively. Finally, tblastn v2.11.0+ [26], Augustus v3.3.2 [34], and HMMER v3.3.1 [35] were used to map the assembled genome sequences to the single-copy orthologous gene database based on the Benchmarking Universal Single-copy Orthologs (BUSCO) evaluation method [31], and the genome integrity was evaluated according to the mapping results.

2.5. Genome Assembly at the Chromosomal Level

To obtain a high-quality genome, we further applied the Hi-C technology for genome assembly at the chromosomal level. First, the muscle tissue cells were treated with 40 mL of 2% formaldehyde solution (Sbjbio, Nanjing, China) for DNA cross-linking, and the sticky ends were generated by restriction endonuclease cleavage. Secondly, biotin-labeled oligonucleotide ends were added during end repair, and adjacent DNA fragments were ligated using nucleic acid ligase. Finally, the cross-linked protein and DNA were released, and biotin-labeled DNA fragments were captured to construct the Hi-C library, which was then sequenced on the Illumina NovaSeq-6000 platform.

The raw Hi-C sequencing data were filtered via standard quality control to obtain clean data, which included multiple types of reads such as valid pairs, contiguous sequences, internal fragments, and PCR repeats, among which only the valid pairs reflect the information of site-to-site interactions on the genome [36]. Therefore, the clean data were further filtered by the `hicup_filter` subroutine to obtain valid pairs, which were used as Di-Tags. The contigs or scaffolds of the same chromosome can be sorted and oriented based on the following criteria: cis interactions are much larger than trans interactions, and the closer the linear distance in cis interactions, the stronger the interactions [36]. Accordingly, the 3D-DNA program [37] was used to assemble Di-Tags and to cluster the assembled contig and scaffold sequences to obtain a chromosome-level genome. The interaction map of the assembled genome was constructed using Juicer v1.6 [38] and visualized via JuiceBox v1.11.08 [39].

2.6. Genome Prediction and Annotation

Genome annotation mainly includes repetitive sequence annotation, gene annotation (prediction of gene structure and function), and non-coding RNA (ncRNA) annotation. Repetitive sequences consist of tandem and interspersed repeats. Tandem repeats in the genome sequence were searched by TRF v4.09 [40], while the interspersed repeats were identified via homology prediction and de novo prediction methods. The homology prediction was based on the homologous repeat database Repbase [41]. RepeatMasker v4.0.9 and RepeatProteinMask v4.0.9 [42] were used to identify sequences with similar repeat sequences of known nucleic acids and amino acids, respectively. The de novo prediction was achieved through the RepeatMasker program based on the creation of a new repeat sequence database using LTR_Finder v1.0.7 [43] and RepeatModeler v1.0.11 [44].

The location and structure of the protein-coding genes were predicted using three methods: homology-based, de novo, and transcriptome-based prediction. The homology-based prediction involved first downloading the genome-wide protein sequences of *Homo sapiens*, *Danio rerio*, *Thunnus maccoyii*, *Thunnus albacares*, *Takifugu rubripes*, *Oryzias latipes*, *Etheostoma spectabile*, *Sander lucioperca*, *Perca flavescens*, and *Larimichthys crocea* from the NCBI database and aligning them to the *E. muticus* genome using `tblastn` (E-value $\leq 1 \times 10^{-5}$). The alignments were then analyzed using GeneWise v2.4.1 [45] to define the protein-coding gene models. De novo prediction was performed via Augustus v3.3.2 and GeneScan v1.0.0 [46]. The RNA-Seq data of the tissues were assembled by Tophat v2.1.1 for comparison and Cufflinks v2.2.1 [47] to obtain the transcripts. Subsequently, the gene sets predicted by these methods were integrated into a non-redundant and more complete gene set through Maker2 v2.31.10 [48], and the results of the Core Eukaryotic Genes Mapping Approach (CEGMA) [49] were integrated using the HiCESAP process to obtain the final reliable gene set. Finally, the `tblastn` program was used to map the gene sequences to non-redundant (NR), Swiss Protein institute (SwissProt) [50], Kyoto Encyclopedia of Genes and Genomes (KEGG) [51], and gene ontology (GO) [52] databases. Pfam [53] and InterPro databases were then used to predict the protein family and conserved functional domains of protein-coding genes.

The transfer RNA (tRNA) sequences in the genome can be searched by tRNAscan-SE v1.3.1 [54] based on their structural characteristics. Since ribosomal RNAs (rRNAs) are highly conserved, the rRNA sequences of the species closely related to *E. muticus* were

selected as reference sequences, and blastn alignment was used to find rRNAs in the genome. MicroRNA (miRNA) and small nuclear RNA (snRNA) sequences in the genome were predicted by infernal software of Rfam based on the covariance model of the Rfam v14.0 family [55].

2.7. Gene Family Identification and Dynamics Analysis and Phylogenetic Tree Construction

Genomic comparative analysis was conducted between *E. muticus* and 19 other selected fishes (*L. savala*, *T. albacares*, *T. maccoyii*, *Epinephelus akaara*, *Epinephelus fuscoguttatus*, *P. flavescens*, *L. crocea*, *S. lucioperca*, *Acanthopagrus latus*, *Cheilinus undulatus*, *Echeneis naucrates*, *Periophthalmus magnuspinnatus*, *T. rubripes*, *Takifugu flavidus*, *Tetraodon nigroviridis*, *Monopterus albus*, *Hippocampus comes*, *Latimeria chalumnae*, and *Rhincodon typus*) to identify the gene family of *E. muticus*. First, the protein-coding genes with less than 30 amino acids were excluded from the genomes of all species, retaining the protein sequence of the longest transcript. Secondly, the similarity relationship among the protein sequences of all species was calculated by all-vs-all blastp v2.11.0+ (E-value = 1×10^{-5}). Finally, orthologous genes were clustered by OrthoMCL v2.0.9 [56] (expansion coefficient = 1.5) based on their similarity to obtain single-copy genes and gene families.

All single-copy genes were aligned with multiple sequences using MAFFT v7.487 [57], and the results were combined into a super alignment matrix. The maximum likelihood (ML) phylogenetic tree of the above 20 species was constructed using RAxML v8.2.12 [58]. Based on the constructed phylogenetic tree, seven divergence times were obtained from the TimeTree database [59] to be used for calibration, including *R. typus* and *L. chalumnae* (442.7–515.5 Mya), *L. chalumnae* and *P. magnuspinnatus* (424.2–440 Mya), *E. muticus* and *T. albacares* (30.1–58.1 Mya), *M. albus* and *P. flavescens* (103.7–176.0 Mya), *A. latus* and *T. rubripes* (59.3–142.9 Mya), *C. undulatus* and *L. crocea* (63.9–114.6 Mya), and *T. maccoyii* and *M. albus* (106–144 Mya). Interspecific divergence times (95% confidence interval) of the above 20 species were estimated according to the seven divergence times using the default parameters of McMcTree v4.9 [60] in the PAML package [61]. Furthermore, the expansion and contraction analyses of the gene family were conducted using CAFE5 v5.0.0 [62], and the GO and KEGG enrichment analyses were further performed to explore the function of the genes and the biological processes involved. GO terms or KEGG pathways with a *p*-value < 0.05 and an FDR (false discovery rate) < 0.05 were considered to be statistically significant.

2.8. Positive Selection and Collinearity Analyses

Three groups of positive selection analyses were set up to screen candidate genes related to the unique traits of *E. muticus*. Group 1 (*E. muticus* and *L. savala*) vs. (*A. latus*, *E. fuscoguttatus*, *E. akaara*, *P. flavescens*, *L. crocea*, and *S. lucioperca*) and group 2 (*E. muticus*) vs. (*A. latus*, *E. fuscoguttatus*, *E. akaara*, *P. flavescens*, *L. crocea*, and *S. lucioperca*) were set to screen for genes associated with the body characteristics of *E. muticus*. Group 3 (*E. muticus* and *L. savala*) vs. (*T. albacares* and *T. maccoyii*) was set to screen the positively selected genes between the trichiurids and closely related species. The target species were identified as the foreground branch, and the remaining species as background branches. The positive selection effects acting on protein-coding sequences were detected by CodeML v4.9 [60] in PAML based on the branch-site model. The protein sequences in each single-copy gene were subjected to multiple alignments using MAFFT v7.487, and the result was subjected to a multiple-sequence alignment of the coding sequences. The likelihood values were calculated using the branch-site model analysis [63] based on two models (Model A and null mode), and the values were further subjected to likelihood ratio tests (LRTs) via the chi2 program in PAML (with a correct *p*-value of FDR < 0.05). The posterior probability of the positive selection was calculated using the Bayes empirical Bayes method (BEB) [64]. Finally, GO and KEGG enrichment analyses were performed on the positively selected genes to explore their gene functions.

There are two types of collinearity analysis: coding sequences collinearity (at the protein level) and genome-wide collinearity (at the DNA level) [65]. These two types of collinearity analyses were performed between *E. muticus* and *L. savala* using JCVI v1.1.22 [66] (for coding sequence collinearity) and Mummer v4.0.0rc1 [67] (for genome-wide collinearity).

3. Results

3.1. Genome Sequencing Data

A total of 475,992,654 raw paired reads (71.4 Gb) were generated by Illumina sequencing, and 421,436,106 clean paired reads (63.03 Gb) were obtained after data filtering and de-redundancy (Table 1). The GC content of the clean reads was 39.92%, and the proportions of base quality at >Q20 and >Q30 were 97.04% and 92.23%, respectively. The mapping results showed that the top 80% of the 29 genera were all Actinopterygii, with Perciformes having the most genera (21 genera, Table S1), including *Epinephelus* (9.58%), *Lateolabrax* (8.23%), *Plectropomus* (6.57%), *Trachurus* (4.78%), *Sparus* (4.55%), *Larimichthys* (4.55%), and *Nibeia* (4.18%), among others. This indicated that the Illumina sequencing data were reliable and not contaminated by external sources. The depth of K-mer was determined to be 79 according to the expected value of the Poisson distribution with K-mer = 17 (Figure S1). Based on this, the estimated genome size of *E. muticus* was 673 Mb, which was revised to 664 Mb, with a genomic heterozygosity rate of 1.18% and a repeat sequence rate of 35.43%.

Table 1. Statistics of the sequencing data of the *E. muticus* genome.

Type	Platform	Library Size (bp)	Raw Data (Gb)	Clean Data (Gb)	Coverage (×)
Illumina Nova	Illumina NovaSeq-6000	350	71.40	63.03	99.1
PacBio SMRT	PacBio Sequel II	15 k	59.57	34.97	48.5
Hi-C	Illumina NovaSeq-6000	350	75.20	74.25	104.4
Illumina RNA-Seq	Illumina NovaSeq-6000	350	7.10	6.71	9.8
ONT RNA-Seq	NanoPromethION	-	9.09	8.61	12.6
Total			222.36	187.57	274.4

3.2. Genome Assembly and Evaluation

In total, 2,337,277 high-quality HiFi reads (34.97 Gb) were obtained by PacBio SMRT sequencing (Table 1). The average length of the HiFi reads was 14,963.26 bp, and their N50 length was 15,643 bp. After assembly error correction and elimination of redundant sequences, 156 contigs were obtained, with the maximum (max) length, N50 length, and N90 length of 50,599,264 bp, 25,347,879 bp, and 4,252,074 bp, respectively (Table S2). The genome size assembled at the contig level was 709.27 Mb, which is close to the estimated value (664 Mb) of the genome survey.

The mapping rate of Illumina clean reads was 99.50%, with an average sequencing depth of 86.61× and a genome coverage rate of 99.93%. However, the mapping rate, average sequencing depth, and genome coverage rate of PacBio HiFi reads were 99.99%, 48.52×, and 99.99%, respectively. The correlation graph between the GC content and average depth distribution showed that the GC content was concentrated around 40.81% (Figure S2), without significant GC content separation, indicating that there was no exogenous pollution in the genome. SNP and InDel identification analyses showed that the homozygous SNP and InDel rates were 0.001% and 0.002%, and the heterozygous SNP and InDel rates were 0.936% and 0.294%, respectively. The extremely low rate of homology SNPs indicated that the assembled genome had a high single-base accuracy.

Based on the BUSCO method, the database with 3640 orthologous single-copy genes was used as a reference to evaluate the integrity of the assembled genome. The results showed that the *E. muticus* genome contained 3534 (97.1%) complete BUSCOs, of which 3471 (95.4%) were complete single-copy BUSCOs, 63 (1.7%) were complete duplicated BUSCOs, 12 (0.3%) were fragmented BUSCOs, and 94 (2.6%) were missing BUSCOs

(Table S3). This suggested that the assembled genome contained over 97.1% of orthologous genes, indicating a high gene coverage rate.

3.3. Hi-C Technology-Assisted Genome Assembly at the Chromosomal Level

A total of 501,341,846 raw paired reads (75.20 Gb) were obtained by Illumina sequencing (Table 1), and 495,369,140 clean reads (74.25 Gb) were obtained after quality control, with 96.54% of Q20 and 91.06% of Q30. The mapping results exhibited that 132,890,440 read1 and 132,890,440 read2 were successfully matched (Table S4). Among them, valid pairs accounted for 70.62% (93,847,523), and the proportion was 51.23% (68,084,802) after de-duplication (Table S5, Figure S3).

