



Article

Whole-Genome Survey Analyses Provide a New Perspective for the Evolutionary Biology of Shimofuri Goby, *Tridentiger bifasciatus*

Xiang Zhao ¹, Yaxian Liu ², Xueqing Du ¹, Siyu Ma ¹, Na Song ¹ and Linlin Zhao ^{3,4,*}

¹ The Key Laboratory of Mariculture, Ocean University of China, Ministry of Education, Qingdao 266003, China; zx15965582296@163.com (X.Z.); duxueqing0226@163.com (X.D.); msy05011@163.com (S.M.); songna624@163.com (N.S.)

² Yantai Laishan Marine Fisheries Supervision and Monitoring Brigade, Yantai 264000, China; liuyaxn@163.com

³ Key Laboratory of Marine Eco-Environmental Science and Technology, First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China

⁴ Marine Ecology and Environmental Science Laboratory, Pilot National Laboratory for Marine Science and Technology, Qingdao 266237, China

* Correspondence: zhaolinin@fio.org.cn

Simple Summary: Gobies form a group of small fish that are extremely adaptable to their environment. However, genomic resources for Gobiidae species are scarce. In this study, high-throughput sequencing technology was used to generate the whole genome data for the shimofuri goby (*Tridentiger bifasciatus*). In addition to providing basic genomic information, the mitochondrial whole genome, microsatellite markers, and single-copy homologous genes are also explored. The PSMC model is used to provide the historical population dynamics of the shimofuri goby. The results of this study provide a reference for the assembly of a complete genome map and biological research of the shimofuri goby.



Citation: Zhao, X.; Liu, Y.; Du, X.; Ma, S.; Song, N.; Zhao, L. Whole-Genome Survey Analyses Provide a New Perspective for the Evolutionary Biology of Shimofuri Goby, *Tridentiger bifasciatus*. *Animals* **2022**, *12*, 1914. <https://doi.org/10.3390/ani12151914>

Academic Editors: Tomasz Szymatola, Artur Gurgul, Ewa Ocloń and Klaudia Pawlina-Tyszko

Received: 8 June 2022

Accepted: 20 July 2022

Published: 27 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Abstract: The shimofuri goby (*Tridentiger bifasciatus*) is a small and highly adaptable goby, distributed along the coasts of China, the Sea of Japan, and the west coastal and estuarine areas of the Northwest Pacific. Next-generation sequencing was used to generate genome-wide survey data to provide essential characterization of the shimofuri goby genome and for the further mining of genomic information. The genome size of the shimofuri goby was estimated to be approximately 887.60 Mb through K-mer analysis, with a heterozygosity ratio and repeat sequence ratio of 0.47% and 32.60%, respectively. The assembled genome was used to identify microsatellite motifs (Simple Sequence Repeats, SSRs), extract single-copy homologous genes and assemble the mitochondrial genome. A total of 288,730 SSRs were identified. The most frequent SSRs were dinucleotide repeats (with a frequency of 61.15%), followed by trinucleotide (29.87%), tetranucleotide (6.19%), pentanucleotide (1.13%), and hexanucleotide repeats (1.66%). The results of the phylogenetic analysis based on single-copy homologous genes showed that the shimofuri goby and *Rhinogobius similis* can be clustered into one branch. The shimofuri goby was originally thought to be the same as the chameleon goby (*Tridentiger trigonocephalus*) due to their close morphological resemblance. However, a complete mitochondrial genome was assembled and the results of the phylogenetic analysis support the inclusion of the shimofuri goby as a separate species. PSMC analysis indicated that the shimofuri goby experienced a bottleneck event during the Pleistocene Glacial Epoch, in which its population size decreased massively, and then it began to recover gradually after the Last Glacial Maximum. This study provides a reference for the further assembly of the complete genome map of the shimofuri goby, and is a valuable genomic resource for the study of its evolutionary biology.

Keywords: whole-genome survey; *Tridentiger bifasciatus*; molecular markers; evolutionary biology; PSMC

1. Introduction

The shimofuri goby (*Tridentiger bifasciatus*) belongs to the Perciformes, Gobiidae and *Tridentiger*. It is a small goby distributed along the coasts of China, the Sea of Japan and the west coastal and estuarine areas of the Northwest Pacific [1]. The shimofuri goby was originally thought to be the same as the chameleon goby (*Tridentiger trigonocephalus*) due to their close morphological resemblance. In 1989, Akihito and Sakamoto identified the shimofuri goby as a separate species, based on differences in its sensory canals, pectoral fins and coloration [2]. After the 1980s, the shimofuri goby invaded the coastal waters and estuaries of the Eastern Pacific Ocean due to the development of pelagic fisheries and the carrying effect of ships' ballast water. It was first recorded in the San Francisco Estuary in 1985, and its range rapidly expanded to become one of the most abundant fish in the region [3–5]. This phenomenon demonstrates the powerful adaptability of the shimofuri goby. Previous studies have shown that the shimofuri goby has a wide range of temperature and salinity tolerances, even more than the environmental adaptation ability of indigenous species in the San Francisco Estuary [6]. Its strong environmental adaptability may be an important reason as to why the shimofuri goby successfully invaded other areas. The shimofuri goby has a short lifespan, high reproductive capacity and plastic life history characteristics [7]. In addition, the dietary composition of the shimofuri goby is very complex, including mollusks, aquatic insects, other macroinvertebrates, shrimp, fish, fish eggs and detritus [8,9]. These characteristics also contribute to the ability of the shimofuri goby to survive in other areas.

To date, there have been several genetic studies of the shimofuri goby. Jin et al. assembled the whole mitochondrial genome of the shimofuri goby by primer walking and preliminarily explored the phylogenetic relationships of Gobiidae [10]. Raphalo et al. used mitochondrial cytochrome C subunit I (*COI*) and ryanodine receptor 3 (*Ryr3*) DNA sequence data to assess the genetic diversity and genetic structure of eight coastal shimofuri goby populations in Zhejiang Province, China, and showed significant genetic divergence between the southern and northern populations [11]. Qin et al. developed and characterized 14 highly polymorphic microsatellite markers in the shimofuri goby using next-generation sequencing technology; notably, these markers are important for shimofuri goby genetics research [12]. In general, shimofuri goby genetic research is still largely lacking. Although gobies have no commercial value, they are key bait organisms for commercially important fish and play an integral role in the material cycle and energy flow in ecosystems. Therefore, it is necessary to reveal its biological, genetic, and evolutionary basis.

