



OPEN

DATA DESCRIPTOR

Gap-free telomere-to-telomere genome assembly of marbled flounder (*Pseudopleuronectes yokohamae*)

Weiwei Zheng^{1,2,4}, Aijun Cui^{1,2,4}, Zhen Meng^{1,2,4}, Ning Zhang³, Yan Jiang^{1,2}, Jichang Zheng^{1,2}, Rixiang Zhao³, Wenteng Xu^{1,2}✉ & Yongjiang Xu^{1,2}✉

Marbled flounder (*Pseudopleuronectes yokohamae*) is becoming a commercially important flatfish species in Northeast Asia due to its strong environmental adaptability and excellent nutritional value. However, no high-quality marbled flounder reference genome was reported to date, which greatly limits the studies of evolutionary and functional genomics. Here, we reported the first gap-free T2T genome in flatfish (marbled flounder), with length of 582.73 Mb (contig N50: 26.29 Mb) combining short reads, PacBio HiFi long reads, ONT ultra-long reads, and Hi-C data. All of the genome sequences were assembled onto 24 chromosomes, and 48 telomeres were identified on both ends of all chromosomes. 99.29% complete BUSCOs were identified, demonstrating a high level of completeness. The average mapping ratio of short reads, PacBio HiFi reads, and ONT ultra-long reads aligned to the genome was more than 99.89%. 121.02 Mb repeating elements and 22,778 protein-coding genes were identified in the genome assembly. These results provide valuable resources for the evolutionary genomics research and the identification of key candidate genes for economic traits in marbled flounder.

Background & Summary

Marbled flounder, *Pseudopleuronectes yokohamae* (FishBase ID: 1877), belonging to Pleuronectiformes, Pleuronectidae, is widely distributed in Northwest Pacific, from southern Hokkaido of Japan to the Yellow Sea, the Gulf of Bo Hai, and the northern part of the East China Sea, including Korea^{1–4}. It was once an important fishery resource, but due to overfishing, environmental pollution, and habitat degradation, its annual catch has decreased substantially from more than 500 t in 1980s to around 50 t in the 2000s^{5–7}. In recent years, it has become an economically important aquaculture species in South Korea, Japan and north of China, because of its tolerance to low temperature, strong adaptability, delicate taste, and high nutritional value^{8,9}.

Extensive studies of marbled flounder have been conducted, including life history and ecology^{4,10}, nutrition⁸, growth⁵, reproductive biology^{4,11}, genetic diversity and population structure^{9,12–14}, sperm cryopreservation¹⁵, and stress response^{16,17}. However, the lack of high-quality reference genome has greatly hindered the comprehensive genetic and functional genomic research. To date, only a contig-level marbled flounder genome with low continuity and completeness (contig N50 length of 2 kb assembled by Illumina sequencing) was reported¹⁸. Therefore, there is an urgent need to obtain a high-quality marbled flounder genome.

The rapid development of sequencing technologies and assembly algorithms, especially long-read sequencing technologies like PacBio high-fidelity (HiFi), and Oxford Nanopore Technologies (ONT) ultra-long sequencing, has made it possible to obtain high-quality telomere-to-telomere (T2T) genome. Several near-T2T or T2T genome assemblies have been reported in fish, such as Chinese sea bass (*Lateolabrax maculatus*)¹⁹, the giant grouper (*Epinephelus lanceolatus*)²⁰, the tomato hind (*Cephalopholis sonnerati*)²¹, African catfish

¹State Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, Shandong, 266071, China. ²Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao Marine Science and Technology Center, Qingdao, Shandong, 266237, China. ³Yantai Zongzhe Marine Science & Technology Co., Ltd., Yantai, Shandong, 265617, China. ⁴These authors contributed equally: Weiwei Zheng, Aijun Cui, Zhen Meng. ✉e-mail: xuwt@ysfri.ac.cn; xuyj@ysfri.ac.cn