The Hi-C-assembled genome at the chromosomal level had 84 contigs (N50 length: 25,078,085 bp) and 60 scaffolds (N50 length: 30,064,390 bp). Among these, 24 sequences (691,679,068 bp) were assembled, while 60 sequences (10,817,913 bp) were not assembled at the chromosomal level, resulting in a genome assembly mounting rate of 98.46% (Table S6). As shown in Figure 2, the species chromosome interaction mapping was consistent with the genome-wide interaction mapping, indicating that the Hi-C-assisted assembly was good.

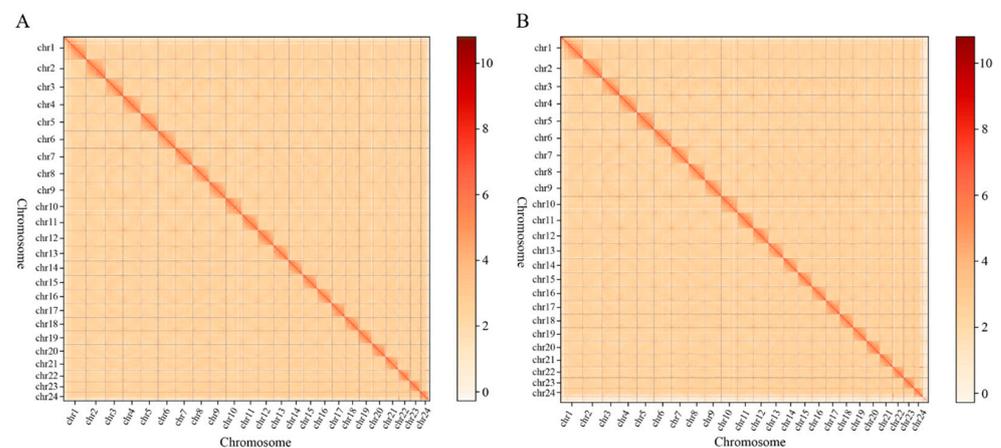


Figure 2. Chromosome interaction mapping (A) and genome-wide interaction mapping (B). The color reflects the intensity of each contact, with deeper colors representing higher intensity.

3.4. Genome Annotation

In total, 72,738,949 bp tandem repeats and 216,702,465 bp interspersed repeats were predicted, accounting for approximately 10.35% and 30.85% of the genome, respectively. Transposable elements (TEs) accounted for the highest number of repeats (102,347,937 bp; 14.57%), followed by long interspersed nuclear elements (LINEs) (36,949,851 bp; 5.26%), long terminal repeats (LTRs) (25,110,984 bp; 3.57%), and short interspersed nuclear elements (SINEs) (9,190,875 bp; 1.31%) (Table S7). A total of 29,052 non-coding RNAs were predicted in the genome, including 879 miRNAs (75,630 bp; 0.0108%), 12,792 tRNAs (968,630 bp; 0.1379%), 13,817 rRNAs (1,993,777 bp; 0.2838%), and 1537 snRNAs (236,834 bp; 0.0337%) (Table S8).

As illustrated in Figure 3, the circle figure of the genome characteristics of *E. muticus* was constructed based on the 24 assembled chromosomes, which showed the distribution of protein-coding genes, repeats, LTR, LINE, and DNA-TE on the 24 chromosomes. Among them, there were 21,492 genes with functional annotation, accounting for 97.92% of the total protein-coding genes (21,949) (Table S9). The basic details of 21,446 genes were obtained from the NR database, while those of 19,425 genes were obtained from the SwissProt databases. Moreover, the biological processes and functions of 15,253 genes were obtained from the GO database, while those of 21,322 genes were derived from the KEGG databases. Annotation information on the functional domains and protein families of 19,218 genes was acquired from the Pfam databases, while that of 19,940 genes was obtained from InterPro

databases. The average lengths of all genes and coding sequences (CDS) were 15,160 bp and 1728 bp, respectively. The average number of exons per gene was 10.27, and the average lengths of exons and introns were 259.53 bp and 1348 bp, respectively (Table 2). The gene structure comparison between *E. muticus* and the other 10 fishes showed that the distribution of exon and CDS lengths was highly consistent among all fishes (Figure S4), indicating their conservation during fish genome evolution.

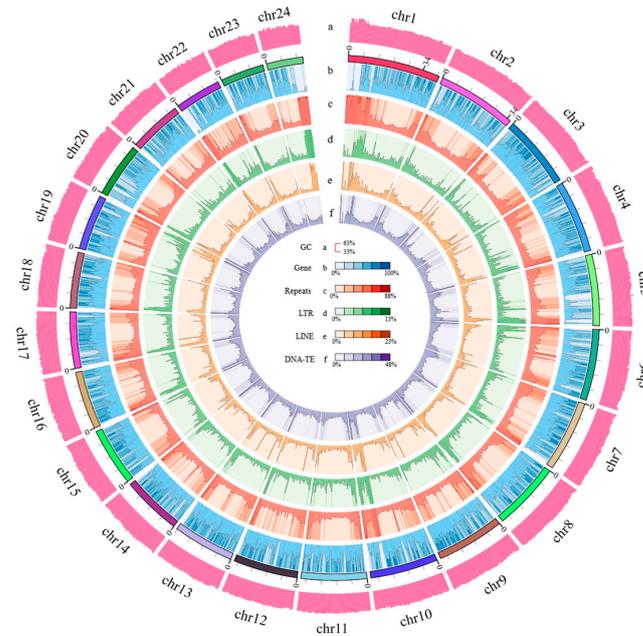


Figure 3. Circle figure of the genomic characteristics of *E. muticus*, including (a) the GC content of the genome, (b) the distribution of genes, (c) the distribution of repeats, (d) the distribution of long tandem repeats, (e) the distribution of long interspersed nuclear elements, and (f) the distribution of DNA transposable elements.

Table 2. Gene structure and parameters predicted by three methods.

Methods	Gene Set	Number	Average Gene Length (bp)	Average CDS Length (bp)	Average Exon Per Gene	Average Exon Length (bp)	Average Intron Length (bp)
De novo	Genscan	36,114	13,128.07	1419.77	7.78	182.56	1727.59
	AUGUSTUS	36,217	9868.25	1270.82	6.96	182.72	1443.6
Homolog	<i>Etheostoma spectabile</i>	30,817	12,295.23	1472.87	8.29	177.59	1483.41
	<i>Homo sapiens</i>	24,338	13,205.29	1321.50	8.24	160.32	1640.46
	<i>Larimichthys crocea</i>	32,085	12,203.04	1478.14	8.27	178.64	1474.12
	<i>Thunnus maccoyii</i>	32,571	12,375.79	1501.14	8.32	180.47	1485.64
	<i>Perca flavescens</i>	31,479	12,242.79	1463.95	8.30	176.36	1475.98
	<i>Takifugu rubripes</i>	28,393	12,451.37	1509.68	8.63	174.96	1433.9
	<i>Thunnus albacares</i>	32,146	12,393.48	1508.74	8.43	179.07	1465.57
	<i>Danio rerio</i>	29,225	12,428.59	1442.84	8.26	174.65	1512.68
	<i>Sander lucioperca</i>	32,372	12,404.23	1482.49	8.32	178.23	1492.16
	<i>Oryzias latipes</i>	30,041	12,304.19	1507.13	8.34	180.76	1471.05
Transcriptome	RNAseq	10,859	17,854.09	1764.30	11.53	312.82	1353.38
	ISOseq	1698	10,846.91	1173.28	9.55	231.66	1010.05
BUSCO		3661	11,770.51	1843.36	12.31	149.74	877.63
MAKER		22,903	14,826.32	1668.48	9.72	259.16	1411.85
HiCESAP		21,949	15,159.85	1727.52	10.27	259.53	1347.92

3.5. Gene Family Clustering, Expansion, and Contraction and Phylogenetic Analyses

The protein-coding genes screened from the genomes of *E. muticus* and 19 other fish species ranged from 20,932 (*L. chalumnae*) to 29,408 (*T. flavidus*) (Figure 4A), and the cluster analysis yielded 3006 single-copy genes (Table 3). In the *E. muticus* genome, 20,826 genes were clustered into 15,686 gene families, of which 6711 were common families and 25 were unique. There were 13,805 gene families shared by *E. muticus* and 3 closely related species (*L. savala*, *T. albacares*, and *T. maccoyii*), while 435 gene families were unique to *E. muticus* (Figure 4B). KEGG enrichment analysis showed that these unique gene families were mainly involved in several pathways, such as neutrophil extracellular trap formation, systemic lupus erythematosus, alcoholism, shigellosis, transcriptional misregulation in cancer, the cytosolic DNA-sensing pathway, and herpes simplex virus 1 infection (Supplementary Material S2).

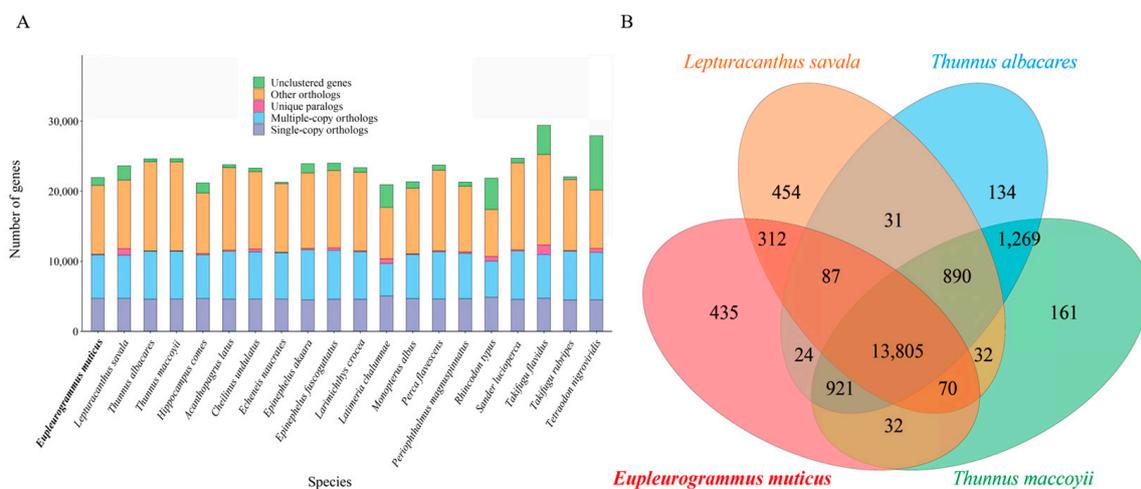


Figure 4. The numbers of homologous genes in 20 fish species (A) and Venn diagram of the homologous gene families between *E. muticus* and three closely related species (B).

Based on the comparison of common ancestors between *E. muticus* and *L. savala*, 143 gene families expanded while 708 gene families contracted during the genomic evolution of *E. muticus* (Figure 5). Additional GO and KEGG enrichment analyses revealed 31 significantly expanded gene families (including 124 genes) and 123 significantly contracted gene families (including 52 genes) (Supplementary Material S3). KEGG enrichment analysis also revealed that the expanded gene families were involved in the following pathways: cytokine–cytokine receptor interaction, viral protein interaction with cytokine and cytokine receptor, chemical carcinogenesis-DNA adducts, Notch signaling pathway, and breast cancer; however, the contracted gene families were mainly involved in ABC transporters, axon guidance, antifolate resistance, antigen processing and presentation, and longevity regulating pathway-multiple pathways (Table 4). According to GO enrichment analysis, the expanded gene families were mainly enriched in various terms, including cellular processes, binding, metabolic processes, cellular metabolic processes, nitrogen compound metabolic processes, primary metabolic processes, and organic substance metabolic processes; however, the contracted gene families were mainly enriched in binding, organic cyclic compound binding, heterocyclic compound binding, catalytic activity, cellular process, ATP binding, and adenylyl nucleotide binding. In general, the expanded gene families were mainly concentrated in metabolic processes, while the contracted families were mainly associated with binding processes.

Table 3. The gene family clustering in 20 fish species.

