scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of the marble goby (*Oxyeleotris marmorata*)

Haiyang Liu[1,2] ✉, Jiajun Yao[1,3], Mengdan Dong[1,3], Meisheng Yi[1,3] & Wei Liu[1,3] ✉

The marble goby (*Oxyeleotris marmorata*) is an important aquaculture species in South China and Southeast Asia due to its strong adaptability and rapid growth rate. However, the lack of comprehensive genomic resources has hindered in-depth genetic studies and breeding programs. In this study, we generated a high-quality, chromosome-level genome assembly of marble goby using PacBio HiFi long-read sequencing and Hi-C technologies. The final assembly spans 1,068.42 Mb, with 95.51% of the genome anchored to 23 chromosomes. The assembly exhibits a contig N50 of 13.72 Mb and a scaffold N50 of 45.09 Mb. Genome completeness was assessed with a BUSCO score of 98.19%. A total of 22,964 protein-coding genes were predicted, with a BUSCO completeness of 96.15%, and 99.25% of the genes were functionally annotated. Approximately 51.80% of the genome is identified as repeat elements. This high-fidelity genome assembly provides a valuable resource for molecular breeding, comparative genomics, and evolutionary studies of marble goby and related species.

Background & Summary

Marble goby (*Oxyeleotris marmorata*), also known as the Thai goby or Thai marble goby, belongs to the order Perciformes, family Eleotridae, and genus *Oxyeleotris*[1,2]. It is native to several Southeast Asian countries, including Thailand, Malaysia and Indonesia, and has gradually been introduced to other regions such as China's Pearl River Delta[3,4]. This species is benthic and burrow-dwelling, typically inhabiting freshwater bodies with clear or moderately flowing water[5]. Among species within the genus *Oxyeleotris*, the marble goby exhibits superior growth rates and maximum body size. Individuals can reach up to 65 cm in length and weigh as much as 5 kg, although the typical market size ranges from 500 to 1000 g. The species is highly valued for its tender, low-fat, and nutritious flesh, and is widely appreciated for its taste[6]. In Asian markets, marble goby commands a premium price, ranging from 10 to 20 USD per kilogram, underscoring its considerable economic importance[7]. Since its successful introduction to China in the 1980s, significant progress has been made in artificial breeding, seedling cultivation, and aquaculture techniques[4]. Over the past few decades, marble goby farming has expanded significantly across South China, establishing it as one of the most important high-value aquaculture species.

Previous research on the marble goby has primarily focused on population genetics, aquaculture, reproduction, and nutritional feed development[4,8,9]. In the field of population genetics, mitochondrial DNA sequencing has been widely used[2,10], along with genetic background analysis using mitochondrial DNA polymorphism[11–13] and microsatellites[14]. Notably, male marble gobies at two years of age exhibit an average body weight that is 56.3% higher than that of females, indicating pronounced sexual dimorphism[15]. However, cytogenetic analyses have not identified any sex-linked heteromorphic chromosomes in this species[16]. Law *et al.* used amplified fragment length polymorphism (AFLP) technology to develop a sequence-characterized amplified region (SCAR) marker specific to females, providing preliminary insights into sex-related genetic markers[17]. Nevertheless, due to the lack of a high-quality chromosome-level reference genome, the precise chromosomal location and functional context of this marker remain unresolved. Additionally, Liu *et al.* identified key genes such as *vasa*, *dmrt1* and *foxl2*, which are involved in germ cell development and sex differentiation in marble gobies[15,18]. However, the specific roles of these genes in sex determination have yet to be fully elucidated. Thus, the development of

[1]School of Marine Sciences, Sun Yat-sen University, Guangzhou, Guangdong, 510275, China. [2]Key Laboratory of Tropical and Subtropical Fishery Resources Application and Cultivation, Ministry of Agriculture and Rural Affairs, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, 510380, China. [3]Guangdong Provincial Key Laboratory of Marine Resources and Coastal Engineering, Guangzhou, Guangdong, 510000, China. ✉e-mail: hyliu@prfri.ac.cn; liuw68@mail.sysu.edu.cn