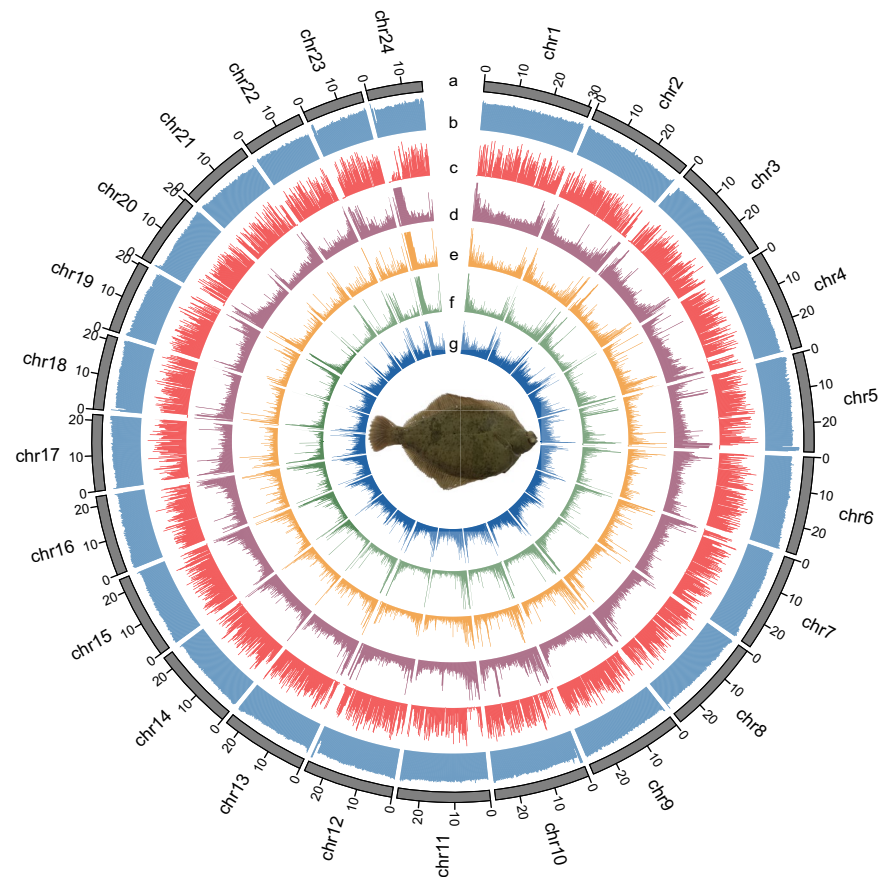


Fig. 1 Features of marbled flounder T2T genome assembly. Tracks from outer to inner represent the chromosomes (a), GC content (b), distribution of gene density (c), distribution of the total repetitive sequences density (d), distribution of LTRs (e), distribution of LINEs (f), distribution of DNE TEs (g).

Library Type	Sequencing Platform	Total Bases (Gb)	Reads Mean Length (bp)	Reads N50 (bp)	Q30 rate (%)
WGS short reads	DNBSEQ-T7	73.96	150	—	97.92
Hi-C	DNBSEQ-T7	58.98	150	—	92.76
Pacbio	Pacbio Revio	31.47	17,407	17,505	—
ONT	PromethION	25.77	101,848	100,000	—
RNA-seq	DNBSEQ-T7	16.53	150	—	97.57

Table 1. Statistics of the sequencing data.

(*Clarias gariepinus*)²², Yadong trout (*Salmo trutta*)²³, Asian icefish (*Protosalanx chinensis*)²⁴, and zig-zag eel (*Mastacembelus armatus*)²⁵.

In the present study, we have assembled a gap-free T2T genome of marbled flounder combining whole genome short-read sequencing, PacBio HiFi sequencing, ONT ultralong sequencing, and Hi-C sequencing data. To our knowledge, this is the first gap-free T2T genome reported in flatfish (Fig. 1). This gap-free T2T genome assembly not only provides a robust foundation for identifying the key candidate genes of economic traits, but also advances the future evolutionary genomics within the family Pleuronectidae.