Species	Genes	Unclustered	Genes	Family	Unique	Unique	Common	Common	Single-Copy
	Number	Genes	In Families	Number	Families	Families Genes	Families	Families Genes	Genes
<i>Eupleurogrammus muticus</i>	21,949	1123	20,826	15,686	25	89	6711	10,925	3006
<i>Lepturacanthus savala</i>	23,625	2040	21,585	15,681	47	926	6711	10,880	3006
<i>Thunnus albacares</i>	24,623	429	24,194	17,161	23	49	6711	11,427	3006
<i>Thunnus maccoyii</i>	24,646	475	24,171	17,180	22	61	6711	11,423	3006
<i>Hippocampus comes</i>	21,175	1439	19,736	14,602	67	170	6711	10,931	3006
<i>Acanthopagrus latus</i>	23,773	405	23,368	16,666	57	168	6711	11,429	3006
<i>Cheilinus undulatus</i>	23,303	521	22,782	15,995	86	410	6711	11,358	3006
<i>Echeneis naucrates</i>	21,275	194	21,081	15,518	21	62	6711	11,220	3006
<i>Epinephelus akaara</i>	23,923	1322	22,601	16,077	66	184	6711	11,659	3006
<i>Epinephelus fuscoguttatus</i>	24,005	1055	22,950	16,191	121	338	6711	11,574	3006
<i>Larimichthys crocea</i>	23,354	660	22,694	16,573	43	118	6711	11,372	3006
<i>Latimeria chalumnae</i>	20,932	3250	17,682	13,208	174	696	6711	9681	3006
<i>Monopterus albus</i>	21,343	915	20,428	15,304	41	99	6711	10,979	3006
<i>Perca flavescens</i>	23,736	739	22,997	16,417	41	122	6711	11,376	3006
<i>Periophthalmus magnuspinnatus</i>	21,293	597	20,696	15,033	53	223	6711	11,140	3006
<i>Rhincodon typus</i>	21,868	4491	17,377	12,554	193	662	6711	10,029	3006
<i>Sander lucioperca</i>	24,714	687	24,027	16,892	60	157	6711	11,484	3006
<i>Takifugu flavidus</i>	29,408	4177	25,231	15,690	261	1349	6711	10,980	3006
<i>Takifugu rubripes</i>	22,064	407	21,657	15,320	42	107	6711	11,445	3006
<i>Tetraodon nigroviridis</i>	27,918	7741	20,177	14,512	222	541	6711	11,285	3006

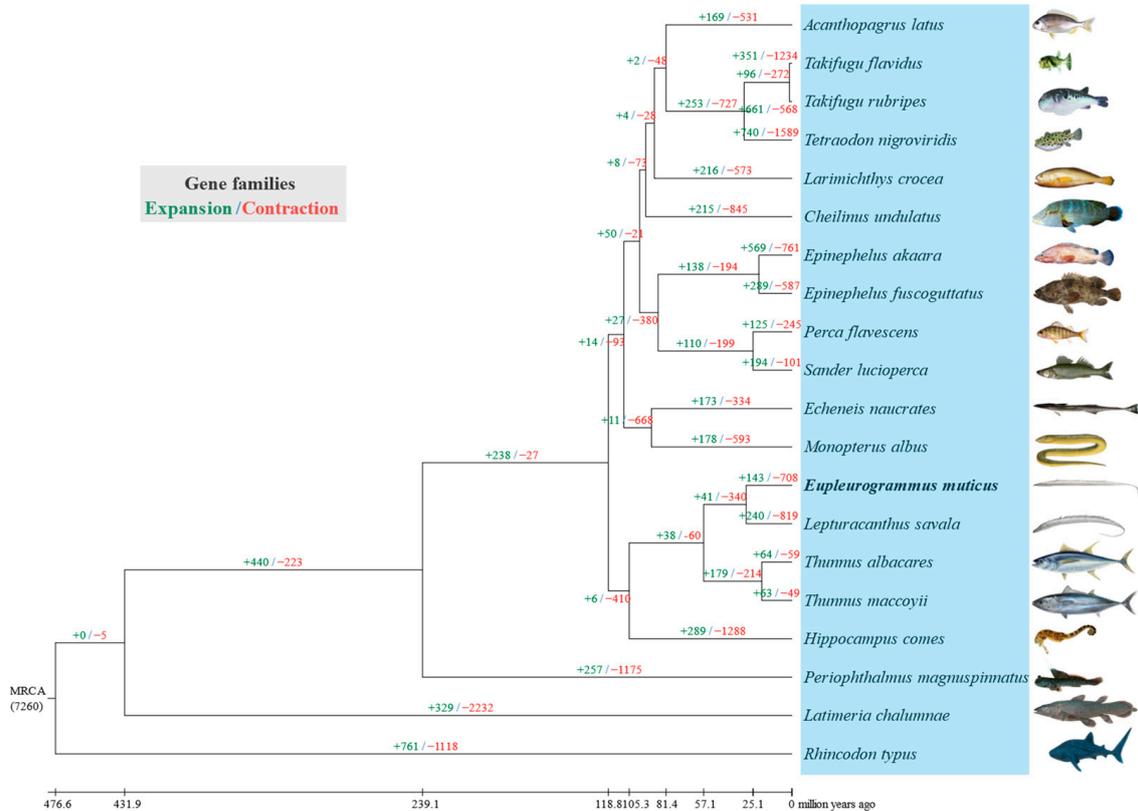


Figure 5. The expanded and contracted gene families of the 20 fish species during in the evolutionary process.

Table 4. KEGG enrichment analysis of expanded and contracted gene families.

1. Expansion (67 Gene Families, Top 20 KEGG Pathways, <i>p</i>-Value < 0.05)		
KEGG Pathways	<i>p</i>-Value	Genes
Ascorbate and aldarate metabolism	2.97×10^{-7}	<i>ugt3, ugt1a1</i>
Pentose and glucuronate interconversions	5.85×10^{-7}	<i>ugt3, ugt1a1</i>
Chemical carcinogenesis—DNA adducts	5.85×10^{-7}	<i>ugt3, ugt1a1</i>
Porphyrin and chlorophyll metabolism	7.94×10^{-7}	<i>ugt3, ugt1a1</i>
Drug metabolism—cytochrome P450	1.21×10^{-6}	<i>ugt3, ugt1a1</i>
Metabolism of xenobiotics by cytochrome P450	1.39×10^{-6}	<i>ugt3, ugt1a1</i>
Notch signaling pathway	1.97×10^{-6}	<i>hes5</i>
Steroid hormone biosynthesis	3.58×10^{-6}	<i>ugt3, ugt1a1</i>
Retinol metabolism	8.61×10^{-6}	<i>ugt3, ugt1a1</i>
Drug metabolism—other enzymes	2.24×10^{-5}	<i>ugt3, ugt1a1</i>
Steroid biosynthesis	1.74×10^{-4}	<i>soat1</i>
Bile secretion	2.24×10^{-4}	<i>ugt3, ugt1a1</i>
Viral protein interaction with cytokine and cytokine receptor	5.96×10^{-4}	<i>ccr3, ccr5</i>
Breast cancer	8.91×10^{-4}	<i>hes5</i>
Hypertrophic cardiomyopathy	0.001042306	<i>ttn</i>
Chemokine signaling pathway	0.001245404	<i>tiam1, ccr3, ccr5</i>
Epithelial cell signaling in Helicobacter pylori infection	0.001703999	<i>ptprz1</i>
Human papillomavirus infection	0.001826957	<i>hes5, dlg1l</i>
Dilated cardiomyopathy	0.002127476	<i>ttn</i>
Cholesterol metabolism	0.004796358	<i>soat1</i>
2. Contraction (123 gene families, Top 20 KEGG pathways, <i>p</i>-value < 0.05)		
KEGG Pathways	<i>p</i>-Value	Genes
ABC transporters	4.31×10^{-20}	<i>abcc8, abcc12, abcc10, abcc5, etc.</i>
Axon guidance	2.48×10^{-19}	<i>epha2, epha8, epha3, epha6, etc.</i>
Antifolate resistance	4.78×10^{-11}	<i>abcc4, abcc5, abcc2, abcc3, etc.</i>
Antigen processing and presentation	2.85×10^{-9}	<i>hspa5, hsc71, hspa8, hsc70</i>
Legionellosis	1.63×10^{-7}	<i>hsc71, hspa8, hsp70</i>
Longevity regulating pathway—multiple species	8.54×10^{-7}	<i>hsc71, hspa8, hsp70</i>
Systemic lupus erythematosus	2.52×10^{-6}	<i>h3f3b, hist2h3d</i>
Protein processing in endoplasmic reticulum	2.71×10^{-6}	<i>hsc71, hspa8, hspa5, hsp70</i>
Toxoplasmosis	5.46×10^{-6}	<i>hsc71, hspa8, hsp70</i>
Spliceosome	5.70×10^{-6}	<i>hsc71, hspa8, hsp70</i>
Measles	1.04×10^{-5}	<i>hsc71, hspa8, hsp70</i>
MAPK signaling pathway	1.06×10^{-5}	<i>hsc71, hspa8, epha2, hsp70</i>
Estrogen signaling pathway	1.89×10^{-5}	<i>hsc71, hspa8, hsp70</i>
Lipid and atherosclerosis	2.53×10^{-5}	<i>hspa5, hsc71, hspa8, hsp70</i>
Neutrophil extracellular trap formation	3.25×10^{-5}	<i>h3f3b, hist2h3d</i>
Alcoholism	4.49×10^{-5}	<i>h3f3b, hist2h3d</i>
Prion disease	1.26×10^{-4}	<i>hsc71, hspa5, hspa8, hsp70</i>
Transcriptional misregulation in cancer	2.07×10^{-4}	<i>h3f3b, hist2h3d</i>
Vitamin digestion and absorption	2.46×10^{-4}	<i>abcc1</i>
Bile secretion	4.76×10^{-4}	<i>abcc4, abcc2, abcc3</i>

The ML phylogenetic tree was constructed based on the sequences of 3006 single-copy genes shared by 20 fish species (Figure S5). The results showed that *E. muticus* first clustered together with *L. savala* and then formed a sister-group relationship with *T. albacares* and *T. maccoyii*, and the nodes of all branches had 100% bootstrap support. Based on the estimated divergence time (Figure 6), the divergence between *E. muticus* and *L. savala* occurred 29.6 (16.7–40.5) million years ago, and they differentiated with *T. albacares* and *T. maccoyii* approximately 57.1 (50.8–60.9) million years ago.

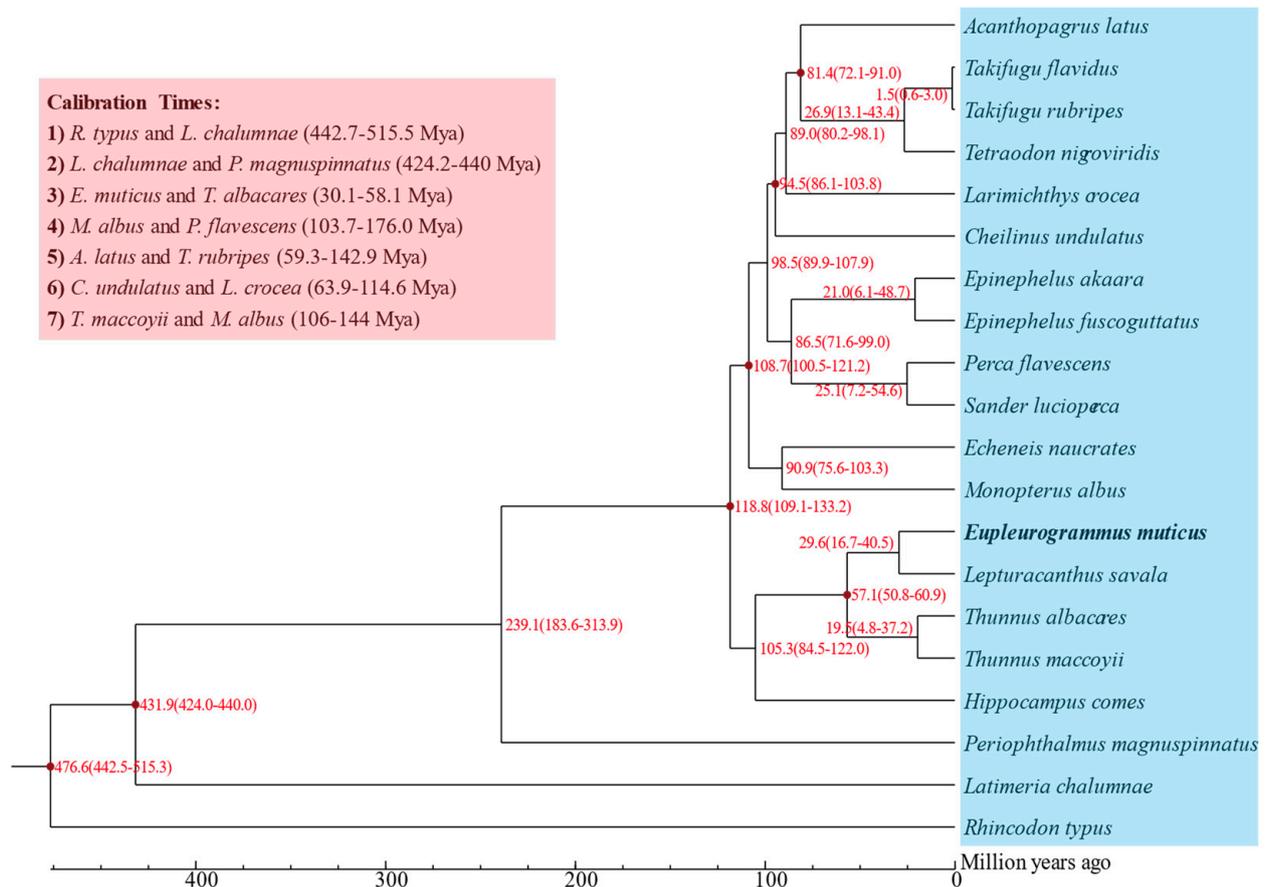


Figure 6. Divergence time estimates of the 20 fish species.