In recent years, the rapid development of high-throughput sequencing technology and various bioinformatics analysis methods has greatly promoted the marine fish genome and evolutionary biology research [13,14]. High-throughput sequencing data with whole-genome survey analyses can provide essential information on the genome size, repeat sequence ratio, genome heterozygosity, and GC content of a study species [15,16]. Compared with traditional methods, the use of genomic data makes it more efficient to assemble organelle genomes and develop microsatellite markers [17,18]. The assembled genomic data can be used to identify single-copy homologous genes [19–21]; in addition, the whole genome of a single species can also be used to predict a population's historical dynamics. In this study, the whole genome data for shimofuri goby were obtained using Illumina sequencing technology. In addition to providing basic genome data, the mitochondrial genome and single-copy homologous genes were extracted in order to reconstruct the phylogenetic relationship with other Gobiidae species. Based on the PSMC model, the historical dynamics of the shimofuri goby population were analyzed. These data greatly supplement the genetic information of shimofuri goby and provide new perspectives on the study of its genomic characteristics and evolutionary biology.

2. Materials and Methods

2.1. Sample Collection and Genome Survey Sequencing

Shimofuri goby samples were collected from Haiyang City, Shandong Province, China, in July 2021. Muscle tissues were stored in 95% ethanol at $-80\text{ }^{\circ}\text{C}$ and total genomic DNA was extracted using the standard phenol-chloroform method. The tested DNA samples were randomly broken into fragments using a Covaris ultrasonic crusher (Covaris, Woburn, MA, USA), and whole library preparation was completed by end repair, the addition of an A-tail, the addition of sequencing junction, purification and PCR amplification. The constructed libraries were sequenced using an Illumina HiSeq 2500 (San Diego, CA, USA, 150-bp paired-end reads, PE150, 350 bp insert size). The library construction and sequencing were performed at Novogene (Beijing, China). The whole-genome sequencing data were deposited in the Short Read Archive (SRA) database of National Center for Biotechnology Information (NCBI) under accession number PRJNA825009.

2.2. Data Analysis Methods

Quality control of raw sequencing data was performed using the FASTP software [22]. After removing low-quality reads, we randomly selected 10,000 pairs of clean reads and mapped them to the NCBI nucleotide (NT) database, displaying the top 80% of matched species. All clean reads were used to perform K-mer analysis. We used the jellyfish software to perform the K-mer calculation and generate the K-mer frequency statistics table [23]. Based on the results of the K-mer analysis, the K-mer depth distribution curve and peak were obtained and used to estimate the size of the genome. The genome size algorithm calculates: $G_{\text{size}} = N_{\text{kmer}} / C_{\text{kmer}}$, where G_{size} is the genome size, N_{kmer} is the total number of K-mer, and C_{kmer} is the peak depth. We consider the case of K-mer depth 1 as the error case, and the error rate obtained from the calculation can be used to correct the genome size. The final calculation method for the genome size is Revised $G_{\text{size}} = G_{\text{size}} \times (1 - \text{Error Rate})$. In addition, we used the GCE software to calculate the genomic heterozygosity and proportion of repetitive sequences [24].

Filtered clean reads were assembled into the completed mitochondrial genome using the Mitofinder software [25] with a reference sequence downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/>, accessed on 27 April 2022) under accession number NC_015992.1. Subsequently, the assembled mitochondrial sequence was annotated using the online fish mitochondrial genome database (<http://mitofish.aori.u-tokyo.ac.jp/>, accessed on 27 April 2022). Then, 13 protein-coding genes in mitochondrial genomes of 27 species belonging to 9 genera of Gobiidae were downloaded from the NCBI database for phylogenetic analysis (Table S1). We concatenated the nucleotide sequences of 13 protein-coding genes. Subsequently, the tandem sequences were aligned using the muscle method in MEGA X [26], and a phylogenetic tree was constructed using the maximum likelihood method based on 1000 resamples.

We used SOAPdenovo2 to preliminarily assemble the clean reads separately into unique contigs and scaffolds [27]. In the initial assembly, we used the default parameters of the SOAPdenovo2 software (without setting the “-u” parameter) to remove alternative contigs. To further assess the assembly quality, we conducted a BUSCO analysis [28]. The Perl script “misa.pl” of the MISA software was used to identify microsatellite motifs in the de novo draft genome [29]. The search parameters were set for the detection of di-, tri-, tetra-, penta-, and hexanucleotide microsatellite motifs with a minimum of 6, 5, 5, 5, and 5 repeats, respectively. The OrthoFinder and Busco5 software were used to search for single-copy direct homologous genes [28,30], and the completed genomes of 10 species of Gobiidae (*Periophthalmus magnuspinnatus* GCA_009829125.1; *Rhinogobius similis*: GCA_019453435.1; *Mugilogobius chulae*: GCA_016735935.1; *Boleophthalmus pectinirostris*: GCA_000788275.1; *Periophthalmodon schlosseri*: GCA_000787095.1; *Scartelaos histophorus*: GCA_000787155.1; *Chaenogobius annularis*: GCA_015082035.1; *Periophthalmus modestus*: GCA_019594935.1; *Lythrypnus dalli*: GCA_011763505.1; *Lesueurigobius sanzi*: GCA_900303255.1) were downloaded from the GenBank database and used for phylogenetic analyses. Subsequently,

extracted single-copy homologous genes of the shimofuri goby were annotated using the eggNOG-mapper software [31] based on eukaryotic ortholog groups (KOG) as well as the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Analysis of population size dynamics of the shimofuri goby was carried out using the PSMC model, as implemented in the PSMC software [32]. The “fq2psmcfa” and “splitfa” tools in the PSMC software were used to create the input file for the PSMC modelling. The PSMC analysis command included the options “-N25” for the number of cycles of the algorithm, “-t15” as the upper limit for the most recent common ancestor (TMRCA), “-r5” for the initial θ/ρ , and “-p 4 + 25 * 2 + 4 + 6” for atomic intervals. The reconstructed population history was plotted using the “psmc_plot.pl” script with the substitution rate “-u 2.5×10^{-8} ” and a generation time of 1 year.

3. Results

3.1. Whole-Genome Sequencing, K-Mer Analysis and Genome Assembly

A total of 74 Gb raw data were generated by the sequencing genome survey library with 350 bp inserts. The Q20, Q30 and GC content of raw data were 96.56, 91.57 and 39.62%, respectively. A total of 73.28 Gb clean data were obtained after filtering. From the NT database mapping results, randomly selected reads were matched to the DNA of closely related species of the shimofuri goby, which proved that the data in this study did not contain obvious exogenous contamination (Figure 1). The 19-mer frequency distribution derived from the sequencing reads is plotted in Figure 2. The K-mer analyses indicated that the peak of 19-mer distribution of the shimofuri goby was at $32\times$, and the error rate was 0.66%; thus, the estimated genome size was 887.60 Mb. The results of the GCE software analysis showed that the heterozygosity ratio and repeat sequence ratio were 0.47% and 32.60%, respectively.