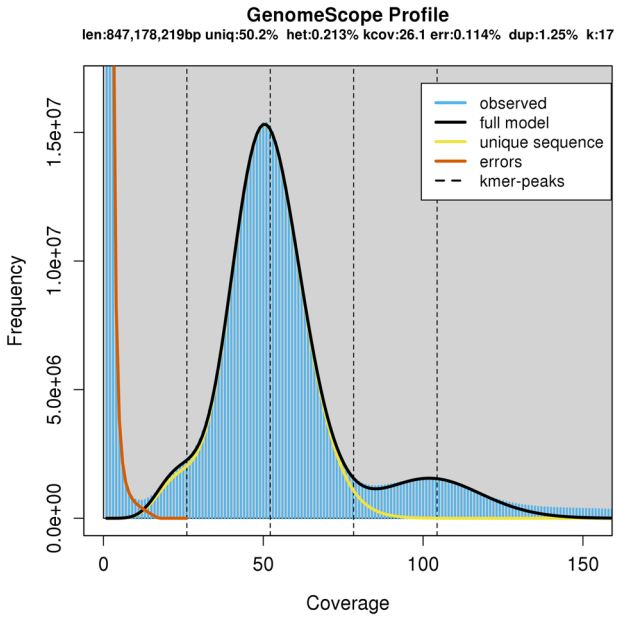

Fig. 1 Genome survey of *Oxyeleotris marmorata* was conducted using 17-mer analysis with GenomeScope2. The estimated genome size was 847.18 Mb, with a heterozygosity rate of 0.213% and repeat content of 49.8%.

a high-quality reference genome is essential for facilitating genetic analyses of economically important traits, identifying candidate genes, and improving breeding efficiency in this species[19–21]. Currently, genomic resources for the genus *Oxyeleotris* are extremely limited, and no assembled genome has been reported for *Oxyeleotris marmorata*.

To address these limitations and establish a foundation for future genetic research, we successfully assembled the first chromosome-level reference genome of marble goby using Illumina short reads, PacBio HiFi long reads, and Hi-C scaffolding data. The assembled genome spans approximately 1,068.42 Mb, with a contig N50 of 13.72 Mb. The final chromosome-level assembly consists of 23 chromosomes, and a total of 22,964 genes were annotated. This high-quality genomic resource provides important insights into the biological characteristics and functional genomic research of the marble goby, and serves as a valuable reference for future phylogenetic and evolutionary studies within the genus *Oxyeleotris*.

Methods

Sample collection and DNA extraction. A mature female marble goby was collected from Zhuhai, Guangdong, China. Genomic DNA was extracted from muscle tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, USA), following the manufacturer's instructions. DNA integrity and purity were assessed through 1% agarose gel electrophoresis, and the concentration was measured using a NanoDrop One spectrophotometer (Thermo Scientific, USA).

Genome sequencing. For short-read sequencing, a 350 bp insert-size library was prepared from sheared DNA using the Illumina DNA Prep kit (Illumina, USA) and sequenced on the NovaSeq 6000 platform (Illumina, USA), generating 56.79 Gb of PE150 raw reads. For long-read sequencing, high-fidelity (HiFi) libraries were constructed with the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, USA) and sequenced on the PacBio Sequel II system, producing 36.15 Gb of continuous long reads with an average length of 15.59 kb. Additionally, Hi-C libraries were prepared from approximately 1 g of marble goby muscle tissue using the GrandOmics Hi-C kit (DpnII digestion) and sequenced on the NovaSeq 6000 platform, resulting in 116.83 Gb of Hi-C data.

RNA extraction and transcriptome sequencing. For comprehensive genome annotation, RNA was extracted from ten tissues (spleen, kidney, brain, muscle, ovary, liver, intestine, heart, gill, and swim bladder). RNA quality was evaluated using both a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and an Agilent 2100 Bioanalyzer. Equal amounts (1 μg per tissue) of RNA were pooled to construct a cDNA library with the TruSeq Stranded mRNA Prep Kit (Illumina), followed by PE150 sequencing mode on the NovaSeq 6000 platform. This resulted in 16.70 Gb of transcriptomic data to support genome annotation.