3.6. Positive Selection and Collinearity Analyses

The results of positive selection analyses (Table 5, Supplementary Material S4) showed that there were 1566 positively selected genes and 21 significantly enriched pathways (mainly Fanconi anemia, non-homologous end-joining, homologous recombination, ferroptosis, and cytokine–cytokine receptor interaction pathways) detected in group 1. In group 2, there were 1022 positively selected genes and 20 significantly enriched pathways (mainly cytokine–cytokine receptor interaction, lysosome, JAK-STAT signaling pathway, Fanconi anemia, and homologous recombination). There were 2300 positively selected genes and 17 significantly enriched pathways (mainly cytokine–cytokine receptor interaction, viral protein interaction with cytokine and cytokine receptor, base excision repair, Fanconi anemia, and complement and coagulation cascades) in group 3. As shown in the two collinearity analysis figures (Figures 7 and S6), the genomes of *E. muticus* and *L. savala* had a high degree of collinearity, with a one-to-one correspondence of the 24 chromosomes of the two species.

Table 5. The results of the positive selection analysis of *E. muticus*.

Group 1 (Genes: 1566; KEGG Pathways: 21, <i>p</i> -Value < 0.05)		
<i>(Eupleurogrammus muticus and Lepturacanthus savala) vs. (Acanthopagrus latus, Epinephelus fuscoguttatus, Epinephelus akaara, Perca flavescens, Larimichthys crocea, and Sander lucioperca)</i>		
KEGG Pathways	<i>p</i> -Value	Genes
Novobiocin biosynthesis	0.00	<i>tat</i>
Fanconi anemia pathway	3.59×10^{-6}	<i>slx1a, eme1, palb2, fanca</i> , etc.
Non-homologous end-joining	0.001790458	<i>polm, nhej1, dntt</i>
Homologous recombination	0.003095449	<i>eme1, blm, brca1, bard1</i> , etc.
Ferroptosis	0.004991427	<i>atg5, atg7, map1lc3b, map1lc3c</i> , etc.
Bacterial secretion system	0.004708158	<i>srp54</i>
Base excision repair	0.006208329	<i>mpg, parp4, smug1, neil3</i> , etc.
RNA transport	0.007638419	<i>EIF2B3, gemin5, nup188, acin1</i> , etc.
Oxidative phosphorylation	0.009939509	<i>ndufb9, ppa1, atp6v0b, atp5g3</i> , etc.
Cell cycle—caulobacter	0.01347929	<i>clpp</i>
Sesquiterpenoid and triterpenoid biosynthesis	0.01347929	<i>sqle</i>
Caffeine metabolism	0.02573447	<i>uox</i>
Aminobenzoate degradation	0.02573447	<i>echS1</i>
Naphthalene degradation	0.02573447	<i>adh5</i>
DNA replication	0.03018216	<i>rnaseh2b, zmcm3, dna2, rfc1</i> , etc.
Autophagy—yeast	0.03827173	<i>rab7, ip6k1, vps8, kras</i> , etc.
Ribosome biogenesis in eukaryotes	0.03438198	<i>EIF6, nob1, pop4, dkc1</i> , etc.
One carbon pool by folate	0.03736458	<i>dhfr, mthfr, mthfd2</i> , etc.
Cytokine–cytokine receptor interaction	0.04263281	<i>il1r1l, il17a, ngfr, tnfrsf1b</i> , etc.
Ubiquinone and other terpenoid-quinone biosynthesis	0.04447989	<i>tat, coq6</i>
Citrate cycle (TCA cycle)	0.04820553	<i>dld, dlat, sdha, suclg1</i> , etc.
Group 2 (Genes: 1022; KEGG Pathways: 20, <i>p</i> -Value < 0.05)		
<i>(Eupleurogrammus muticus) vs. (Acanthopagrus latus, Epinephelus fuscoguttatus, Epinephelus akaara, Perca flavescens, Larimichthys crocea, and Sander lucioperca)</i>		
KEGG Pathways	<i>p</i> -Value	Genes
Novobiocin biosynthesis	0.00	<i>tat</i>
Fanconi anemia pathway	7.11×10^{-8}	<i>palb2, fance, faap100, blm</i> , etc.
Homologous recombination	0.000144	<i>eme1, xrcc3, palb2, brip1</i> , etc.
Mismatch repair	0.00209	<i>msh3, pold1, exo1</i>
Cytokine–cytokine receptor interaction	0.00485	<i>il1r1l, ngfr, prlr, il22ra1</i> , etc.
Aminobenzoate degradation	0.0105	<i>ehhadh</i>
Lysosome	0.012	<i>lamp2, lipa, man2b1, ppt1</i> , etc.
Base excision repair	0.0127	<i>mpg, parp4, pold1</i>
Notch signaling pathway	0.0151	<i>cir1, dtx3, dtx2, maml2</i> , etc.
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	0.0157	<i>pigk, pigo, pgap1</i>
Ribosome biogenesis in eukaryotes	0.0163	<i>droscha, riok2, dkc1, rpp40</i> , etc.
Non-homologous end-joining	0.0203	<i>nhej1</i>
Tropane, piperidine and pyridine alkaloid biosynthesis	0.0337	<i>tat</i>
Caprolactam degradation	0.0337	<i>ehhadh</i>
JAK-STAT signaling pathway	0.0358	<i>ccnd3, il22ra1, prlr, il12b</i> , etc.
Ribosome	0.0361	<i>rps26, rpl29, mrps18c, rpl23</i> , etc.
Aminoacyl-tRNA biosynthesis	0.0397	<i>sars, sepsecs, wars2</i>
Nucleotide excision repair	0.0432	<i>ercc3, ercc5, pold1, ccnh</i>
Phenylalanine, tyrosine and tryptophan biosynthesis	0.0437	<i>tat</i>
Phenylalanine metabolism	0.0461	<i>tat, mif</i>

Table 5. Cont.

Group 3 (Genes: 2300; KEGG Pathways: 17, <i>p</i> -Value < 0.05)		
<i>(Eupleurogrammus muticus and Lepturacanthus savala) vs. (Thunnus albacares and Thunnus maccoyii)</i>		
KEGG Pathways	<i>p</i> -Value	Genes
Cytokine–cytokine receptor interaction	5.95×10^{-5}	<i>ngfr, ccr6, il17a, il22ra1</i> , etc.
Fanconi anemia pathway	6.21×10^{-4}	<i>eme1, palb2, fanca, bripl1</i> , etc.
Base excision repair	7.34×10^{-4}	<i>mpg, lig3, smug1, nthl1</i> , etc.
Viral protein interaction with cytokine and cytokine receptor	8.07×10^{-4}	<i>il6, ccr6, tnfsf14, il22ra1</i> , etc.
Lipoic acid metabolism	0.009138648	<i>lipt2</i>
Sphingolipid metabolism	0.01278764	<i>psap, plpp3, cerk, cers5</i> , etc.
DNA replication	0.01548808	<i>rnaseh2b, pole2, dna2, prim2</i> , etc.
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	0.01839599	<i>pigk, pigo, pigw, pigt</i> , etc.
Complement and coagulation cascades	0.01912686	<i>c5, c9, c1s, c1qa</i> , etc.
Ribosome	0.02090745	<i>mrpl16, rps10, mrpl21, mrpl11</i> , etc.
Ubiquinone and other terpenoid-quinone biosynthesis	0.02202827	<i>tat, vkorc111, coq6</i>
Nucleotide excision repair	0.02460608	<i>rbx1, gtf2h3, pole2, ccnh</i> , etc.
Biotin metabolism	0.02567046	<i>hlcs</i>
RNA transport	0.03514058	<i>EIF2B3, nup188, gemin5, gemin6</i> , etc.
JAK-STAT signaling pathway	0.03880553	<i>prl, il6, lifr, csf2rb</i> , etc.
Caffeine metabolism	0.04809981	<i>uox</i>
Apoptosis—multiple species	0.04856226	<i>ngfr, cyc, tnfrsfla, diablo</i> , etc.

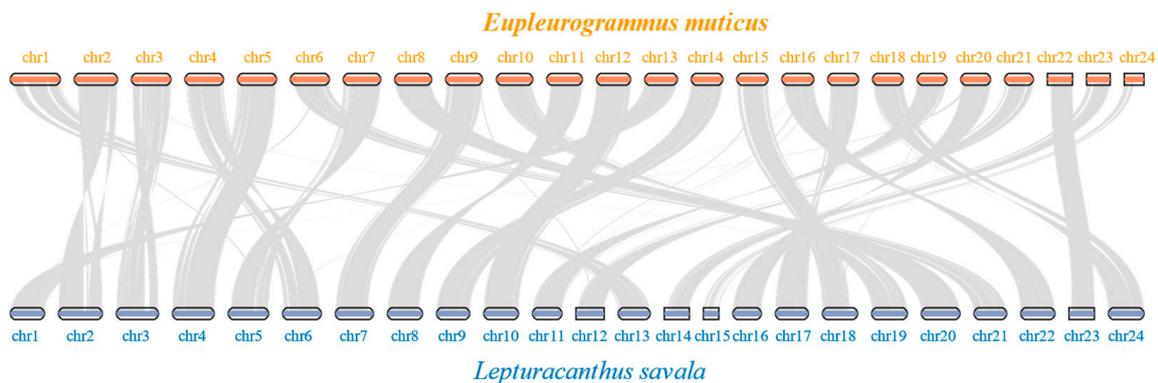


Figure 7. Collinearity analysis of *E. muticus* and *L. savala* based on coding sequences.

We further analyzed the functions of the positively selected genes and gene families by combining the biological characteristics, geographic distribution, and habitat environmental features of *E. muticus*, and confirmed that two gene families (*ccr3* and *ccr5*), 29 genes (*atg5*, *atg7*, *map1lc3b*, *map1lc3c*, *ids*, *lipa*, *gla*, *man2b1*, *glns*, *dnase2*, *ppt1*, *lamp2*, *il1rl1*, *il17a*, *c1s*, *c1qa*, *c9*, *mif*, *fanca*, *fance*, *fanci*, *faap100*, *eme1*, *bripl1*, *blm*, *slx1a*, *polh*, *palb2*, and *brca1*), and six pathways (autophagosome, lysosome, cytokine–cytokine receptor interaction, complement and coagulation cascades, virus protein interaction with cytokine and cytokine receptor, and Fanconi anemia pathway) play important roles in the whip-like tail formation, innate immune evolution, and DNA repair mechanisms of *E. muticus*.

4. Discussion

4.1. Characterization of the *E. muticus* Genome

In this study, we combined PacBio SMRT-Seq, Illumina Nova-Seq, and Hi-C technologies to obtain the first chromosome-level genome assembly of *E. muticus*. Since high coverage is one of the necessary conditions for sequencing error correction, genome coverage is a key indicator for measuring the efficiency of high-throughput genome sequencing technology [68]. Contig N50 value reflects the size and potential continuity of the genome

assembly and is an important parameter for determining the quality of genome assembly [69]. In this study, 222.36 Gb of raw sequencing data were obtained from different sequencing technologies, with Q20 and Q30 values above 95% and 90%, respectively, and a genome coverage of $274.4\times$. The contig N50 generated by the assembly was 25.07 Mb, and the proportion of genes with a complete comparison to BUSCO reached 97.1%. These high coverage and large contig N50 values showed the accuracy of the sequencing data and the high quality of the assembled *E. muticus* genome. The size of the final assembled genome was 709.27 Mb, with a GC content of 40.81%, a heterozygosity rate of 1.18%, and a repetitive sequence rate of 35.43%. These values were generally close to the results of the *E. muticus* genome survey (670 Mb, 41.68%, 1.26%, 35.33%) reported by Song et al. [70], but differed significantly from the results of the *L. savala* genome (790.02 Mb, 39.03%, 0.53%, 40.54%) as determined by Wu et al. [23]. The rate of repetitive sequences is a key factor influencing the genome size of species [71]. This rate was lower in the *E. muticus* genome than that of *L. savala*, which may be the main reason for the significantly lower genome size of the former. However, the heterozygosity rate of the *E. muticus* genome was 2.4 times higher than that of *L. savala*, indicating that *E. muticus* may have a relatively high level of genetic variation. Meng et al. [22] demonstrated that the randomly amplified polymorphic DNA (RAPD) polymorphism rate and genetic polymorphism were higher in *E. muticus* from the Yellow Sea than that of *Trichiurus lepturus*.