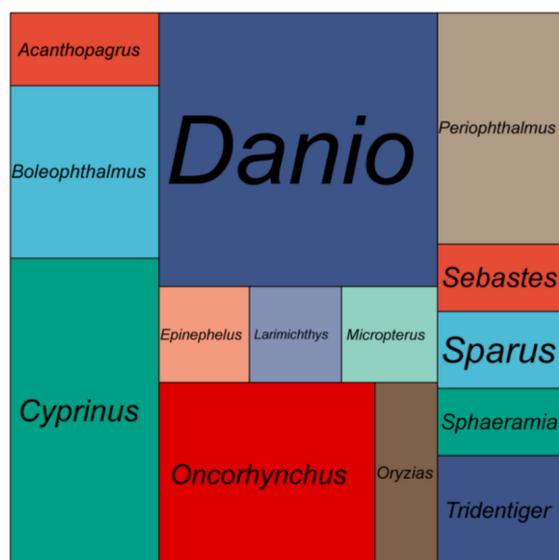


Figure 1. Mapping statistics of clean reads to the NT database. Area size represents the number of reads mapped to that species.

The filtered clean reads were used to assemble the draft genome. The total length, total number of sequences, max length of sequences, length of N50 and length of N90 of the contig and scaffold level genomes are shown in Table 1. The N50 length of contig and scaffold genomes were 691 and 1203, respectively. The N90 length of contig and scaffold genomes were both 127. The BUSCO analysis indicated that 25.8%, 37.5% and 36.7% of 3354 BUSCO genes were complete, fragmented and missing, respectively (see Figure 3). The average GC content of the draft genome was 38.30%. In Figure 4, the red regions represent relatively dense portions of the scatter plot.

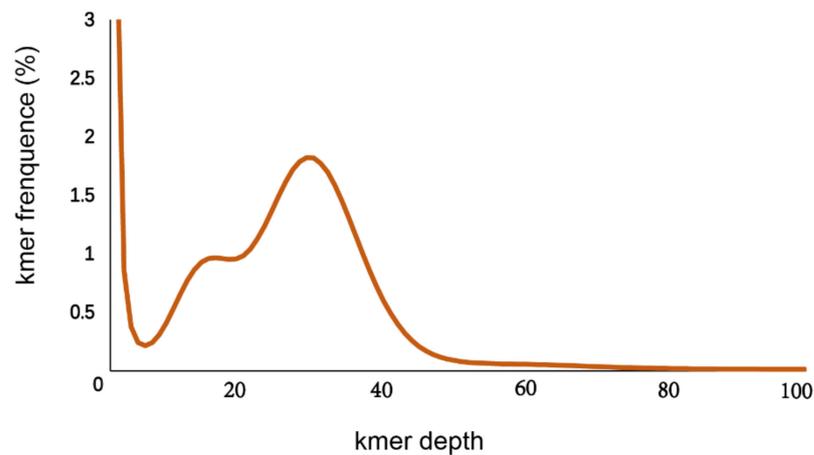


Figure 2. K-mer (19-mer) analysis for estimating the genome size of the shimofuri goby.

Table 1. The genome result of assembly for the shimofuri goby using Illumina clean data.

	Total Length (bp)	Total Number	Max Length (bp)	N50 Length (bp)	N90 Length (bp)
Contig	830,520,379	2,382,088	25,967	691	127
Scaffold	818,382,753	1,814,709	42,478	1203	127

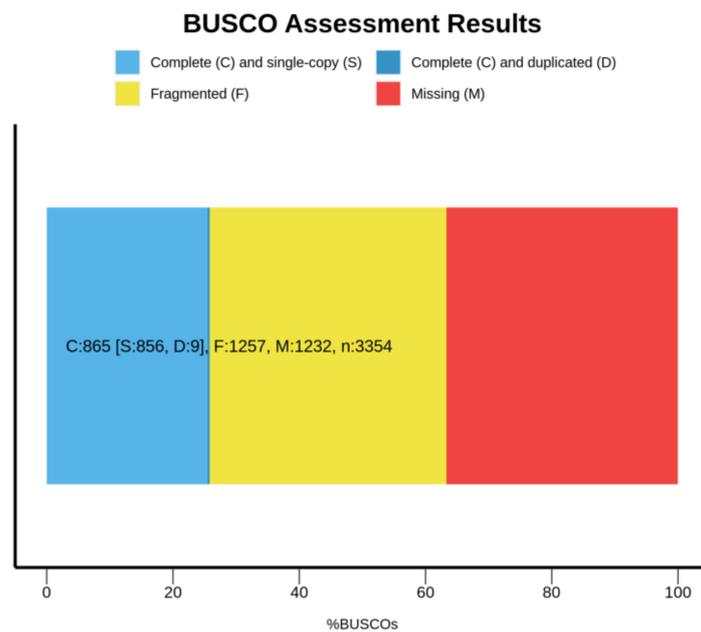


Figure 3. Assembly evaluation summary by BUSCO.

3.2. Identification and Characterization of SSR for the Genome of the Shimofuri Goby

From the 818,382,753 bp scaffold-level draft genome, a total of 288,730 SSRs were identified, which included 223,927 SSR-containing sequences. However, only 44,813 sequences contained more than one SSRs, and 30,549 SSRs were present in compound formation (Table 2). The SSRs distribution frequency in this genome was estimated to be approximately 347.10 SSRs per Mb. The motif types of SSRs included 61.15% dinucleotide, 29.87% trinucleotide, 6.19% tetranucleotide, 1.13% pentanucleotide and 1.66% hexanucleotide repeats (Figure 5a). Among the dinucleotides, the most frequent motifs were AC (14.27%) and CA (13.99%), followed by AT (11.73%) and TG (11.03%) (Figure 5b). Of the trinucleotides, the most frequent motif was AAT (10.88%), while CCT (1.30%) was the least frequent

trinucleotides motif (Figure 5c). Of the tetranucleotide repeats, the AAAT repeat (4.57%) was the most abundant motif, and the 5-fold repeat was predominant for all tetranucleotide repeats. (Figure 5d).