Methods

Sample collection, library construction and sequencing. A two-year-old female marbled flounder was collected for sampling from Zongzhe Marine Technology Co., Ltd, Yantai, Shandong, China. Fresh muscle tissues were sampled and stored at −80 °C for Genomic DNA (gDNA) extraction. Ten different tissues, including muscle, brain, pituitary, gill, liver, spleen, kidney, intestine, skin, and gonad, were collected for RNA extraction. For whole genome short-read sequencing and Hi-C sequencing, the gDNA was extracted using the TIANamp Marine Animals DNA Kit (Tiangen, China). Furthermore, for ONT ultra-long sequencing and PacBio HiFi sequencing, a standard sodium dodecyl sulfate (SDS) extraction method was used for the gDNA isolation. For RNA-seq sequencing, RNA was extracted using RNA Easy Fast Tissue/Cell Kit (Tiangen, China). The quality and quantification of gDNA and RNA was determined using agarose gel electrophoresis, NanoDrop One

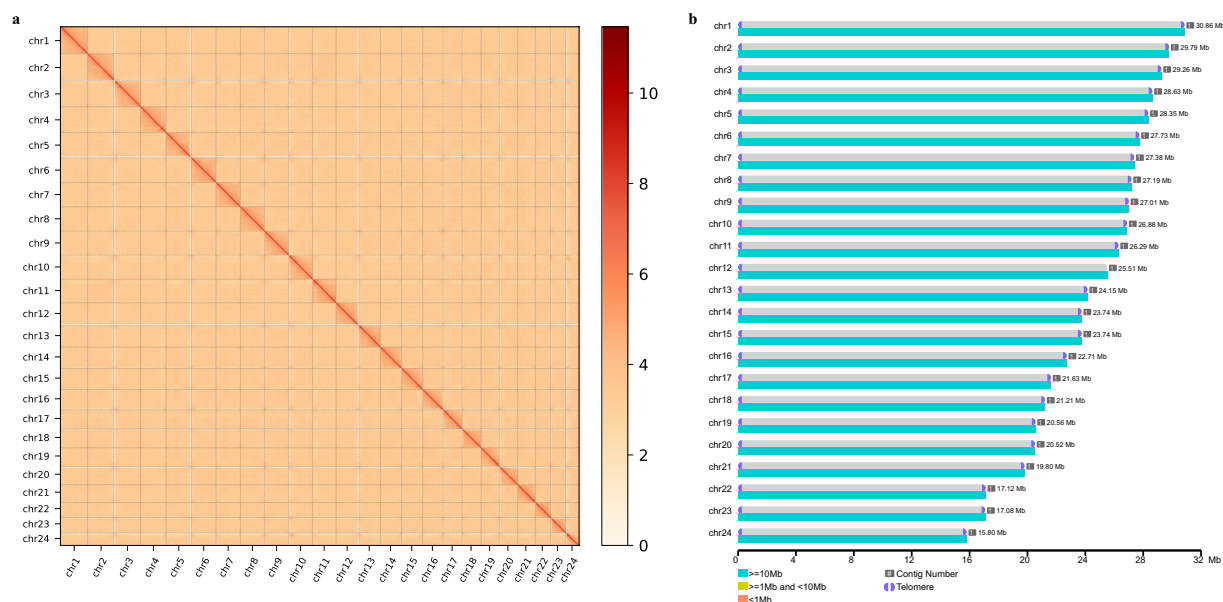


Fig. 2 Hi-C heatmap and T2T genome display map of marbled flounder. **(a)** The Hi-C heatmap of chromosome interactions in marbled flounder. **(b)** Distribution map of telomeres and centromeres on the T2T marbled flounder genome. Purple semicircles indicate predicted telomeres.

spectrophotometer (Thermo Fisher Scientific, USA), and Qubit 3.0 Fluorometer (Life Technologies, USA) to meet the requirements for library construction.