The number of chromosomes in the assembled genome of *E. muticus* was 24, consistent with that of *L. savala*, *T. albacares*, and *T. maccoyii* [23]. The ML tree results indicated that the phylogenetic relationships of these four species were consistent with their morphological classification [72]. The number of protein-coding genes annotated in the *E. muticus* genome was 21,949, which was significantly fewer than that in *L. savala* (23,625), *T. albacares* (BioProject: PRJEB47267) (24,623), and *T. maccoyii* (BioProject: PRJEB46021) (24,659). This may be related to their genome sizes (709.27 Mb vs. 790.02 Mb, 792.1 Mb, 782.4 Mb). Although a high level of collinearity was observed between the genomes of *E. muticus* and *L. savala*, there were varying degrees of gene deletions on chromosomes 2, 3, and 24 in the *E. muticus* genome compared to *L. savala*. This may be an important reason for the differences in biological characteristics between the two species.

4.2. Role of Autophagy-Related Genes in the Formation of Whip-like Tail in *E. muticus*

Autophagy refers to several processes by which cytoplasmic substances are introduced into the lysosome for degradation by autophagosomes [73,74]. Autophagy is involved in cell apoptosis and tissue remodeling during embryogenesis [75], responsible for the degradation of normal proteins to reorganize cells during animal metamorphosis and development [76]. It has been confirmed that autophagy is an important part of organ degeneration in arthropods and organ metamorphosis remodeling in most lepidopteran larvae [77,78]. A previous study reported that autophagosomes were present in the silk gland organ of the silkworm (*Bombyx mori*) during metamorphosis [79]. Moreover, the expression levels of both BmAtg8 and BmAtg12 proteins in the silk gland of the fifth instar larvae of *B. mori* were obviously up-regulated [80]. This indicates that autophagy plays a crucial role in the differentiation and degeneration of silk glands in the silkworm.

In the positive selection analyses, we screened several genes related to autophagosome formation (*atg5*, *atg7*, *map1lc3b*, and *map1lc3c*) and lysosome-related genes (*ids*, *lipa*, *gla*, *man2b1*, *glns*, *dnase2*, *ppt1*, and *lamp2*). These genes are crucial for the autophagy process in *E. muticus*, and may be involved in the autophagy-mediated degradation of certain organs. The protein encoded by *atg5* plays a core role in autophagy [81], and reducing or knocking out this protein could down-regulate or completely inhibit autophagy [82]. This interacts with the Atg12 protein to form an Atg12-Atg5 conjugate, which participates in the elongation of the isolation membranes during autophagosome formation [83]. As an essential element of autophagy, *atg7* encodes an E1-like activating enzyme involved in the two ubiquitin-like systems required for autophagy [84,85]. The absence of Atg7 protein could impair the degradation of the inner autophagosomal membrane (IAM) af-

ter the fusion of autophagosomes and lysosomes [86]. The *map1lc3b* and *map1lc3c* genes encode the Map1lc3 protein, an ortholog of the yeast autophagosome protein Atg8 [87], mediating autophagosome membrane formation as a ubiquitin-like modifier [88]. During the midgut remodeling in the larvae of *B. mori*, the expression levels of autophagy-related genes (*atg5*, *atg6*, and *atg8*) were significantly up-regulated [89]. Similarly, there was a significant increase in the expression levels of Atg5 and Lc3 proteins (orthologue of Atg8 protein) in the injury regeneration site of the caudal fin of zebrafish (*D. rerio*), indicating that *atg*-mediated autophagy is key for caudal fin regeneration in zebrafish [90]. In the *L. savala* genome, autophagy-related genes such as *atg3*, *atg4c*, and *atg12*, which are associated with the formation of its elongated whip-like tail, were also detected by positive selection analyses [23]. Meanwhile, the *ids*, *lipa*, *gla*, *man2b1*, *glns*, *dnase2*, and *ppt1* genes detected in this study encode various acid hydrolases that function in lysosomes. The *lamp2* gene encodes lysosomal-associated membrane protein 2, which protects lysosomal membranes from degradation by hydrolases [91]. These genes are essential for the lysosomal degradation of cellular substances and the maintenance of intracellular stability and autophagy [92]. Based on these findings, we speculate that the positive selection genes involved in autophagy may play a crucial role in the formation of the whip-like tail in *E. muticus*, especially the *atg5*, *atg7*, *map1lc3b*, and *map1lc3c* genes.

4.3. Evolution of Innate Immune System in *E. muticus*

Through positive selection and gene family expansion analyses, we identified several genes (*il1rl1*, *il17a*, *c1s*, *c1qa*, *c9*, *mif*, *ccr3*, and *ccr5*) related to innate immunity in the *E. muticus* genome. The *il1rl1* and *il17a* genes are associated with interleukins and are involved in the expression and regulation of inflammatory immune responses [93,94]. The *il1rl1* gene encodes the interleukin 1 receptor-like 1 protein (also known as St2) [95], which induces an immune response through its only ligand, interleukin-33 [96]. Conversely, *il17a* encodes the interleukin-17A protein [97] and plays an important role in the innate immune response, adaptive immunity, and immune defense against bacteria in teleost fishes [98]. The *c1s*, *c1qa*, and *c9* genes belong to the serum complement system-related genes, and *c1s* encodes a serine protease (C1s) [99], while *c1qa* encodes the A-chain polypeptide of serum complement subcomponent C1q [100]. Both C1s and C1q are components of the serum complement system C1, which defends against microbial infections and maintains immune homeostasis in organisms [101]. The complement C9 protein encoded by *c9* is critical for the innate immune response of teleost fishes against pathogen invasion [102]. The *mif* gene encodes the macrophage migration inhibitory factor [103], which enhances the resistance to bacterial invasion and promotes innate immune responses in golden pompano (*Trachinotus ovatus*) [104]. The *ccr3* and *ccr5* genes encode chemokine receptor family proteins that coordinate immune cell localization and function in the immune response of aquatic animals [105]. These genes were mainly enriched in pathways such as cytokine–cytokine receptor interaction, complement and coagulation cascades, and virus protein interaction with cytokine and cytokine receptors, indicating that they are crucial for the innate immunity of *E. muticus*. This may also indicate that the immune system of *E. muticus* has evolved. Similarly, several immune-related genes (e.g., *cfi*, *c1qa*, *vtn* genes, and the *il* gene family) detected in the *L. savala* genome were also significantly enriched in the complement and coagulation cascades and cytokine–cytokine receptor interaction pathways [23]. Moreover, the *c1qa* gene and the *il* gene families were identified in both *E. muticus* and *L. savala*, indicating that these two genes are conserved and are important for the immune system of trichiurids.

It has been demonstrated that pathogens, including bacteria, viruses, and parasites, could enhance evolutionary selective pressure on the host and promote the evolution of the host's immune system [106,107]. You et al. [108] observed that several immune-domain-containing genes, including the Toll-like receptor 13 (*tlr13*) gene family (the family of innate immune receptors), were commonly present in the genomes of four representative mudskippers [109]. These genes may provide a special immune defense against novel pathogens

encountered by mudskippers when living on land. In the genome of brown-marbled grouper (*E. fuscoguttatus*), Yang et al. [110] identified several expanded gene families (such as *nirc3*, *igl*, *trim25*, *fcr1*, and *trim35*) closely related to its antiviral infection, suggesting that the grouper has undergone adaptive evolution for disease resistance. In other words, the pathogen-rich habitats might have been the driving force behind its immune system's evolution. Another recent transcriptome analysis found that the peanut worm (*Sipunculus nudus*) has evolved different molecular mechanisms of immune responses to various habitat environments of the intertidal zone [111]. *E. muticus* has a wide range of habitats across the tropical, subtropical, and temperate zones of the Indo-West Pacific Oceans, rich in bacteria and diverse viruses. As previously reported, the East China Sea shelf sediments contained at least 13 bacteria phyla [112] and had abundant viral communities in their surface waters, including at least 1029 virus species [113]. Moreover, the bacterial communities in the deep seafloor sediments of the western Pacific warm pool contained more than five groups (e.g., α -/ β -Proteobacteria) [114], and a highly diversified bacterial community (seven dominant bacterial groups) was determined in the equatorial region of the East Indian Ocean and adjacent Bay of Bengal waters, with seven dominant bacterial phyla (e.g., Proteobacteria, Bacteroidetes, Actinobacteria) [115]. However, our study could not clarify the correlation between the immune system of *E. muticus* and its habitat environments due to the lack of relevant research data on its pathogens and diseases. Nonetheless, the innate immune evolution of *E. muticus* detected in this study provides a natural barrier against the invasion of several pathogens in its vast habitats, which may be important for maintaining the widespread distribution and population stability of *E. muticus*.

4.4. Important Role of DNA Repair-Related Genes in Maintaining Genome Stability of *E. muticus*

Genome integrity and stability are prerequisites for the survival and reproduction of species [116]. The genome of a species is constantly affected by internal and external factors during its life history, leading to DNA damage (e.g., DNA replication errors, ultraviolet radiation, environmental or reagent contamination) [117]. Therefore, repairing damaged DNA is important for maintaining the stability of a species' genome [118]. The Fanconi anemia pathway is essential for repairing damaged DNA and is primarily used to repair the interstrand DNA cross-linking damage [119]. Interstrand DNA cross-linking is a common type of DNA damage caused by ultraviolet light (UV), aldehydes produced by cellular metabolism, and exogenous DNA cross-linking agents, which block DNA replication, transcription, and recombination [118,120]. In our positive selection analyses, the genes (*fanca*, *fance*, *fanci*, *faap100*, *eme1*, *brip1*, *blm*, *slx1a*, *polh*, *palb2*, and *brca1*) related to DNA repair in *E. muticus* were significantly enriched in the Fanconi anemia pathway. The proteins encoded by *fanca*, *fance*, *fanci*, and *faap100* are components of the Fanconi anemia core complex, which are essential for DNA repair and maintenance of genomic stability [121,122]. *Fanca* protein activates interstrand DNA cross-link repair by monoubiquitination of *Fancd2* [123]. *Fance* protein forms a ternary complex with *Fancc* and *Fancd2* proteins and plays a role in the DNA damage response [124]. *Fanci* protein is critical for the repair of DNA double-strand breaks and interstrand DNA cross-links [125,126]. *Faap100* protein regulates *Fancd2* monoubiquitination and the stability of the Fanconi anemia core complex, which could significantly affect the DNA damage response associated with Fanconi anemia [127]. Furthermore, the proteins encoded by *eme1*, *brip1*, *blm*, *slx1a*, and *polh* are important enzymes involved in the DNA repair process. Among them, the *Eme1* and *Mus81* proteins form a DNA endonuclease that cleaves the branching DNA structures [128], and the lack of *Eme1* would lead to chromosomal instability in mouse clonal cells [129]. The *brip1* gene encodes a 5' to 3' DNA helicase (acting in DNA double-strand break repair) required to maintain chromosomal stability [130]. Like helicase encoded by *blm*, *Blm* RecQ participates in DNA replication and repair [131,132]. The *palb2* gene encodes the partner and localizer of *Brca2*, which plays a key role in the homologous recombination repair by localizing to DNA damage sites [133]. As a functional unit component of the homologous recombination and DNA damage repair [134], the *Brca1* protein encoded by *brca1* participates in DNA

damage repair and transcriptional regulation [135,136]. Thus, it is evident that the proteins encoded by these genes are important components of the DNA repair system of *E. muticus*.

Similarly, genes related to DNA repair (*polm*, *prkdc*, *bard1*, *brca1*, *nbn*, *xrcc2*, *eme2*, and *faap100*) were also screened in the positive selection analyses of the *L. savala* genome [23]. These genes have an essential contribution to the recombination of homologous chromosomes and the maintenance of genomic stability in the *L. savala* genome [23]. Based on a comparative analysis of the genomes of chondrichthyan and teleost fishes, Marra et al. [137] found that most positively selected genes associated with DNA damage response, DNA repair, translesion DNA synthesis, and ubiquitination widely exist in chondrichthyan fishes (e.g., white shark (*Carcharodon carcharias*), whale shark (*R. typus*), and elephant shark (*Callorhinchus milii*)), and are important in maintaining the genomic stability of sharks. Further positive selection analyses showed that some core histone genes were involved in the DNA damage response and associated histone epigenetic modifications in white shark, and that genes (*fgg*, *extl2*, and *krt18*) and terms (e.g., angiogenesis, VEGFA-VEGFR2 signaling network, epidermal growth factor receptor) related to the stronger wound healing in white shark were also identified [137]. This revealed the pivotal role of the white shark's cancer-fighting, long lifespan, and superior wound-healing ability in maintaining its genome stability and conservation genes, which perpetuates the long-term existence of the species [137]. In a genomic analysis of the Yap hadal snailfish, 34 positively selected genes (e.g., *rad52*, *rad9a*, *ercc1*, *exo1*, *pms1*, and *polk*) were identified to be significantly enriched in the DNA repair pathway, and its copies of *rad51* and *xrcc2* were higher than those of other teleost fishes [138]. Two proteins encoded by *rad51* and *xrcc2* play key roles in repairing DNA double-strand breaks [139] and DNA damage [140], respectively. Recently, 22 significantly co-expanded gene families of the two deep-sea anemones (*Alvinactis idseensis* and *Paraphelliactis xishaensis*) were found to be associated with DNA repair and cell membrane [141]. High hydrostatic pressure can cause DNA breaks and damage and affect cell membrane fluidity, protein stability, and the cytoskeleton [141]. Thus, the identification of these genes signifies that the Yap hadal snailfish and deep-sea anemones have a greater ability to repair DNA, which is critical for their adaptation to the high hydrostatic pressure in their living environment [138]. Based on these analyses, we infer that the genes related to DNA repair in *E. muticus* may have important contributions in maintaining its genome stability and survival. Although no studies have been reported on the DNA repair mechanisms in trichiurids, we screened multiple genes related to DNA repair in the genomes of both *E. muticus* and *L. savala*, indicating that the trichiurids may have well-developed DNA repair mechanisms. However, the importance of these genes in these species remains to be explored.