Genome size and heterozygosity estimation. The genome size of the marble goby was estimated through k-mer analysis of Illumina clean reads. A 17-mer frequency table was generated using Jellyfish v2.3.0[22] and subsequently analyzed with GenomeScope v2.0[23]. The analysis revealed an estimated genome size of approximately 847.18 Mb, with a heterozygosity rate of 0.213% and repeat content of 49.8% (Fig. 1).

Feature	Value
Assembly genome length (Mb)	1,068.42
GC (%)	40.15
contigs number	640
contig N50 (Mb)	13.72
contig N90 (Mb)	0.97
Longest contig (Mb)	33.87
scaffolds number	332
scaffold N50 (Mb)	45.09
scaffold N90 (Mb)	36.56
Longest scaffold (Mb)	64.39
Number of chromosomes	23

Table 1. Summary statistics of *Oxyeleotris marmorata* genome assembly.

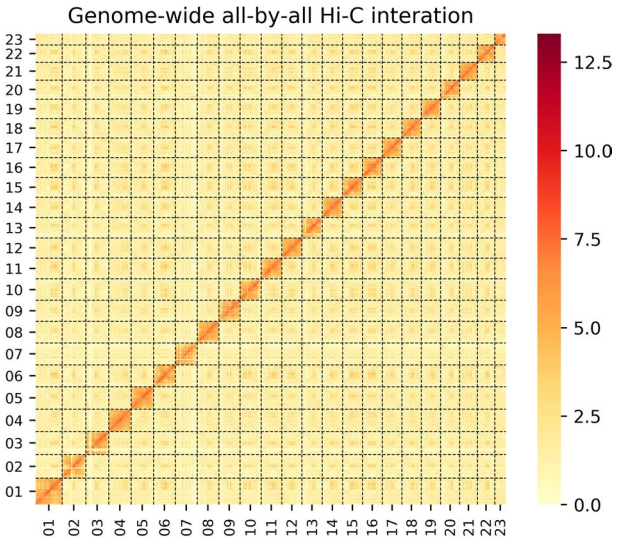

Fig. 2 Hi-C interactive heatmap of 23 chromosomes in *Oxyeleotris marmorata* genome.

Genome assembly. The genome was assembled using Hifiasm v0.19.5[24] with default parameters, producing 640 contigs with a total length of 1,068.42 Mb. The maximum contig size was 33.87 Mb, with an N50 of 13.72 Mb (Table 1). For scaffolding, we employed an integrated approach combining Juicer v1.6[25] and 3D-DNA v201008[26]. After indexing the contig-level genome with BWA v0.7.17[27], Hi-C reads were mapped to generate restriction site data. The 3D-DNA algorithm was used for initial chromosome assembly, followed by manual refinement of chromosome boundaries and scaffold correction using Juicerbox v1.11.08[28]. This process resolved 23 chromosomes (Figs. 2, 3) and yielded a final assembly of 332 scaffolds, with a maximum scaffold size of 64.39 Mb and an N50 size of 45.09 Mb (Table 1, Table S1).

Repeat annotation. We conducted a comprehensive repeat analysis using multiple computational methods. For tandem repeats, GMATA v2.2.1[29] and Tandem Repeats Finder v4.10.0[30] were used to identify simple sequence repeats (SSRs) and broader tandem repeats, which covered 0.74% and 1.18% of the genome, respectively. Dispersed repeats were characterized using an integrated pipeline. First, MITEs[31] were detected with MITE-hunter to create a MITE library. The genome was then hardmasked and analyzed with RepeatModeler v2.0.6[32] to generate a *de novo* repeat library (RepMod.lib). TEclass was used to classify unknown repeats in RepMod.lib, which was subsequently combined with MITE.lib and Repbase v19.06[33] to construct a combined repeat library. RepeatMasker v4.1.6[34] applied this library for genome-wide repeat annotation. The analysis revealed that 41.44% of the genome consists of dispersed repeats (Table 2). DNA transposons were the most abundant (22.30%), followed by LINEs (10.29%), LTR retrotransposons (5.58%), and SINEs (1.83%). In total, repetitive sequences accounted for 553,425,772 bp (51.80%) of the genome (Table 2, Fig. 4).