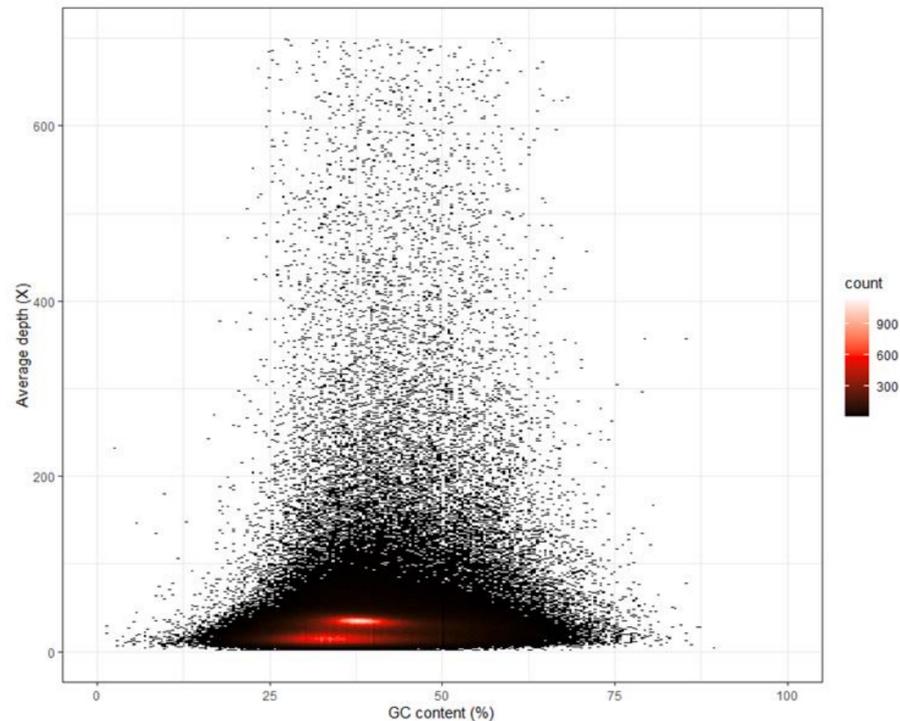


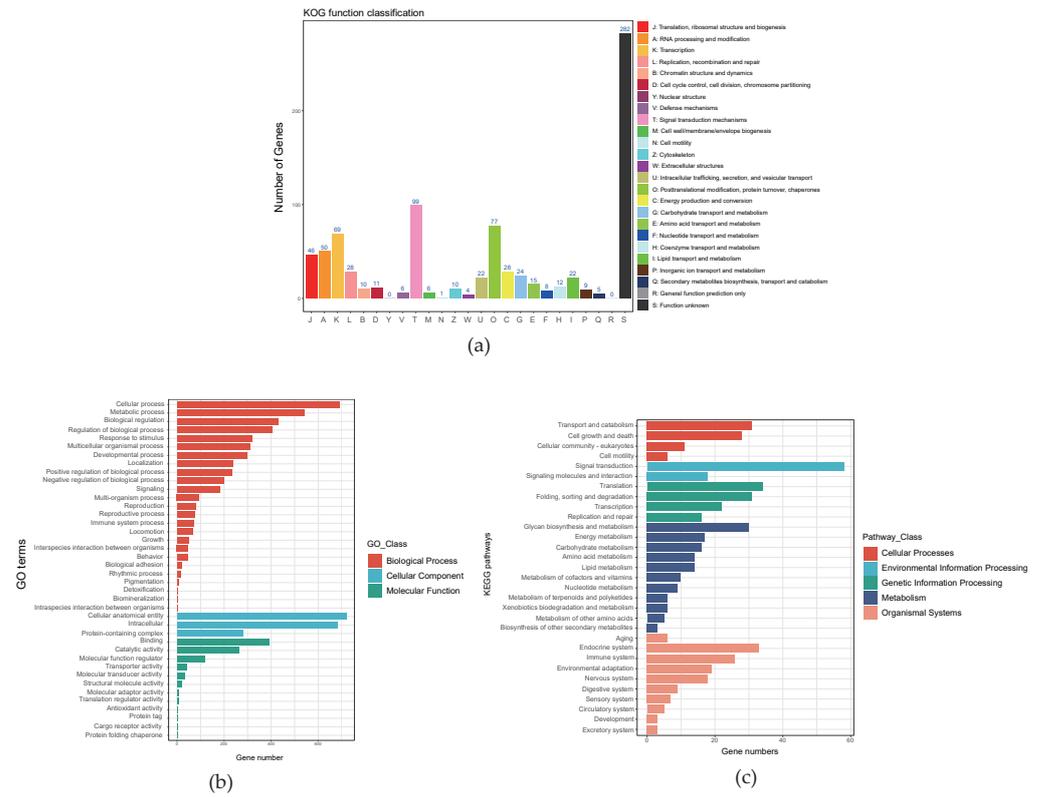
Figure 4. GC content and depth correlation analysis of the shimofuri goby. The x axis denotes the percentage GC content and the y axis represents sequencing depth.

Table 2. Statistics of microsatellite recognition results.

Statistical Items	Numbers
Total number of sequences examined	1,814,709
Total number of identified SSRs	288,730
Number of sequences containing more than 1 SSR	48,813
Number of SSRs present in compound formation	30,529

3.3. Identification of Single-Copy Homologous Genes in the Shimofuri Goby Genome

In this study, 856 single-copy homologous genes were identified in the shimofuri goby genome using the Busco5 software. The results of the KOG annotation demonstrated that the function of 282 single-copy homologous genes was unknown. The main known functions of single-copy homologous genes of the shimofuri goby were translation, ribosomal structure, and biogenesis, RNA processing and modification, transcription, signal transduction mechanisms, posttranslational modification, protein turnover and chaperones (Figure 6a). The results of the GO annotation showed that the main terms associated with single-copy homologous genes were cellular process, cellular anatomical entity and binding (Figure 6b). The results of KEGG annotation showed that the main pathways of single-copy homologous genes were transport and catabolism, signal transduction, translation, glycan biosynthesis and metabolism and endocrine system (Figure 6c). Both the Busco5 and OrthoFinder software were used to identify single-copy homologous genes in 10 goby species downloaded from NCBI (see Section 2 for NCBI accession numbers). We identified 295 single-copy homologous genes longer than 200 bp in the 11 goby species. OrthoFinder defaults to mafft for multiple sequence concatenation and fasttree for species evolution tree inference. The results of the phylogenetic analysis indicated that the species evolutionary