5. Conclusions

In this study, a high-quality *E. muticus* genome was assembled at the chromosomal level for the first time via PacBio sequencing and Hi-C technology. The assembled genome size was 709.27 Mb, with a contig N50 value of 25.07 Mb, and contained 21,949 protein-coding genes and 24 chromosomes. Positive selection analyses revealed that the genes (*atg5*, *atg7*, *map1lc3b*, and *map1lc3c*) related to autophagosome formation may be the key factors associated with the whip-like tail of *E. muticus*. The rapid evolution of the innate immune-related genes (*il1rl1*, *il17a*, *c1s*, *c1qa*, *c9*, *mif*, *ccr3*, and *ccr5*) was detected in *E. muticus*. This innate immune evolution provides a natural barrier for *E. muticus* against the invasion of various pathogens in its vast habitats. Moreover, the genes (*fanca*, *fance*, *fanci*, *faap100*, *eme1*, *brip1*, *blm*, *slx1a*, *polh*, *palb2*, and *brca1*) related to DNA repair mechanisms were also identified in *E. muticus*. These genes are important in maintaining the stability of the *E. muticus* genome and ensuring the survival and reproduction of the species. Thus, our study provides important basic data for exploring the genetic and evolutionary mechanisms of *E. muticus* at the genomic level and is an invaluable reference for the genomic studies of other trichiurids.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ani14030434/s1>. Supplementary Material S1: Figure S1: The K-mer analysis of *E. muticus* genome; Figure S2: The correlation graph between the GC content and average depth distribution based on Illumina clean reads (A) and PicBio HiFi reads (B); Figure S3: The number of reads in the valid match result; Figure S4: Statistical graphs of the gene set obtained via gene structure prediction (comparison graph of gene elements with closely related species); Figure S5: The phylogenetic tree of 20 fish species; Figure S6: Whole-genome sequence-based collinearity analysis of *E. muticus* and *L. savala*; Table S1: Blast results of the NCBI nucleotide database; Table S2: Sequence statistics of genome assembly results; Table S3: The results of the BUSCO evaluation; Table S4: The statistics of matching results; Table S5: Filtering statistics of the successfully paired matches; Table S6: The results of the Hi-C-assisted assembly; Table S7: The interspersed repeat classification results; Table S8: Annotation and detailed data of the non-coding RNA; Table S9: Summary of functional annotations of the predicted genes. Supplementary Material S2: Results of enrichment analysis of unique gene families in *E. muticus*. Supplementary Material S3: Results of enrichment analysis of the contraction and expansion gene families in *E. muticus*. Supplementary Material S4: Results of the positive selection analysis of *E. muticus*.

Author Contributions: Conceptualization, R.-X.W. and F.-Y.H.; methodology, F.-Y.H., R.-X.W. and B.-B.M.; software, F.-Y.H., R.-X.W. and B.-B.M.; validation, F.-Y.H. and R.-X.W.; formal analysis, F.-Y.H., R.-X.W. and B.-B.M.; investigation, Q.-H.W. and Z.-B.L.; resources, R.-X.W., Q.-H.W. and Z.-B.L.; data curation, F.-Y.H. and R.-X.W.; writing—original draft preparation, F.-Y.H. and R.-X.W.; writing—review and editing, F.-Y.H., R.-X.W. and S.-F.N.; visualization, F.-Y.H. and R.-X.W.; supervision, R.-X.W. and S.-F.N.; project administration, R.-X.W.; funding acquisition, R.-X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant no. 31372532).

Institutional Review Board Statement: The experimental animal protocols in the present study were reviewed and approved by the Animal Experimental Ethics Committee of Guangdong Ocean University, China (approval number: 0901-2021, approval date: 5 September 2021). Experiment procedures were performed in accordance with the Provisions and Regulations for the National Experimental Animal Management Regulations (China, July 2013) and the Experimental Animal Policies and Regulations of Guangdong Province (China, October 2010).

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome assembly data of *Eupleurogrammus muticus* were deposited at NCBI under BioProject ID: PRJNA1040966 (Submitted), BioSample accession: SAMN38271036 (Submitted). The raw read sequence accession numbers: SRR26857548, SRR26857549, SRR26857550, SRR26857546, SRR26857547.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Froese, R.; Pauly, D. FishBase. World Wide Web Electronic Publication. Version (10/2023). Available online: <http://www.fishbase.org> (accessed on 3 November 2023).
2. Shao, K.T. The Fish Database of Taiwan. Available online: <http://fishdb.sinica.edu.tw> (accessed on 26 September 2023).
3. Nakamura, I.; Parin, N.V. An annotated and illustrated catalogue of the Snake Mackerels, Snoeks, Escolars, Gemfishes, Sackfishes, Domine, Oilfish, Cutlassfishes, Scabbardfishes, Hairtails and Frostfishes known to date. *FAO Fish. Synopsis*. **1993**, *125*, 61–107.
4. He, X.; Luo, Z.; Zhao, C.; Huang, L.; Yan, Y.; Kang, B. Species Composition, Growth, and Trophic Traits of Hairtail (Trichiuridae), the Most Productive Fish in Chinese Marine Fishery. *Animals* **2022**, *12*, 3078. [[CrossRef](#)]
5. Chen, G.B.; Li, Y.Z.; Zhao, X.Y.; Chen, Y.Z.; Jin, X.S. Acoustic assessment of five groups commercial fish in South China Sea. *Acta Oceanol. Sin.* **2006**, *28*, 128–134.
6. Luo, B.Z. Life history patterns and geographical variation of ecological parameters for marine fishes in the coastal waters of China. *Oceanol. Limnol. Sin.* **1992**, *23*, 63–73.
7. Wang, K.L.; Zhang, P.J.; Liu, L.Y.; You, F.; Xu, C.; Wang, J.F. Research on the population biochemical genetic structure and identification of ribbonfish in the coastal waters of China. *Acta Oceanol. Sin.* **1994**, *16*, 93–104.
8. Zhang, Q.Y.; Hong, W.S.; Chen, S.X. Stock changes and resource protection of the large yellow croaker (*Larimichthys crocea*) and ribbon fish (*Trichiurus japonicus*) in coastal waters of China. *J. Appl. Oceanogr.* **2017**, *36*, 438–445. [[CrossRef](#)]

9. Liu, Y.; Cheng, J.H.; Jia, S.G. Characteristics of submarine water temperature distribution of *Trichiurus haumela* in the East China Sea and Southern Yellow Sea with the improvement of the analysis methods. *J. Fish. China* **2021**, *45*, 871–886. [[CrossRef](#)]
10. Memon, K.H.; Qun, L.; Kalhor, M.A.; Chang, M.S.; Baochao, L.; Memon, A.M.; Hyder, S.; Tabassum, S. Growth and Mortality Parameters of Hairtail *Lepturacanthus savala* from Pakistan Waters. *Pak. J. Zool.* **2016**, *48*, 829–837.
11. Hsu, K.C.; Yi, M.R.; Gu, S.; He, X.-B.; Luo, Z.-S.; Kang, B.; Lin, H.-D.; Yan, Y.-R. Composition, Demographic History, and Population Structures of *Trichiurus*. *Front. Mar. Sci.* **2022**, *9*, 875042. [[CrossRef](#)]
12. James, P. *MBAI Memoir No. 1: The Ribbon-Fishes of the Family Trichiuridae of India*; Western Printers & Printers, Bombay-13: Madras State, India, 1967; pp. 87–122.
13. Liu, J.; Wu, R.X.; Kang, B.; Ma, L. *Fishes of Beibu Gulf*; Science Press: Beijing, China, 2016; p. 339.
14. Pauly, D.; Christensen, V.V.; Dalsgaard, J.; Froese, R.; Torres, F. Fishing down Marine Food Webs. *Science* **1998**, *279*, 860–863. [[CrossRef](#)] [[PubMed](#)]
15. Miqueleiz, I.; Miranda, R.; Ariño, A.H.; Ojea, E. Conservation-Status Gaps for Marine Top-Fished Commercial Species. *Fishes* **2022**, *7*, 2. [[CrossRef](#)]
16. Liu, S.; Zu, D.; Liu, Q.; Dai, F.; Ma, Q.; Zhuang, Z. Isolation and characterization of polymorphic microsatellite markers for *Eupleurogrammus muticus*. *Conserv. Genet. Resour.* **2015**, *7*, 487–488. [[CrossRef](#)]
17. Xu, B.; Jin, X. Variations in fish community structure during winter in the southern Yellow Sea over the period 1985–2002. *Fish. Res.* **2005**, *71*, 79–91. [[CrossRef](#)]
18. Tucker, D.W. Studies on the trichiuroid fishes—3. A preliminary revision of the family Trichiuridae. *Bull. Br. Mus. (Nat. Hist.) Zool.* **1956**, *4*, 112–128. [[CrossRef](#)]
19. Rizvi, A.F.; Chakraborty, S.K.; Deshmukh, V.D. Stock assessment of small head hair tail *Eupleurogrammus muticus* (Gray) (Pisces/Trichiuridae) from Mumbai coast. *Indian J. Mar. Sci.* **2003**, *32*, 85–88.
20. Burhanuddin, A.I.; Iwatsuki, Y. Comparative of Meristic and Morphometric Characters between Two Smallhead Hairtail Fishes *Eupleurogrammus muticus* (Gray, 1831) and *E. Glossodon* (Bleeker, 1860) (Perciformes: Trichiuridae). *Biota* **2006**, *11*, 142–145.
21. Eighani, M.; Daliri, M.; Paighambari, S.Y.; Alizadeh, E. Length-weight relationship and GSI index of smallhead hairtail, *Eupleurogrammus muticus* (Gray, 1831), northern Persian Gulf, Hormozgan coastal waters. *J. Appl. Ichthyol.* **2014**, *30*, 257–258. [[CrossRef](#)]
22. Meng, Z.N.; Zhuang, Z.M.; Jin, X.S.; Tang, Q.S.; Su, Y.Q. Analysis of RAPD and mitochondrial 16S rRNA gene sequences from *Trichiurus lepturus* and *Eupleurogrammus muticus* in the Yellow Sea. *Prog. Nat. Sci.* **2003**, *11*, 1170–1176. [[CrossRef](#)]
23. Wu, R.X.; Miao, B.B.; Han, F.Y.; Niu, S.F.; Liang, Y.S.; Liang, Z.B.; Wang, Q.H. Chromosome-Level Genome Assembly Provides Insights into the Evolution of the Special Morphology and Behaviour of *Lepturacanthus savala*. *Genes* **2023**, *14*, 1268. [[CrossRef](#)]
24. Sambrook, J.; Russell, D.W. The Inoue Method for Preparation and Transformation of Competent *E. coli*: “Ultra-Competent” Cells. *CSH Protoc.* **2006**, 2006, pdb.prot3944. [[CrossRef](#)]
25. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. *fastp*: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)] [[PubMed](#)]
26. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)] [[PubMed](#)]
27. Liu, B.; Shi, Y.; Yuan, J.; Hu, X.; Zhang, H.; Li, N.; Li, Z.; Chen, Y.; Mu, D.; Fan, W. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv* **2013**, arXiv:1308.2012.
28. Gordon, S.P.; Tseng, E.; Salamov, A.; Zhang, J.; Meng, X.; Zhao, Z.; Kang, D.; Underwood, J.; Grigoriev, I.V.; Figueroa, M.; et al. Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. *PLoS ONE* **2015**, *10*, e0132628. [[CrossRef](#)] [[PubMed](#)]
29. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [[CrossRef](#)]
30. Roach, M.J.; Schmidt, S.A.; Borneman, A.R. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinform.* **2018**, *19*, 460. [[CrossRef](#)]
31. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
32. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
34. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **2006**, *34*, W435–W439. [[CrossRef](#)]
35. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **2013**, *41*, e121. [[CrossRef](#)]
36. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)] [[PubMed](#)]