Gene prediction and function assignment. Gene prediction was performed through an integrative approach combining transcriptomic evidence, homology-based methods, and ab initio predictions. For transcriptome-guided annotation, Illumina RNA-seq reads were aligned to the genome using HISAT2 v2.2.1[35], followed by transcript assembly with StringTie v2.2.3[36]. Transcript models were further refined using PASA v2.3.3[37], and coding sequences were predicted with TransDecoder v5.7.1[38]. Homology-based predictions were generated by aligning protein sequences from five evolutionarily related teleost species

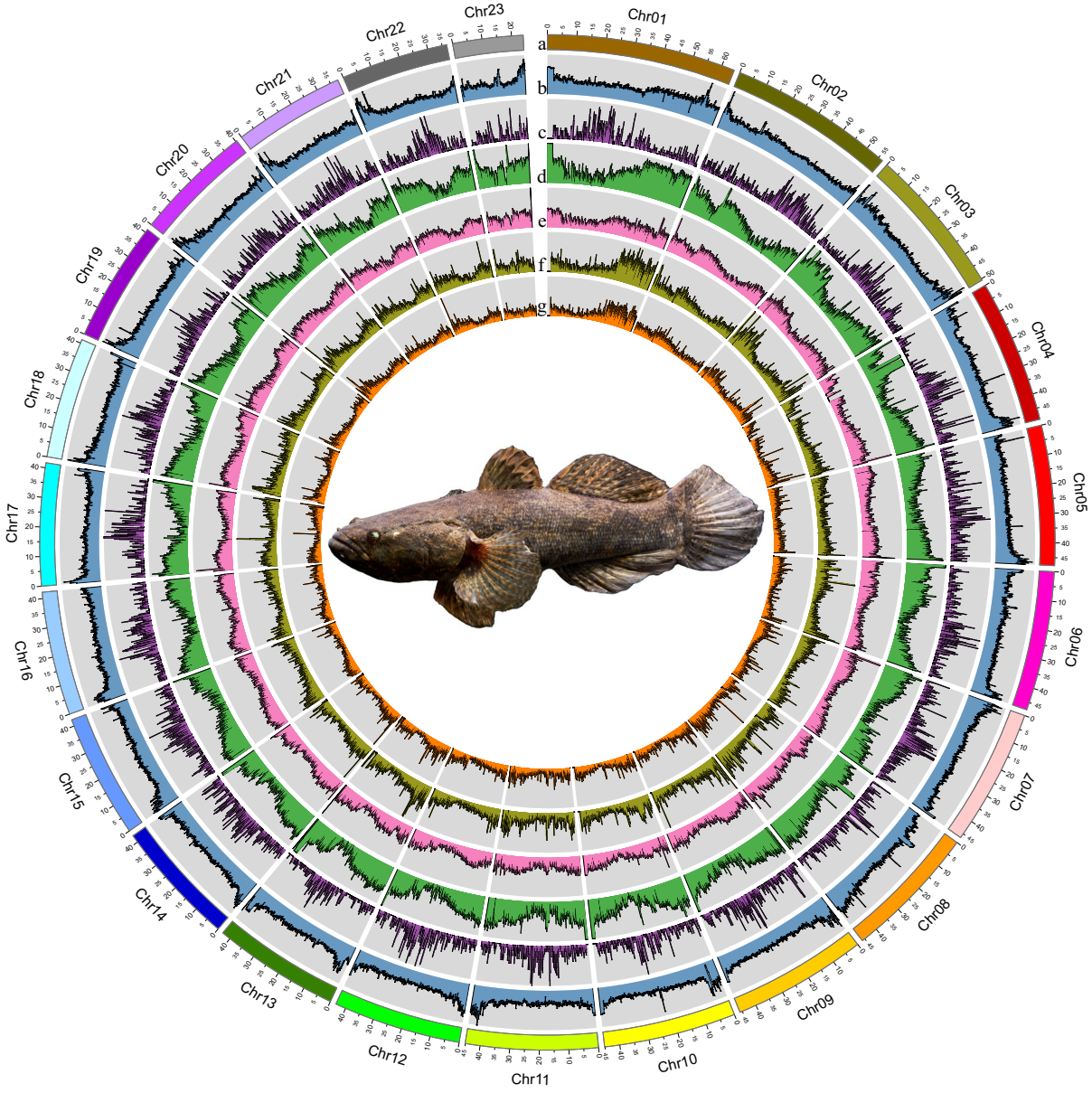

Fig. 3 Features of the *Oxyeleotris marmorata* genome. From outside to inside: (**a**) The 23 pseudo-chromosomes, (**b**) GC content, (**c**) Gene density, (**d**) Repeats content, (**e**) LTR content, (**f**) LINE content, (**g**) DNA-TE content.

Type		Number	Length	% of the genome
Transposable elements	LINE	527,519	109,974,357	10.29
	SINE	218,628	19,556,198	1.83
	LTR	387,914	59,656,698	5.58
	DNA	1,659,577	238,273,688	22.30
	RC	70,427	15,313,100	1.43
	Total	2,864,065	442,774,041	41.44
Tandem Repeats		71,484	12,643,223	1.18
Simple repeats		48,436	7,911,036	0.74
Other		16,091	5,701,831	0.53
Unknown		428,682	83,265,206	7.79