For short-read sequencing, a 200–400 bp short-read library was constructed using MGIEasy Universal DNA Library Prep Kit V1.0 (MGI, China) following the standard protocol, and the qualified library was sequenced on the DNBSEQ-T7RS platform (MGI, China). As a result, a total of 73.96 Gb (128×) of short-read sequencing data was generated (Table 1). For PacBio HiFi sequencing, a SMRTbell library was prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, CA, USA) according to the manufacturer's instruction, and the qualified library then was performed on the PacBio Revio platform (Pacific Biosciences, CA, USA) for sequencing with the CCS mode. Finally, we obtained 31.47 Gb (54×) of PacBio HiFi sequencing data with an average and N50 read length of 17.41 kb and 17.51 kb, respectively (Table 1). The Hi-C library construction included cross-link with formaldehyde, DNA digestion, end repair and biotin labelling, blunt-end ligation, and DNA purification and shearing. After quality control, the library was implemented on DNBSEQ-T7RS platform for sequencing. In total, 58.98 Gb (102×) of Hi-C sequencing data was obtained (Table 1). For ONT ultra-long sequencing, the library was produced using the Oxford Nanopore SQK-ULK001 ligation kit, and the purified library was subsequently sequenced on a PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK). Finally, 25.77 Gb (44×) of ONT ultralong sequencing data with mean and N50 read length of 101.84 kb and 100.00 kb, respectively (Table 1). To assist in annotating the coding genes, RNA from ten different tissues was mixed into a total RNA for RNA sequencing. A pair-ended sequencing library was constructed and sequenced on DNBSEQ-T7RS platform. A total of 16.53 Gb RNA-seq data were generated (Table 1).

Genome assembly and telomere identification. We first evaluated the genome heterozygosity (0.71%) of marbled flounder using K-mer analysis. The primary contigs of marbled flounder genome were initially assembled using PacBio HiFi reads, ONT reads, and Hi-C data by Hifiasm (v0.19.9)²⁶ with the parameter of “--n-hap 2”. The Purge Haplotigs pipeline (v1.1.3)²⁷ with the parameter “-a 70--depth 250” was used to identify and remove haplotypic duplication of the primary genome. To obtain the chromosome-level genome, Hi-C reads were aligned to the contigs using BWA (v0.7.12)²⁸ and filtered out low quality reads using a HiC-Pro pipeline (v3.1.0)²⁹. After de-duplication, the valid pairs were used to anchor, order, and orient scaffolds onto chromosomes by juicer (v1.6)³⁰ and 3D-DNA pipeline (v180992)³¹. Subsequently, the Hi-C contact map was interactively visualized and manually corrected using Juicebox (v1.91)³² based on the fact that the interaction within chromosomes is greater than that between chromosomes, the interaction at close range is greater than that at long range (Fig. 2a). Then, all ONT ultra-long reads were mapped to the chromosome-level genome using minimap²³ to extend the telomere sequences and were utilized to fill gaps by TGS-GapCloser³⁴ to lengthen the contigs. After telomere extension and gap filling, Pilon (v1.23)³⁵ was used to polish the genome assembly with short-read data. Finally, we obtained a gap-free T2T marbled flounder genome assembly with length of 582.73 Mb (contig N50 length of 26.29 Mb), which significantly improved the completeness and continuity compared to the published contig-level marbled flounder genome (GCA_000787555.1)¹⁸ (contig N50 length of 0.002 Mb) (Table 2). We further identified the occurrences of characteristic sequences (AACCCT/AGGGTT) in the telomere region by searching the whole genome using quartet (v1.2.5)³⁶. As a result, a total of 48 telomeres were screened on both ends of 24 chromosomes (Fig. 2b and Table 3).

	GCA_051314325.1	GCA_000787555.1 ¹⁸
Total genome length (Mb)	582.73	547.80
Total chromosome length (Mb)	582.73	—
Number of chromosome	24	—
Number of contigs	24	525,502
Number of scaffolds	24	—
Contig N50 (Mb)	26.29	0.002
Scaffold N50 (Mb)	26.29	—

Table 2. Statistics of marbled flounder genome assemblies.