37. Dudchenko, O.; Batra, S.S.; Omer, A.D.; Nyquist, S.K.; Hoeger, M.; Durand, N.C.; Shamim, M.S.; Machol, I.; Lander, E.S.; Aiden, A.P.; et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **2017**, *356*, 92–95. [[CrossRef](#)] [[PubMed](#)]
38. Durand, N.C.; Shamim, M.S.; Machol, I.; Rao, S.S.P.; Huntley, M.H.; Lander, E.S.; Aiden, E.L. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **2016**, *3*, 95–98. [[CrossRef](#)] [[PubMed](#)]
39. Durand, N.C.; Robinson, J.T.; Shamim, M.S.; Machol, I.; Mesirov, J.P.; Lander, E.S.; Aiden, E.L. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **2016**, *3*, 99–101. [[CrossRef](#)] [[PubMed](#)]
40. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
41. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [[CrossRef](#)] [[PubMed](#)]
42. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **2009**, *25*, 4.10.1–4.10.14. [[CrossRef](#)]
43. Xu, Z.; Wang, H. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **2007**, *35*, W265–W268. [[CrossRef](#)]
44. Price, A.L.; Jones, N.C.; Pevzner, P.A. De novo identification of repeat families in large genomes. *Bioinformatics* **2005**, *21*, i351–i358. [[CrossRef](#)]
45. Birney, E.; Clamp, M.; Durbin, R. GeneWise and Genomewise. *Genome Res.* **2004**, *14*, 988–995. [[CrossRef](#)]
46. Burge, C.; Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **1997**, *268*, 78–94. [[CrossRef](#)]
47. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. [[CrossRef](#)]
48. Holt, C.; Yandell, M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **2011**, *12*, 491. [[CrossRef](#)]
49. Parra, G.; Bradnam, K.; Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **2007**, *23*, 1061–1067. [[CrossRef](#)]
50. Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28*, 45–48. [[CrossRef](#)]
51. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]
52. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
53. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [[CrossRef](#)] [[PubMed](#)]
54. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)] [[PubMed](#)]
55. Griffiths-Jones, S.; Moxon, S.; Marshall, M.; Khanna, A.; Eddy, S.R.; Bateman, A. Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **2005**, *33*, D121–D124. [[CrossRef](#)]
56. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* **2003**, *13*, 2178–2189. [[CrossRef](#)]
57. Nakamura, T.; Yamada, K.D.; Tomii, K.; Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **2018**, *34*, 2490–2492. [[CrossRef](#)] [[PubMed](#)]
58. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
59. Kumar, S.; Stecher, G.; Suleski, M.; Hedges, S.B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **2017**, *34*, 1812–1819. [[CrossRef](#)] [[PubMed](#)]
60. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
61. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **1997**, *13*, 555–556. [[CrossRef](#)]
62. Mendes, F.K.; Vanderpool, D.; Fulton, B.; Hahn, M.W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **2021**, *36*, 5516–5518. [[CrossRef](#)]
63. Yang, Z.; Nielsen, R. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites along Specific Lineages. *Mol. Biol. Evol.* **2002**, *19*, 908–917. [[CrossRef](#)]
64. Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes Empirical Bayes Inference of Amino Acid Sites under Positive Selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [[CrossRef](#)] [[PubMed](#)]
65. Li, S.; Zhao, G.; Han, H.; Li, Y.; Li, J.; Wang, J.; Cao, G.; Li, X. Genome collinearity analysis illuminates the evolution of donkey chromosome 1 and horse chromosome 5 in perissodactyls: A comparative study. *BMC Genom.* **2021**, *22*, 665. [[CrossRef](#)] [[PubMed](#)]
66. Tang, H.; Bowers, J.E.; Wang, X.; Ming, R.; Alam, M.; Paterson, A.H. Synteny and Collinearity in Plant Genomes. *Science* **2008**, *320*, 486–488. [[CrossRef](#)]
67. Delcher, A.; Salzberg, S.; Phillippy, A. Using MUMmer to Identify Similar Regions in Large Sequence Sets. *Curr. Protoc. Bioinform.* **2003**, *00*, 10.3.1–10.3.18. [[CrossRef](#)] [[PubMed](#)]

68. Sims, D.; Sudbery, I.; Illott, N.E.; Heger, A.; Ponting, C.P. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat. Rev. Genet.* **2014**, *15*, 121–132. [[CrossRef](#)] [[PubMed](#)]
69. Earl, D.; Bradnam, K.; St John, J.; Darling, A.; Lin, D.; Fass, J.; Yu, H.O.K.; Buffalo, V.; Zerbino, D.R.; Diekhans, M. Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Res.* **2011**, *21*, 2224–2241. [[CrossRef](#)]
70. Song, N.; Zhao, X.; Cai, C.; Gao, T. Profile of the genomic characteristics and comparative studies of five Trichiuridae species by genome survey sequencing. *Front. Mar. Sci.* **2022**, *9*, 962307. [[CrossRef](#)]
71. Kazazian, H.H.J. Mobile Elements: Drivers of Genome Evolution. *Science* **2004**, *303*, 1626–1632. [[CrossRef](#)]
72. Nelson, J.S.; Grande, T.C.; Wilson, M.V.H. *Fishes of the World*, 5th ed.; John Wiley & Sons: New York, NY, USA, 2016; pp. 13–526. [[CrossRef](#)]
73. He, C.; Klionsky, D.J. Regulation Mechanisms and Signaling Pathways of Autophagy. *Annu. Rev. Genet.* **2009**, *43*, 67–93. [[CrossRef](#)]
74. Ichimiya, T.; Yamakawa, T.; Hirano, T.; Yokoyama, Y.; Hayashi, Y.; Hirayama, D.; Wagatsuma, K.; Itoi, T.; Nakase, H. Autophagy and Autophagy-Related Diseases: A Review. *Int. J. Mol. Sci.* **2020**, *21*, 8974. [[CrossRef](#)]
75. Cecconi, F.; Levine, B. The Role of Autophagy in Mammalian Development: Cell Makeover Rather than Cell Death. *Dev. Cell* **2008**, *15*, 344–357. [[CrossRef](#)]
76. Di Bartolomeo, S.; Nazio, F.; Cecconi, F. The Role of Autophagy During Development in Higher Eukaryotes. *Traffic* **2010**, *11*, 1280–1289. [[CrossRef](#)]
77. Malagoli, D.; Abdalla, F.C.; Cao, Y.; Feng, Q.; Fujisaki, K.; Gregorc, A.; Matsuo, T.; Nezis, I.P.; Papassideri, I.S.; Sass, M.; et al. Autophagy and its physiological relevance in arthropods: Current knowledge and perspectives. *Autophagy* **2010**, *6*, 575–588. [[CrossRef](#)]
78. Romanelli, D.; Casati, B.; Franzetti, E.; Tettamanti, G. A Molecular View of Autophagy in Lepidoptera. *Biomed. Res. Int.* **2014**, *2014*, 902315. [[CrossRef](#)] [[PubMed](#)]
79. Matsuura, S.; Tashiro, Y. Cup-shaped Mitochondria in the Posterior Silk Gland of *Bombyx mori* in the Prepupal Stadium. *Cell Struct. Funct.* **1976**, *1*, 137–145. [[CrossRef](#)]
80. Zhang, X.; Hu, Z.Y.; Li, W.F.; Li, Q.R.; Deng, X.J.; Yang, W.Y.; Cao, Y.; Zhou, C.Z. Systematic cloning and analysis of autophagy-related genes from the silkworm *Bombyx mori*. *BMC Mol. Biol.* **2009**, *10*, 50. [[CrossRef](#)]
81. Changotra, H.; Kaur, S.; Yadav, S.S.; Gupta, G.L.; Parkash, J.; Duseja, A. ATG5: A central autophagy regulator implicated in various human diseases. *Cell Biochem. Funct.* **2022**, *40*, 650–667. [[CrossRef](#)] [[PubMed](#)]
82. Ye, X.; Zhou, X.J.; Zhang, H. Exploring the Role of Autophagy-Related Gene 5 (ATG5) Yields Important Insights Into Autophagy in Autoimmune/Autoinflammatory Diseases. *Front. Immunol.* **2018**, *9*, 2334. [[CrossRef](#)]
83. Mizushima, N.; Yamamoto, A.; Hatano, M.; Kobayashi, Y.; Kabeya, Y.; Suzuki, K.; Tokuhisa, T.; Ohsumi, Y.; Yoshimori, T. Dissection of Autophagosome Formation Using Apg5-Deficient Mouse Embryonic Stem Cells. *J. Cell Biol.* **2001**, *152*, 657–668. [[CrossRef](#)]
84. You, Z.; Xu, Y.; Wan, W.; Zhou, L.; Li, J.; Zhou, T.; Shi, Y.; Liu, W. TP53INP2 contributes to autophagosome formation by promoting LC3-ATG7 interaction. *Autophagy* **2019**, *15*, 1309–1321. [[CrossRef](#)]
85. Collier, J.J.; Guissart, C.; Olahova, M.; Sasorith, S.; Piron-Prunier, F.; Suomi, F.; Zhang, D.; Martinez-Lopez, N.; Leboucq, N.; Bahr, A.; et al. Developmental Consequences of Defective ATG7-Mediated Autophagy in Humans. *N. Engl. J. Med.* **2021**, *384*, 2406–2417. [[CrossRef](#)]
86. Tsuboyama, K.; Koyama-Honda, I.; Sakamaki, Y.; Koike, M.; Morishita, H.; Mizushima, N. The ATG conjugation systems are important for degradation of the inner autophagosomal membrane. *Science* **2016**, *354*, 1036–1041. [[CrossRef](#)]
87. He, H.; Dang, Y.; Dai, F.; Guo, Z.; Wu, J.; She, X.; Pei, Y.; Chen, Y.; Ling, W.; Wu, C.; et al. Post-translational Modifications of Three Members of the Human MAP1LC3 Family and Detection of a Novel Type of Modification for MAP1LC3B. *J. Biol. Chem.* **2003**, *278*, 29278–29287. [[CrossRef](#)] [[PubMed](#)]
88. Weidberg, H.; Shvets, E.; Shpilka, T.; Shimron, F.; Elazar, Z. LC3 and GATE-16/GABARAP subfamilies are both essential yet act differently in autophagosome biogenesis. *EMBO J.* **2010**, *29*, 1792–1802. [[CrossRef](#)] [[PubMed](#)]
89. Franzetti, E.; Huang, Z.J.; Shi, Y.X.; Xie, K.; Deng, X.J.; Li, J.P.; Li, Q.R.; Yang, W.Y.; Zeng, W.N.; Casartelli, M.; et al. Autophagy precedes apoptosis during the remodeling of silkworm larval midgut. *Apoptosis* **2012**, *17*, 305–324. [[CrossRef](#)] [[PubMed](#)]
90. Varga, M.; Sass, M.; Papp, D.; Takacs-Vellai, K.; Kobolak, J.; Dinnyes, A.; Klionsky, D.J.; Vellai, T. Autophagy is required for zebrafish caudal fin regeneration. *Cell Death Differ.* **2014**, *21*, 547–556. [[CrossRef](#)]
91. Eskelinen, E.L.; Schmidt, C.K.; Neu, S.; Willenborg, M.; Fuertes, G.; Salvador, N.; Tanaka, Y.; Lüllmann-Rauch, R.; Hartmann, D.; Heeren, J.; et al. Disturbed Cholesterol Traffic but Normal Proteolytic Function in LAMP-1/LAMP-2 Double-deficient Fibroblasts. *Mol. Biol. Cell* **2004**, *15*, 3132–3145. [[CrossRef](#)]
92. Yim, W.W.; Mizushima, N. Lysosome biology in autophagy. *Cell Discov.* **2020**, *6*, 6. [[CrossRef](#)]
93. Subramaniam, S.; Stansberg, C.; Cunningham, C. The interleukin 1 receptor family. *Dev. Comp. Immunol.* **2004**, *28*, 415–428. [[CrossRef](#)] [[PubMed](#)]
94. Adamopoulos, I.E.; Pflanz, S. The emerging role of Interleukin 27 in inflammatory arthritis and bone destruction. *Cytokine Growth Factor Rev.* **2013**, *24*, 115–121. [[CrossRef](#)]
95. Pascual-Figal, D.A.; Januzzi, J.L. The Biology of ST2: The International ST2 Consensus Panel. *Am. J. Cardiol.* **2015**, *115*, 3B–7B. [[CrossRef](#)]