Low complexity		730	179,763	0.02
Total Repeats		3,508,811	553,425,772	51.80

Table 2. Repetitive sequences in the genome of *Oxyeleotris marmorata*.

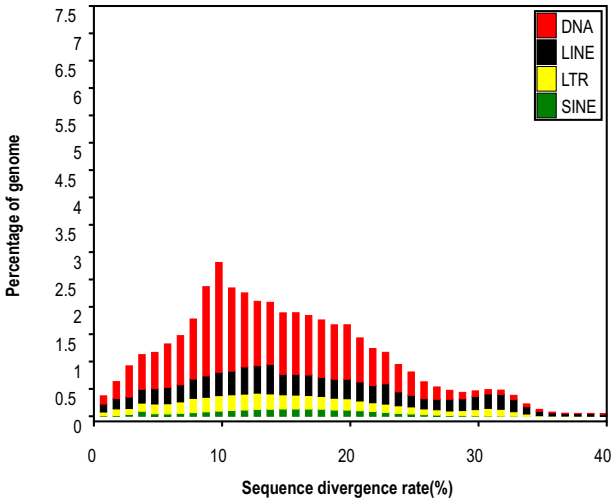

Fig. 4 Four types of TE sequence divergence distribution diagram annotated by RepeatMasker.

Item	Number	Average length (bp)
Gene	22,964	18,806
Exon	9.97 (average)	176
Intron	8.97 (average)	1,900
Database	**Number**	**Percentage (%)**
Swissprot	21,533	93.77
KEGG	14,917	64.96
KOG	15,565	67.78
GO	15,602	67.94
NR	22,624	98.52
ALL	22,791	99.25

Table 3. Gene structures and function annotation.

(*Boleophthalmus pectinirostris*[39], *Epinephelus lanceolatus*[40], *Lates calcarifer*[41], *Danio rerio*[42], and *Oryzias latipes*[43]) using miniprot[44] with stringent alignment criteria. For *ab initio* gene prediction, we used BRAKER v2.1.6[45], integrating Augustus v3.5.0[46] and GeneMark-ETP v1.31[47], trained on reference proteins from the OrthoDB v12[48] database. The three evidence sources were integrated using EVidenceModeler v2.1.0[49], resulting in a final consensus set of 22,964 high-confidence protein-coding genes, with an average gene length of 18,806 bp, an average coding sequence length of 1,755 bp, and an average of 9.97 exons per gene (Table 3).