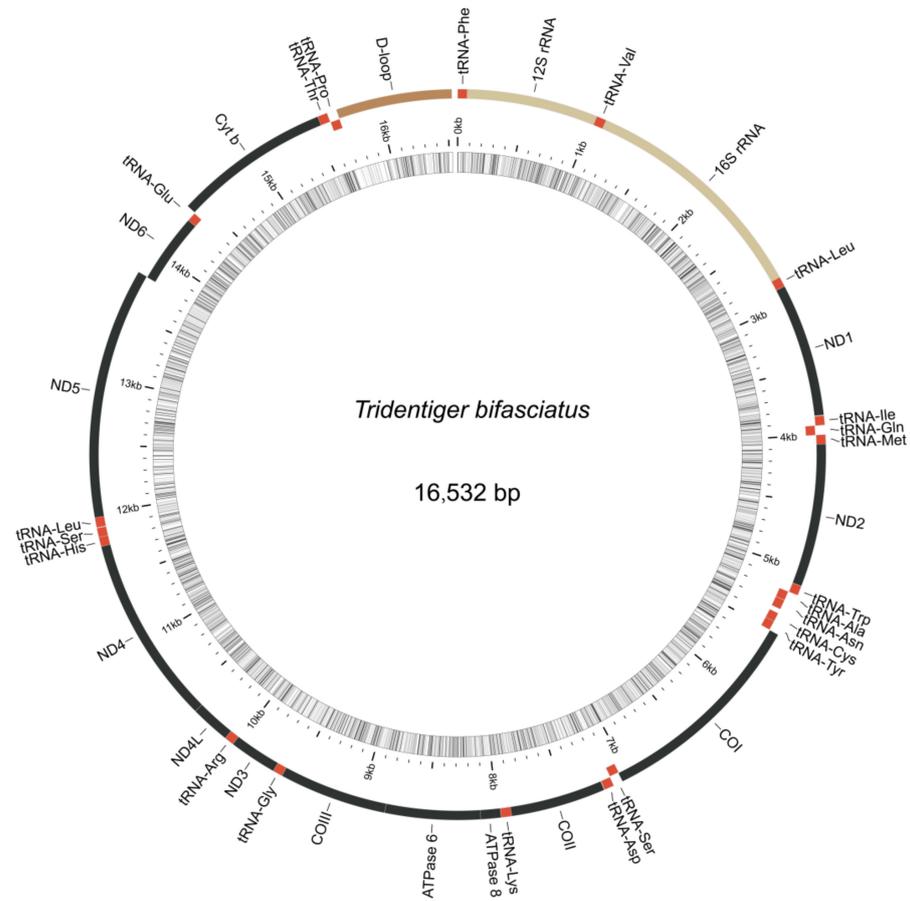


Figure 8. Mitochondrial genome organization of the shimofuri goby. Note: photo of the shimofuri goby from fishbase database (<https://www.fishbase.de/>, accessed on 27 April 2022), taken by Suzuki, Toshiyuki.

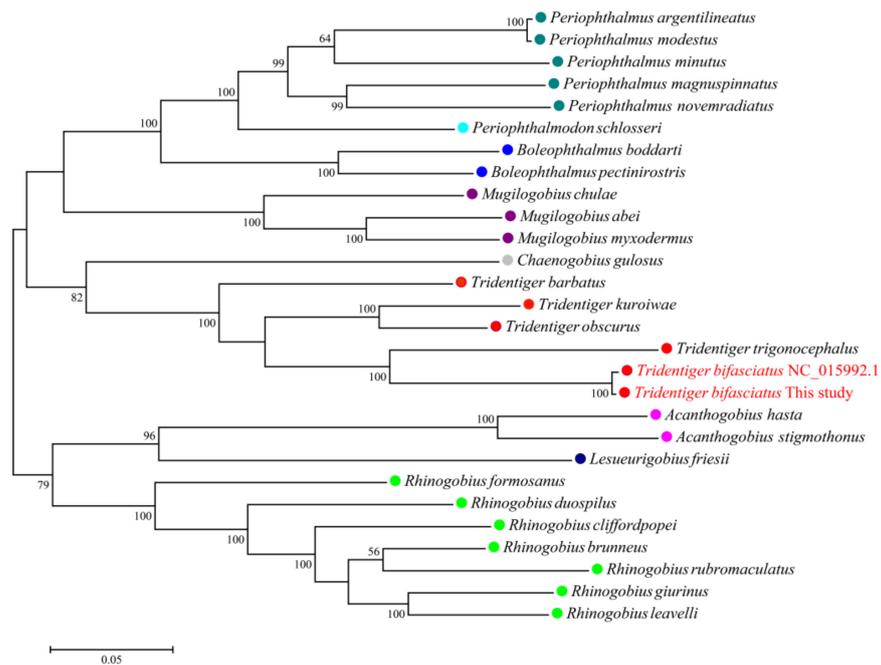


Figure 9. Maximum likelihood phylogenetic trees inferred from mitochondrial genomes of Gobiidae species based on 13 protein-coding genes. The same color represents species of the same genus.

3.5. Population Size Dynamics of the Shimofuri Goby

We used the PSMC model to infer historical changes in the effective population size (N_e ; see Figure 10). The PSMC results indicated that the shimofuri goby has experienced one bottleneck effect in the past million years, and its effective population size has been fluctuating. After the last interglacial period ($1.1\text{--}1.3 \times 10^5$ years ago; a relatively warm geological period), the effective population size reached its peak and then began to decrease continuously. Near the last glacial maximum (2.1×10^4 years ago, an extremely cold geological period), the effective population size of the shimofuri goby reached its minimum, then began to recover.

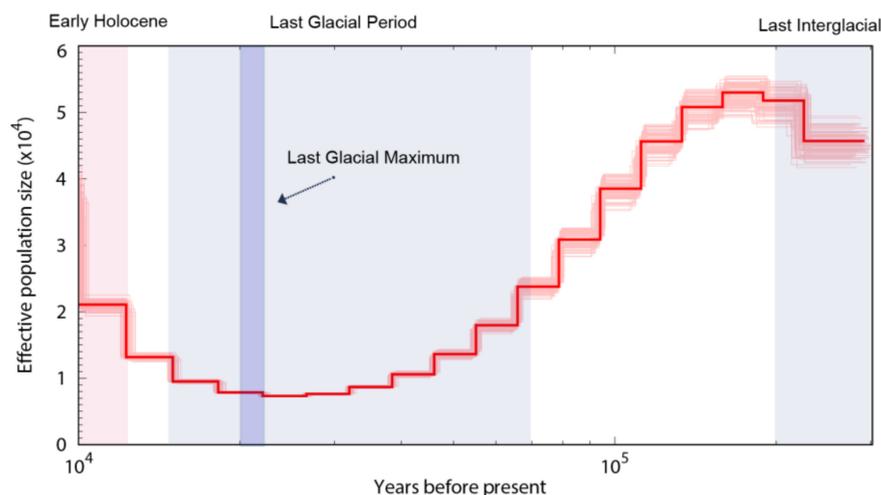


Figure 10. Demographic history of the shimofuri goby in this study. Note: PSMC estimates of demographic changes in effective population size (N_e) over time inferred from the draft genome sequences of the shimofuri goby. Thick lines represent the median and thin light lines correspond to 100 rounds of bootstrapping.