96. Karaesmen, E.; Hahn, T.; Dile, A.J.; Rizvi, A.A.; Wang, J.; Wang, T.; Haagenson, M.D.; Preus, L.; Zhu, Q.; Liu, Q.; et al. Multiple functional variants in the *IL1RL1* region are pretransplant markers for risk of GVHD and infection deaths. *Blood Adv.* **2019**, *3*, 2512–2524. [[CrossRef](#)]
97. McGeachy, M.J.; Cua, D.J. Th17 Cell Differentiation: The Long and Winding Road. *Immunity* **2008**, *28*, 445–453. [[CrossRef](#)]
98. Wang, X.; Jiang, X.; Zhu, L.; Yuan, G.; Li, L.; Pei, C.; Kong, X. Molecular characterizations, immune modulation, and antibacterial activity of interleukin-17A/F1a and interleukin-17A/F1b in common carp *Cyprinus carpio*. *Fish Shellfish Immunol.* **2022**, *127*, 561–571. [[CrossRef](#)]
99. Gal, P.; Ambrus, G.; Zavodszky, P. C1s, the Protease Messenger of C1. *Immunobiology* **2002**, *205*, 383–394. [[CrossRef](#)]
100. Kishore, U.; Reid, K.B. C1q: Structure, function, and receptors. *Immunopharmacology* **2000**, *49*, 159–170. [[CrossRef](#)]
101. Gaboriaud, C.; Thielens, N.M.; Gregory, L.A.; Rossi, V.; Fontecilla-Camps, J.C.; Arlaud, G.J. Structure and activation of the C1 complex of complement: Unraveling the puzzle. *Trends Immunol.* **2004**, *25*, 368–373. [[CrossRef](#)] [[PubMed](#)]
102. Morgan, B.P. The Human Complement System in Health and Disease. *Ann. Rheum. Dis.* **1998**, *57*, 581. [[CrossRef](#)]
103. Sumaiya, K.; Langford, D.; Natarajaseenivasan, K.; Shanmughapriya, S. Macrophage migration inhibitory factor (MIF): A multifaceted cytokine regulated by genetic and physiological strategies. *Pharmacol. Ther.* **2022**, *233*, 108024. [[CrossRef](#)]
104. Zhang, Z.; Hu, X.; Diao, Q.; Zhang, P.; Wu, Y.; Cao, Z.; Zhou, Y.; Liu, C.; Sun, Y. Macrophage migration inhibitory factor (MIF) of golden pompano (*Trachinotus ovatus*) is involved in the antibacterial immune response. *Dev. Comp. Immunol.* **2022**, *133*, 104445. [[CrossRef](#)] [[PubMed](#)]
105. Alejo, A.; Tafalla, C. Chemokines in teleost fish species. *Dev. Comp. Immunol.* **2011**, *35*, 1215–1222. [[CrossRef](#)] [[PubMed](#)]
106. Fumagalli, M.; Sironi, M.; Pozzoli, U.; Ferrer-Admetlla, A.; Pattini, L.; Nielsen, R. Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLOS Genet.* **2011**, *7*, e1002355. [[CrossRef](#)]
107. Ebert, D.; Fields, P.D. Host–parasite co-evolution and its genomic signature. *Nat. Rev. Genet.* **2020**, *21*, 754–768. [[CrossRef](#)]
108. You, X.; Bian, C.; Zan, Q.; Xu, X.; Liu, X.; Chen, J.; Wang, J.; Qiu, Y.; Li, W.; Zhang, X.; et al. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nat. Commun.* **2014**, *5*, 5594. [[CrossRef](#)]
109. Oldenburg, M.; Krüger, A.; Ferstl, R.; Kaufmann, A.; Nees, G.; Sigmund, A.; Bathke, B.; Lauterbach, H.; Suter, M.; Dreher, S.; et al. TLR13 Recognizes Bacterial 23S rRNA Devoid of Erythromycin Resistance–Forming Modification. *Science* **2012**, *337*, 1111–1115. [[CrossRef](#)]
110. Yang, Y.; Wang, T.; Chen, J.; Wu, L.; Wu, X.; Zhang, W.; Luo, J.; Xia, J.; Meng, Z.; Liu, X. Whole-genome sequencing of brown-marbled grouper (*Epinephelus fuscoguttatus*) provides insights into adaptive evolution and growth differences. *Mol. Ecol. Resour.* **2022**, *22*, 711–723. [[CrossRef](#)] [[PubMed](#)]
111. Li, J.; Wen, J.; Hu, R.; Pei, S.; Li, T.; Shan, B.; Huang, H.; Zhu, C. Transcriptome Responses to Different Environments in Intertidal Zones in the Peanut Worm *Sipunculus nudus*. *Biology* **2023**, *12*, 1182. [[CrossRef](#)] [[PubMed](#)]
112. Liu, M.H.; Wang, J.X.; Yu, K.C.; Jiang, R.; Liu, X.H.; Wang, S.B.; Liu, X.Z. Community Structure and geographical distribution of bacterial on surface layer sediments in the East China Sea. *Oceanol. Limnol. Sin.* **2015**, *46*, 1119–1131.
113. Wu, S.; Zhou, L.; Zhou, Y.; Wang, H.; Xiao, J.; Yan, S.; Wang, Y. Diverse and unique viruses discovered in the surface water of the East China Sea. *BMC Genom.* **2020**, *21*, 441. [[CrossRef](#)]
114. Jing, Z.; Zeng, R. Bacterial community in deep seafloor sediments from the western Pacific “warm pool”. *Sci. China Ser. D-Earth Sci.* **2008**, *51*, 618–624. [[CrossRef](#)]
115. Wang, J.; Kan, J.; Borecki, L.; Zhang, X.; Wang, D.; Sun, J. A snapshot on spatial and vertical distribution of bacterial communities in the eastern Indian Ocean. *Acta Oceanol. Sin.* **2016**, *35*, 85–93. [[CrossRef](#)]
116. Feng, J.X.; Riddle, N.C. Epigenetics and genome stability. *Mamm. Genome* **2020**, *31*, 181–195. [[CrossRef](#)]
117. Giglia-Mari, G.; Zotter, A.; Vermeulen, W. DNA damage response. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a000745. [[CrossRef](#)]
118. Sancar, A.; Lindsey-Boltz, L.A.; Unsal-Kacmaz, K.; Linn, S. Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. *Annu. Rev. Biochem.* **2004**, *73*, 39–85. [[CrossRef](#)] [[PubMed](#)]
119. Kottmann, M.C.; Smogorzewska, A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **2013**, *493*, 356–363. [[CrossRef](#)] [[PubMed](#)]
120. Wu, R.A.; Semlow, D.R.; Kamimae-Lanning, A.N.; Kochenova, O.V.; Chistol, G.; Hodskinson, M.R.; Amunugama, R.; Sparks, J.L.; Wang, M.; Deng, L.; et al. TRAP1 is a master regulator of DNA interstrand crosslink repair. *Nature* **2019**, *567*, 267–272. [[CrossRef](#)]
121. Liang, F.; Miller, A.S.; Tang, C.; Maranon, D.; Williamson, E.A.; Hromas, R.; Wiese, C.; Zhao, W.; Sung, P.; Kupfer, G.M. The DNA-binding activity of USP1-associated factor 1 is required for efficient RAD51-mediated homologous DNA pairing and homology-directed DNA repair. *J. Biol. Chem.* **2020**, *295*, 8186–8194. [[CrossRef](#)]
122. Huang, Y.; Leung, J.W.; Lowery, M.; Matsushita, N.; Wang, Y.; Shen, X.; Huong, D.; Takata, M.; Chen, J.; Li, L. Modularized Functions of the Fanconi Anemia Core Complex. *Cell Rep.* **2014**, *7*, 1849–1857. [[CrossRef](#)]
123. Benitez, A.; Liu, W.; Palovcak, A.; Wang, G.; Moon, J.; An, K.; Kim, A.; Zheng, K.; Zhang, Y.; Bai, F.; et al. FANCA Promotes DNA Double-Strand Break Repair by Catalyzing Single-Strand Annealing and Strand Exchange. *Mol. Cell* **2018**, *71*, 621–628.e4. [[CrossRef](#)] [[PubMed](#)]
124. Gordon, S.M.; Alon, N.; Buchwald, M. FANCC, FANCE, and FANCD2 Form a Ternary Complex Essential to the Integrity of the Fanconi Anemia DNA Damage Response Pathway. *J. Biol. Chem.* **2005**, *280*, 36118–36125. [[CrossRef](#)]

125. Smogorzewska, A.; Matsuo, S.; Vinciguerra, P.; McDonald, E.R., 3rd; Hurov, K.E.; Luo, J.; Ballif, B.A.; Gygi, S.P.; Hofmann, K.; D'Andrea, A.D.; et al. Identification of the FANCI Protein, a Monoubiquitinated FANCD2 Paralog Required for DNA Repair. *Cell* **2007**, *129*, 289–301. [[CrossRef](#)]
126. Shah, R.B.; Kernan, J.L.; van Hoogstraten, A.; Ando, K.; Li, Y.; Belcher, A.L.; Mininger, I.; Bussenault, A.M.; Raman, R.; Ramanagoudr-Bhojappa, R.; et al. FANCI functions as a repair/apoptosis switch in response to DNA crosslinks. *Dev. Cell* **2021**, *56*, 2207–2222.E7. [[CrossRef](#)]
127. Ling, C.; Ishiai, M.; Ali, A.M.; Medhurst, A.L.; Neveling, K.; Kalb, R.; Yan, Z.; Xue, Y.; Oostra, A.B.; Auerbach, A.D.; et al. FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway. *EMBO J.* **2007**, *26*, 2104–2114. [[CrossRef](#)]
128. Haber, J.E.; Heyer, W.D. The Fuss about Mus81. *Cell* **2001**, *107*, 551–554. [[CrossRef](#)]
129. Abraham, J.; Lemmers, B.; Hande, M.P.; Moynahan, M.E.; Chahwan, C.; Ciccio, A.; Essers, J.; Hanada, K.; Chahwan, R.; Khaw, A.K.; et al. EME1 is involved in DNA damage processing and maintenance of genomic stability in mammalian cells. *EMBO J.* **2003**, *22*, 6137–6147. [[CrossRef](#)]
130. Cantor, S.B.; Bell, D.W.; Ganesan, S.; Kass, E.M.; Drapkin, R.; Grossman, S.; Wahrer, D.C.; Sgroi, D.C.; Lane, W.S.; Haber, D.A.; et al. BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function. *Cell* **2001**, *105*, 149–160. [[CrossRef](#)]
131. Langland, G.; Elliott, J.; Li, Y.; Creaney, J.; Dixon, K.; Groden, J. The BLM Helicase Is Necessary for Normal DNA Double-Strand Break Repair. *Cancer Res.* **2002**, *62*, 2766–2770. [[PubMed](#)]
132. Nimmonkar, A.V.; Genschel, J.; Kinoshita, E.; Polaczek, P.; Campbell, J.L.; Wyman, C.; Modrich, P.; Kowalczykowski, S.C. BLM–DNA2–RPA–MRN and EXO1–BLM–RPA–MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev.* **2011**, *25*, 350–362. [[CrossRef](#)] [[PubMed](#)]
133. Xia, B.; Sheng, Q.; Nakanishi, K.; Ohashi, A.; Wu, J.; Christ, N.; Liu, X.; Jasin, M.; Couch, F.J.; Livingston, D.M. Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. *Mol. Cell* **2006**, *22*, 719–729. [[CrossRef](#)] [[PubMed](#)]
134. Westermarck, U.K.; Reyngold, M.; Olshen, A.B.; Baer, R.; Jasin, M.; Moynahan, M.E. BARD1 Participates with BRCA1 in Homology-Directed Repair of Chromosome Breaks. *Mol. Cell. Biol.* **2003**, *23*, 7926–7936. [[CrossRef](#)] [[PubMed](#)]
135. Morris, J.R.; Solomon, E. BRCA1: BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. *Hum. Mol. Genet.* **2004**, *13*, 807–817. [[CrossRef](#)] [[PubMed](#)]
136. Wu-Baer, F.; Ludwig, T.; Baer, R. The UBXN1 Protein Associates with Autoubiquitinated Forms of the BRCA1 Tumor Suppressor and Inhibits Its Enzymatic Function. *Mol. Cell. Biol.* **2010**, *30*, 2787–2798. [[CrossRef](#)] [[PubMed](#)]
137. Marra, N.J.; Stanhope, M.J.; Jue, N.K.; Wang, M.; Sun, Q.; Pavinski Bitar, P.; Richards, V.P.; Komissarov, A.; Rayko, M.; Kliver, S.; et al. White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4446–4455. [[CrossRef](#)] [[PubMed](#)]
138. Copenhaver, G.P.; Mu, Y.; Bian, C.; Liu, R.; Wang, Y.; Shao, G.; Li, J.; Qiu, Y.; He, T.; Li, W.; et al. Whole genome sequencing of a snailfish from the Yap Trench (~7000 m) clarifies the molecular mechanisms underlying adaptation to the deep sea. *PLoS Genet.* **2021**, *17*, e1009530. [[CrossRef](#)]
139. Sullivan, M.R.; Bernstein, K.A. RAD-ical New Insights into RAD51 Regulation. *Genes* **2018**, *9*, 629. [[CrossRef](#)]
140. Liu, N.; Lamerdin, J.E.; Tebbs, R.S.; Schild, D.; Tucker, J.D.; Shen, M.R.; Brookman, K.W.; Siciliano, M.J.; Walter, C.A.; Fan, W.; et al. XRCC2 and XRCC3, New Human Rad51-Family Members, Promote Chromosome Stability and Protect against DNA Cross-Links and Other Damages. *Mol. Cell* **1998**, *1*, 783–793. [[CrossRef](#)]
141. Zhou, Y.; Liu, H.; Feng, C.; Lu, Z.; Liu, J.; Huang, Y.; Tang, H.; Xu, Z.; Pu, Y.; Zhang, H. Genetic adaptations of sea anemone to hydrothermal environment. *Sci. Adv.* **2023**, *9*, each0474. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.