Functional annotation of the predicted protein sequences was conducted using Diamond v2.1.10[50] against the SwissProt[51], KEGG[52], KOG[53], GO[54], and NR[55] databases, with an e-value cutoff of 1e-5. A total of 22,791 genes were annotated, representing 99.25% of all inferred genes (Fig. 5 and Table 3).

Data Records

The raw sequencing reads of all libraries have been deposited into NCBI SRA database via the accession number PRJNA1115919[56]. The assembled genome has been deposited at Genbank under the accession number GCA_040285405.1[57]. Moreover, data of the genome annotations, predicted coding sequences and protein sequences are available at Figshare[58].

Technical Validation

Assessment of genome assembly. The accuracy of the marble goby genome assembly was evaluated by assessing its completeness using the conserved metazoan gene set 'actinopterygii_odb10' from BUSCO (v5.4.3)[59]. The analysis demonstrated high completeness, with an overall completeness of 98.19%. Specifically, 97.14% of the genes were complete and single-copy, 1.04% were complete and duplicated, 0.71% were fragmented, and 1.10% were missing. These findings indicate the high quality of the marble goby genome assembly (Table 4). Next, the assembly accuracy was assessed by calculating the mapping rates of PacBio, Illumina, and RNA-seq reads aligned to the final assembly using Minimap2 v2.24[60], BWA v0.7.17[27] and HISAT2 v2.2.1[35]. The assembly exhibited high mapping rates for PacBio (99.85%), Illumina (99.13%), and RNA-seq (95.38%) reads. These comprehensive analyses validate the high quality of our genome assembly.

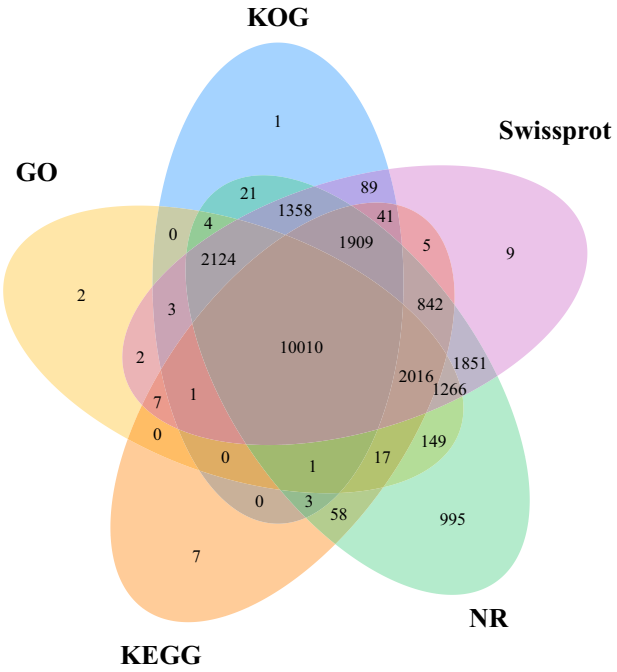

Fig. 5 Venn diagram of function annotations from various databases.

Type	Number	Percent (%)
Complete BUSCOs (C)	3574	98.19
Complete and single-copy BUSCOs (S)	3536	97.14
Complete and duplicated BUSCOs (D)	38	1.04
Fragmented BUSCOs (F)	26	0.71
Missing BUSCOs (M)	40	1.10
Total BUSCO groups searched	3,640	100.00

Table 4. BUSCO analysis of the genome assembly and genes.

Code availability

No special codes or scripts were used in this work, and Data processing was carried out based on the protocols and manuals of the corresponding bioinformatics software.