4. Discussion

At present, whole-genome resources of Gobiidae species are relatively scarce, and whole-genome survey sequencing is expected to become a rapid and low-cost technology for evolutionary biology research. In this study, a genome-wide survey of the shimofuri goby was conducted using whole-genome sequencing. The effective rate, Q20 and Q30 of the raw data were greater than 90%, indicating that the sequencing data were of good quality and could be used for subsequent analysis. The heterozygosity ratio and repeat sequence ratio in the shimofuri goby genome were 0.47% ($<0.5\%$) and 32.60% ($<50\%$), respectively. Therefore, we speculate that the shimofuri goby genome belongs to the common diploid genome. The size of the shimofuri goby genome (887.60 Mb) was similar to that of the most published species of Gobiidae, being smaller than those of *Boleophthalmus pectinirostris* (955.75 Mb) and *Mugilogobius chulae* (1003.36 Mb) [33,34]. Differences in genome size may be due to the differences in repeat sequence content and transposon insertion events [35].

The obtained filtered clean reads were used to assemble the draft genome. To date, this is the first shimofuri goby genome assembled based on second-generation whole-genome sequencing data. The information of this genome provides fundamental data for research on the evolutionary biology of the shimofuri goby, as well as a reference for the further exploration of the genomic features of this species. The N50 and N90 lengths of the shimofuri goby scaffold draft genome were low relative to the complete genome of a species due to the limitations of assembling genomes using Illumina second-generation sequencing data alone [36]. It is necessary to use third-generation sequencing data and Hi-C technology in future studies in order to assemble the genome of the shimofuri goby [37,38]. The shimofuri goby scaffold draft genome was used to identify SSRs. The percentage of

dinucleotide repeats was the highest and, as the repeat motif length increased, the number of loci decreased, similarly to that reported in other studies [39,40]. Previous studies have shown that long repetitive sequences have a higher mutation rate, which may lead to an increase in sequence instability, thus allowing for a decrease in the number of repeats [41]. Overall, motifs containing A or T were more abundant than those containing C or G, which is consistent with the results of genome-wide SSR studies of *Scatophagus argus* [42]. These SSR data will be important for the development of molecular markers in the shimofuri goby, although further validation studies using various populations are needed.

The shimofuri goby scaffold draft genome was used to identify single-copy homologous genes. Most of the single-copy genes belong to the housekeeper genes in the organism and, so, their functions are related to the regulation of various life activities of the organism [43]. Single-copy homologous genes are very conserved during speciation, and species formation may lead to the differentiation of homologous genes [44]. Hundreds of single-copy homologous genes obtained from genome-wide data may provide more accurate phylogenetic relationships between species than individual genetic markers [45]. The identification of single-copy homologous genes based on whole-genome survey sequencing data and phylogenetic analysis may become an effective method for the study of evolutionary relationships. Whole-genome sequencing data certainly contain not only nuclear genome sequences but also mitochondrial genome sequences. The large number of mitochondria in animal cells leads to a significantly higher mitochondrial genome sequencing depth than nuclear genes, which provided us with a sufficient amount of data to assemble a fine mitochondrial genome. The published shimofuri goby mitochondrial genome was assembled using the primer-walking method [10]. In this study, we assembled the shimofuri goby mitochondrial genome for the first time based on whole-genome sequencing data, which was consistent with the genetic composition of the published shimofuri goby mitochondrial genome. The GC content of the mitochondrial genome was 44.15%, showing obvious AT-bias, which is consistent with that observed in most fish [46]. Among the 13 protein-coding genes, the COI gene starts with GTG, the ATP6 gene starts with ATA, and the other 11 genes start with ATG. The translation initiation efficiency of these three initiation codons is high, among which ATG is the most efficient initiation codon [47]. Phylogenetic analysis based on 13 protein-coding genes showed that the shimofuri goby and the chameleon goby (*Tridentiger trigonocephalus*) were clustered together, but their evolutionary dendritic lengths did not coincide—suggesting that, although the sequences of the two species are extremely similar (0.17% sequence divergence), they did not diverge at the same time [48]. Therefore, the results of this study support the classification of the shimofuri goby as a separate species. In addition, the species of *Tridentiger* were closely related to the *Chaenogobius gulosus* and were divided into two large branches with species of the *Rhinogobius*. These results present some differences with species trees constructed based on single-copy homologous genes. We speculate that the 13 protein-coding genes of the mitochondrial genome have some limitations for the estimation of the evolutionary relationships between the taxonomic status of species. Furthermore, there are differences in the base substitution rates between mitochondrial genes and nuclear genes, which may cause some errors in estimating evolutionary relationships [49].

Population history dynamics have long been a hot topic of research in genetic and evolutionary biology. We used the PSMC model to infer historical changes in the effective population size of the shimofuri goby. Our results showed that the shimofuri goby experienced one bottleneck event during the Pleistocene Glacial Epoch, when its population size decreased massively and began to recover gradually after the Last Glacial Maximum. It has also been shown, in the study of Raphalo et al., that the shimofuri goby community off the coast of Zhejiang, China, may have expanded abruptly after experiencing a bottleneck event during the Pleistocene Glacial Epoch [11]; notably, the Zhejiang shimofuri goby population presented a low genetic diversity index compared to other marine fishes [11,50,51]. The accumulation of new mutations is very slow for this species, and genetic diversity is difficult to recover once it has declined [52]. Therefore, we speculate that a large amount

of genetic diversity was lost in the shimofuri goby population during the contraction period. Although the population size of the shimofuri goby rapidly increased after the Last Glacial Maximum., the genetic diversity of the shimofuri goby population did not increase sufficiently, as there are highly similar “redundant copies” in the increased number of individuals from a genetic point of view.

5. Conclusions

In this study, the genomic survey analysis of the shimofuri goby was performed using high-throughput sequencing and genomic information, such as genome size, heterozygosity, GC content and repeat sequence ratio. We showed that the shimofuri goby genome is a common diploid genome. Furthermore, we screened a large number of SSR markers from the shimofuri goby genome. A phylogenetic tree constructed based on single-copy homologous genes clarified the phylogenetic relationships of some Gobiidae species. The assembled mitochondrial genome not only complemented the available shimofuri goby genetic resources, but also further supported its taxonomic status. PSMC analysis showed that the effective population size of the shimofuri goby decreased sharply during the Pleistocene ice age, and then recovered gradually. This study provides a reference for the further assembly of the shimofuri goby genome and offers a new perspective on its evolutionary biology. Future genome survey studies should consider six components: basic genome characterization, genome assembly, SSR marker screening, mitochondrial genome assembly, single-copy homology gene screening and PSMC analysis.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ani12151914/s1>, Table S1: NCBI accession numbers of 13 protein-coding genes of the mitochondrial genomes of 27 species in 9 genera of Gobiidae.