Received: 16 December 2024; Accepted: 26 August 2025;
Published online: 30 September 2025

References

1. Froese, R. & Pauly, D. (Fisheries Centre, University of British Columbia Los Baños, Philippines, 2010).
2. Xu, Y., Hu, Y., Bao, B. & Gong, X. Complete mitochondrial DNA sequence of marble goby, Oxyeleotris marmorata (Bleeker, 1852). *Mitochondrial DNA Part A* **27**, 817–818 (2016).
3. Hoa, N. P. & Yi, Y. Prey ingestion and live food selectivity of marble goby (Oxyeleotris marmorata) using rice field prawn (Macrobrachium lanchesteri) as prey. *Aquaculture* **273**, 443–448 (2007).
4. Xuegeng, W., Jie, Y. U., Yutao, M., Kangsen, M. A. I. & Lei, W. Research progress and prospect of marble goby (Oxyeleotris marmorata). *Journal of fisheries of china* **47**, 069601–069601 (2023).
5. Lim, L.-S., Yee, C.-W., Tan, K. A., Liew, H. J. & Mukai, Y. Sensory-mediated feeding behaviour in the larvae of marble goby (Oxyeleotris marmorata). *Behavioural Processes* **213**, 104969 (2023).
6. Lam, S. S., Ambak, M. A., Jusoh, A. & Law, A. T. Waste excretion of marble goby (Oxyeleotris marmorata Bleeker) fed with different diets. *Aquaculture* **274**, 49–56 (2008).
7. Luong, V. C., Yi, Y. & Lin, C. K. Cove culture of marble goby (Oxyeleotris marmorata Bleeker) and carps in Tri An Reservoir of Vietnam. *Aquaculture* **244**, 97–107 (2005).
8. Darwis, M., Shaleh, S. R. M. & Senoo, S. Daily food intake, feeding activity and growth of marble goby, Oxyeleotris marmoratus juveniles reared under different salinity levels. *Aquaculture Science* **57**, 185–191 (2009).
9. Tan, O. K. K. & Lam, T. J. Induced breeding and early development of the marble goby (Oxyeleotris marmorata, Blk.). *Aquaculture* **2**, 411–423 (1973).
10. Yang, Z.-Y., Liang, H.-W., Li, Z., Wang, D. & Zou, G.-W. Mitochondrial genome of the Marbled goby (Oxyeleotris marmorata). *Mitochondrial DNA Part A* **27**, 1073–1074 (2016).
11. Ha, H. C. *et al.* Population structure of marble goby Oxyeleotris marmorata (Bleeker) in Southeast Asia inferred from mitochondrial DNA. *Aquaculture Science* **59**, 383–391 (2011).