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

Funding: This research was supported by the opening foundation of the Observation and Research Station of Bohai Strait Eco-Corridor, MNR, grant No. BH202203.

Institutional Review Board Statement: This study does not involve endangered and live animals. The shimofuri goby samples used in this study were caught during fishery production and were dead at the time of collection. All animal experiments were conducted following the guidelines and approval of the Animal Research and Ethics Committees of the Ocean University of China.

Informed Consent Statement: Not applicable.

Data Availability Statement: The whole-genome sequencing data were deposited in the Short Read Archive (SRA) database of National Center for Biotechnology Information (NCBI) under accession number PRJNA825009.

Acknowledgments: We would like to express deep gratitude to Changdong Liu of Ocean University of China and his team for providing the samples for this study.

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

References

1. Meng, L.; Moyle, P.B.; Herbold, B. Changes in abundance and distribution of native and introduced fishes of suisun marsh. *Trans. Am. Fish. Soc.* **1994**, *123*, 498–507. [[CrossRef](#)]
2. Sakamoto, K. Reexamination of the status of the striped goby. *Jpn. J. Ichthyol.* **1989**, *36*, 100–112. [[CrossRef](#)]
3. O’grady, K.; Cowx, I. Nonindigenous fishes introduced into inland waters of the united states. *Fish. Manag. Ecol.* **2000**, *2*, 172–175. [[CrossRef](#)]
4. Matern, S.A.; Moyle, P.B.; Pierce, L.C. Native and alien fishes in a california estuarine marsh: Twenty-one years of changing assemblages. *Trans. Am. Fish. Soc.* **2002**, *131*, 797–816. [[CrossRef](#)]
5. Matern, S.A.; Fleming, K.J. Invasion of a third Asian goby, *Tridentiger bifasciatus*, into California. *Calif. Fish. Game* **1995**, *81*, 71–76.

6. Matern, S.A. Using temperature and salinity tolerances to predict the success of the shimofuri goby, a recent invader into California. *Trans. Am. Fish. Soc.* **2001**, *130*, 592–599. [[CrossRef](#)]
7. Qin, J.; Xiang, M.; Jia, M.; Cheng, F.; Xie, S. Combined opportunistic and equilibrium life-history traits facilitate successful invasions of the shimofuri goby (*Tridentiger bifasciatus*). *Aquat. Invasions* **2020**, *15*, in press.
8. Kim, H.; Yu, N.L.; Jeong, J.M.; Kim, H.J.; Baeck, G.W. Diet composition of juvenile *Trachurus japonicus* in the coastal waters of Geumdo Yeosu, Korea. *Bull. Korean Soc. Fish. Technol.* **2015**, *51*, 637–643. [[CrossRef](#)]
9. Qin, J.; Xie, S.; Cheng, F. Broad diet composition and seasonal feeding variation facilitate successful invasion of the shimofuri goby (*Tridentiger bifasciatus*) in a water transfer system. *Water* **2020**, *12*, 3411. [[CrossRef](#)]
10. Jin, X.; Wang, R.; Wei, T.; Tang, D.; Xu, T. Complete mitochondrial genome sequence of *Tridentiger bifasciatus* and *Tridentiger barbatus* (Perciformes, Gobiidae): A mitogenomic perspective on the phylogenetic relationships of Gobiidae. *Mol. Biol. Rep.* **2015**, *42*, 253–265. [[CrossRef](#)]
11. Raphalo, E.M.; Gu, C.; Lü, Z.; Chen, Y. Genetic structure of intertidal shimofuri goby in the Zhejiang coastal sea, China: Implications for management. *Mar. Biodivers.* **2018**, *49*, 1–9. [[CrossRef](#)]
12. Qin, J.; Victor, S.B.; Cheng, F.; Xie, S. Development and characterization of 14 novel microsatellite markers for an invasive goby (*Tridentiger bifasciatus*) in water transfer system. *J. Appl. Ichthyol.* **2021**, *37*, 314–317.
13. Chen, B.; Sun, Z.; Lou, F.; Gao, T.X.; Song, N. Genomic characteristics and profile of microsatellite primers for *Acanthogobius ommaturus* by genome survey sequencing. *Biosci. Rep.* **2020**, *40*, BSR20201295. [[CrossRef](#)] [[PubMed](#)]
14. Xu, S.; Zhang, H.; Gao, T. Comprehensive whole genome survey analyses of male and female brown-spotted flathead fish *Platycephalus* sp.1. *Genomics* **2020**, *112*, 4742–4748. [[CrossRef](#)]
15. Bi, Q.; Zhao, Y.; Cui, Y.; Wang, L. Genome survey sequencing and genetic background characterization of yellow horn based on next-generation sequencing. *Mol. Biol. Rep.* **2019**, *46*, 4303–4312. [[CrossRef](#)]
16. Lu, M.; An, H.; Li, L. Genome Survey Sequencing for the Characterization of the Genetic Background of *Rosa roxburghii* Tratt and Leaf Ascorbate Metabolism Genes. *PLoS ONE* **2016**, *11*, e0147530. [[CrossRef](#)] [[PubMed](#)]
17. Ziya Motalebipour, E.; Kafkas, S.; Khodaeiaminjan, M.; Çoban, N.; Gözel, H. Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: Development of novel SSR markers and genetic diversity in *Pistacia* species. *BMC Genom.* **2016**, *17*, 998. [[CrossRef](#)]
18. Song, H.; Zhang, Y.X.; Yang, M.J.; Sun, J.C.; Zhang, T.; Wang, H.Y. Genome survey on invasive veined rapa whelk (*Rapana venosa*) and development of microsatellite loci on large scale. *J. Genet.* **2018**, *97*, e79–e86. [[CrossRef](#)]
19. Gilbert, P.S.; Chang, J.; Pan, C.; Sobel, E.M.; Sinsheimer, J.S.; Faircloth, B.C.; Alfaro, M.E. Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Mol. Phylogenetics Evol.* **2015**, *92*, 140–146. [[CrossRef](#)]
20. Van Dam, M.H.; Henderson, J.B.; Esposito, L.; Trautwein, M. Genomic Characterization and Curation of UCEs Improves Species Tree Reconstruction. *Syst. Biol.* **2021**, *70*, 307–321. [[CrossRef](#)]
21. Koprek, T.; Rangel, S.; McElroy, D.; Louwerse, J.D.; Williams-Carrier, R.E.; Lemaux, P.G. Transposon-mediated single-copy gene delivery leads to increased transgene expression stability in barley. *Plant Physiol.* **2001**, *125*, 1354–1362. [[CrossRef](#)]
22. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)] [[PubMed](#)]
23. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [[CrossRef](#)]
24. Liu, B.; Shi, Y.; Yuan, J.; Hu, X.; Zhang, H.; Li, N. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *Quant. Biol.* **2013**, *35*, 62–67.
25. Allio, R.; Schomaker-Bastos, A.; Romiguier, J.; Prosdocimi, F.; Nabholz, B.; Delsuc, F. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol. Ecol. Resour.* **2020**, *20*, 892–905. [[CrossRef](#)] [[PubMed](#)]
26. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
27. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 18. [[CrossRef](#)]
28. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
29. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)]
30. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [[CrossRef](#)]
31. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [[CrossRef](#)] [[PubMed](#)]
32. Li, H.; Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **2011**, *475*, 493–496. [[CrossRef](#)]