12. Syaifudin, M., Jubaedah, D., Taqwa, F. H. & Octaviani, R. Phylogenetic of marble goby (Oxyeleotris marmorata Blkr.) in South Sumatra Based on cytochrome C oxidase subunit I (COI) gene. *Genetics of Aquatic Organisms* **6** (2021).
13. Zhao, C. *et al*. Population genetic diversity of marble goby (Oxyeleotris marmoratus) inferred from mitochondrial DNA and microsatellite analysis. *Journal of Genetics* **96**, 65–71 (2017).
14. Luo, J., Zhu, X. P., Peng, Y. H. & Yin, S. W. Isolation and characterization of 16 microsatellite loci in marble goby (Oxyeleotris marmoratus). *Genetics and Molecular Research* **12**, 2020–2023 (2013).
15. Liu, W. *et al*. A novel germline and somatic cell expression of two sexual differentiation genes, Dmrt1 and Foxl2 in marbled goby (Oxyeleotris marmorata). *Aquaculture* **516**, 734619 (2020).
16. Pinthong, K. *et al*. A first karyological analysis of the sand goby, Oxyeleotris marmoratus (Teleostei, Eleotridae) in Thailand by Ag-NOR Staining Technique. *Cytologia* **79**, 127–132 (2014).
17. Law, D. *et al*. Use of Amplified Fragment Length Polymorphism and Sequence Characterized Amplified Region Marker for Identifying the Sex of the Oxyeleotris marmorata. *Pertanika Journal of Tropical Agricultural Science* **44** (2021).
18. Liu, W. *et al*. Molecular characterization of vasa homologue in marbled goby, Oxyeleotris marmorata: Transcription and localization analysis during gametogenesis and embryogenesis. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **229**, 42–50 (2019).
19. Carminati, M.-V. *et al*. Novel Megaptera novaeangliae (Humpback whale) haplotype chromosome-level reference genome. *Scientific Data* **11**, 1113 (2024).
20. Li, R., Leiva, C., Lemer, S., Kirkendale, L. & Li, J. Photosymbiosis shaped animal genome architecture and gene evolution as revealed in giant clams. *Communications Biology* **8**, 7 (2025).
21. Sun, J. *et al*. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nat Commun* **11**, 1657 (2020).
22. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
23. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432 (2020).
24. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
25. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems* **3**, 95–98 (2016).
26. Dudchenko, O. *et al*. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754–1760 (2009).
28. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3**, 99–101 (2016).
29. Wang, X. & Wang, L. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in plant science* **7**, 215951 (2016).
30. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).
31. Han, Y. & Wessler, S. R. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res* **38**, e199–e199 (2010).
32. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).
33. Jurka, J. *et al*. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* **110**, 462–467 (2005).
34. Hubley, R. *et al*. The Dfam database of repetitive DNA families. *Nucleic Acids Res* **44**, D81–D89 (2016).
35. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907–915 (2019).
36. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290–295 (2015).
37. Haas, B. J. *et al*. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666 (2003).
38. Haas, B. J. *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494–1512 (2013).
39. Bian, C. *et al*. Genomics comparisons of three chromosome-level mudskipper genome assemblies reveal molecular clues for water-to-land evolution and adaptation. *Journal of Advanced Research* **58**, 93–104 (2024).
40. Zhou, Q. *et al*. A chromosome-level genome assembly of the giant grouper (Epinephelus lanceolatus) provides insights into its innate immunity and rapid growth. *Mol Ecol Resour* **19**, 1322–1332 (2019).
41. Vij, S. *et al*. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *Plos Genet* **12**, e1005954 (2016).
42. Howe, K. *et al*. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
43. Dong, Z. *et al*. A high-quality chromosome-level genome assembly of the Chinese medaka Oryzias sinensis. *Scientific Data* **11**, 322 (2024).
44. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014 (2023).
45. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* **3**, lqaa108 (2021).
46. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**, W435–W439 (2006).
47. Bruna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP: automatic gene finding in eukaryotic genomes in consistency with extrinsic data. *BioRxiv*, 2023-2001 (2023).
48. Kuznetsov, D. *et al*. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res* **51**, D445–D451 (2023).
49. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* **9**, 1–22 (2008).
50. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021).
51. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**, 45–48 (2000).
52. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
53. Tatusov, R. L. *et al*. The COG database: an updated version includes eukaryotes. *BMC bioinformatics* **4**, 1–14 (2003).
54. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000).
55. Kanz, C. *et al*. The EMBL nucleotide sequence database. *Nucleic Acids Res* **33**, D29–D33 (2005).
56. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRP511840 (2024).
57. *NCBI GenBank*. https://identifiers.org/ncbi/insdc.gca:GCA_040285405.1 (2024).
58. Liu, H. Chromosome-level genome assembly of the marble goby (*Oxyeleotris marmorata*). *Figshare*. https://doi.org/10.6084/m9.figshare.28006505 (2024).

59. 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* **31**, 3210–3212 (2015).
60. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

Acknowledgements

This work is supported by the National Key R&D Program of China (2022YFD2401600), National Natural Science Foundation of China (32273115, 32473153); China Agriculture Research System of MOF and MARA (CARS-46); China-ASEAN Maritime Cooperation Fund (CAMC-2018F); Central Public-interest Scientific Institution Basal Research Fund, CAFS (2023TD37); Guangdong Province Rural Revitalization Strategy Special Fund (2022-SPY-00-016); National Freshwater Genetic Resource Center (FGRC18537).

Author contributions

All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-05878-w.

Correspondence and requests for materials should be addressed to H.L. or W.L.