33. Cai, L.; Liu, G.; Wei, Y.; Zhu, Y.; Li, J.; Miao, Z.; Chen, M.; Yue, Z.; Yu, L.; Dong, Z.; et al. Whole-genome sequencing reveals sex determination and liver high-fat storage mechanisms of yellow stripe goby (*Mugilogobius chulae*). *Commun. Biol.* **2021**, *4*, 15. [[CrossRef](#)] [[PubMed](#)]
34. 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](#)] [[PubMed](#)]
35. Shao, F.; Han, M.; Peng, Z. Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* **2019**, *9*, 15399. [[CrossRef](#)]
36. Zhao, L.; Xu, S.; Han, Z.; Liu, Q.; Ke, W.; Liu, A.; Gao, T. Chromosome-Level Genome Assembly and Annotation of a Sciaenid Fish, *Argyrosomus japonicus*. *Genome Biol. Evol.* **2021**, *13*, evaa246. [[CrossRef](#)]
37. Belton, J.M.; McCord, R.P.; Gibcus, J.H.; Naumova, N.; Zhan, Y.; Dekker, J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **2012**, *58*, 268–276. [[CrossRef](#)]
38. Kumar, K.R.; Cowley, M.J.; Davis, R.L. Next-Generation Sequencing and Emerging Technologies. *Semin. Thromb. Hemost.* **2019**, *45*, 661–673. [[CrossRef](#)]
39. Jo, E.; Lee, S.J.; Choi, E.; Kim, J.; Lee, S.G.; Lee, J.H.; Kim, J.H.; Park, H. Whole genome survey and microsatellite motif identification of *Artemia franciscana*. *Biosci. Rep.* **2021**, *41*, BSR20203868. [[CrossRef](#)]
40. Jo, E.; Cho, Y.H.; Lee, S.J.; Choi, E.; Kim, J.; Kim, J.H.; Chi, Y.M.; Park, H. Genome survey and microsatellite motif identification of *Pogonophryne albipinna*. *Biosci. Rep.* **2021**, *41*, BSR20210824. [[CrossRef](#)]
41. Harr, B.; Schl ötterer, C. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* **2000**, *155*, 1213–1220. [[CrossRef](#)] [[PubMed](#)]
42. Huang, Y.; Jiang, D.; Li, M.; Mustapha, U.F.; Tian, C.; Chen, H.; Huang, Y.; Deng, S.; Wu, T.; Zhu, C.; et al. Genome Survey of Male and Female Spotted Scat (*Scatophagus argus*). *Animals* **2019**, *9*, 1117. [[CrossRef](#)] [[PubMed](#)]
43. Huang, Y.; Feulner, P.; Eizaguirre, C.; Lenz, T.L.; Bornberg-Bauer, E.; Milinski, M.; Reusch, T.; Chain, F. Genome-Wide Genotype-Expression Relationships Reveal Both Copy Number and Single Nucleotide Differentiation Contribute to Differential Gene Expression between Stickleback Ecotypes. *Genome Biol. Evol.* **2019**, *11*, 2344–2359. [[CrossRef](#)] [[PubMed](#)]
44. Fitch, W.M. Distinguishing homologous from analogous proteins. *Syst. Zool.* **1970**, *19*, 99–113. [[CrossRef](#)]
45. Delsuc, F.; Brinkmann, H.; Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **2005**, *6*, 361–375. [[CrossRef](#)]
46. Consuegra, S.; John, E.; Verspoor, E.; Leaniz, C.D. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genet. Sel. Evol.* **2015**, *47*, 58. [[CrossRef](#)] [[PubMed](#)]
47. Hecht, A.; Glasgow, J.; Jaschke, P.R.; Bawazer, L.A.; Munson, M.S.; Cochran, J.R.; Endy, D.; Salit, M. Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Res.* **2017**, *45*, 3615–3626. [[CrossRef](#)]
48. Morel, B.; Schade, P.; Lutteropp, S.; Williams, T.A.; Szöllösi, G.J.; Stamatakis, A. SpeciesRax: A Tool for Maximum Likelihood Species Tree Inference from Gene Family Trees under Duplication, Transfer, and Loss. *Mol. Biol. Evol.* **2022**, *39*, msab365. [[CrossRef](#)]
49. Martin, A.P.; Kessing, B.D.; Palumbi, S.R. Accuracy of estimating genetic distance between species from short sequences of mitochondrial DNA. *Mol. Biol. Evol.* **1990**, *7*, 485–488.
50. Canales-Aguirre, C.B.; Ferrada-Fuentes, S.; Galleguillos, R.; Oyarzun, F.X.; Buratti, C.C.; Hernández, C.E. High genetic diversity and low-population differentiation in the Patagonian sprat (*Sprattus fuegensis*) based on mitochondrial DNA. *Mitochondrial DNA. Part. A* **2018**, *29*, 1148–1155. [[CrossRef](#)]
51. Debes, P.V.; Zachos, F.E.; Hanel, R. Mitochondrial phylogeography of the European sprat (*Sprattus sprattus* L., Clupeidae) reveals isolated climatically vulnerable populations in the Mediterranean Sea and range expansion in the northeast Atlantic. *Mol. Ecol.* **2008**, *17*, 3873–3888. [[CrossRef](#)] [[PubMed](#)]
52. Latta, L.C.; Fisk, D.L.; Knapp, R.A.; Pfrender, M.E. Genetic resilience of daphnia populations following experimental removal of introduced fish. *Conserv. Genet.* **2010**, *11*, 1737–1745. [[CrossRef](#)]