Chromosome	Length(bp)	Number of Telomeres	Number of gaps
chr1	30,859,922	1	0
chr2	29,635,121	1	0
chr3	29,259,407	1	0
chr4	28,634,317	1	0
chr5	28,352,015	1	0
chr6	27,732,592	1	0
chr7	27,379,010	1	0
chr8	27,124,858	1	0
chr9	27,007,596	1	0
chr10	26,883,915	1	0
chr11	26,290,307	1	0
chr12	25,510,882	1	0
chr13	24,151,384	1	0
chr14	23,743,414	1	0
chr15	23,742,429	1	0
chr16	22,713,765	1	0
chr17	21,626,437	1	0
chr18	21,213,813	1	0
chr19	20,557,026	1	0
chr20	20,517,144	1	0
chr21	19,800,744	1	0
chr22	17,115,885	1	0
chr23	17,079,772	1	0
chr24	15,795,157	1	0

Table 3. Statistics of marbled flounder chromosomes and telomeres.

Repeat annotation. Repetitive sequences were predicted by a strategy of combining *de novo* prediction and homologous alignments. *de novo* repeat elements were predicted using RepeatMasker (v4.0.5)³⁷ based on the *de novo* database generated by RepeatModeler (v1.0.8)³⁸ and LTR-FINDER (v1.0.6)³⁹. Homologue prediction was performed using RepeatMasker (v4.0.5) and RepeatProteinMask (v4.0.5) based on the Repbase library (v21.01)⁴⁰. Furthermore, Tandem Repeats Finder (v4.0.7)⁴¹ was used to detect the tandem repeats. Finally, a total of 121.02 Mb of non-redundant repetitive sequences representing approximately 20.77% of the marbled flounder genome were predicted (Fig. 1 and Table 4). Among these repeats, DNA transposons (9.49%) were the most abundant, followed by long interspersed elements (LINEs, 3.42%), long terminal repeats (LTRs, 3.21%), and short interspersed nuclear elements (SINEs, 0.65%) (Fig. 1 and Table 4).

Protein-coding gene prediction and functional annotation. Gene structure prediction was performed through a combination of *de novo* prediction, homology-based prediction, and transcriptome-based prediction. For *de novo* prediction of the gene structures, GENSCAN (v1.0)⁴² and AUGUSTUS (v3.3.2)⁴³ were used with default settings. For homology-based prediction, the protein sequences from 5 closely related representative fish species, including *Platichthys flesus*, *Pleuronectes platessa*, *Hippoglossus hippoglossus*, *Limanda limanda*, and *H. stenolepis*, were downloaded from NCBI and were then aligned to the marbled flounder T2T genome assembly to predict gene structure by minipro-t(v0.11-r234)⁴⁴ and Liftoff (v1.6.3)⁴⁵ based on the homologous evidences. Moreover, RNA-seq data (accession number: SRR33264204) from ten mixed tissues were aligned to the starry flounder T2T genome using Hisat2 (v2.1.0)⁴⁶ for transcriptome-based prediction. The mapped reads were assembled into transcripts using StringTie (v2.1.4)⁴⁷. Subsequently, the merged transcripts were subjected to TransDecoder (v5.7.0,

Type	Rebase TEs		TE proteins		De novo		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	34,032,084	5.84	3,906,782	0.67	31,573,748	5.42	55,324,428	9.49
LINE	13,059,971	2.24	6,635,836	1.14	11,490,591	1.97	19,928,938	3.42
SINE	2,523,779	0.43	0	0.00	1,875,205	0.32	3,801,300	0.65
LTR	10,345,205	1.77	3,310,206	0.57	9,552,529	1.64	18,701,028	3.21
Satellite	3,122,222	0.54	0	0.00	3,276,266	0.56	6,312,932	1.08
Simple_repeat	0	0.00	0	0.00	0	0.00	0	0.00
Other	2,117	0.00	0	0.00	0	0.00	2,117	0.00
Unknown	613,113	0.11	10,068	0.00	26,311,504	4.52	26,745,069	4.59
Total	56,542,486	9.70	13,853,832	2.38	82,560,658	14.17	121,017,057	20.77

Table 4. Statistics of repetitive sequences in marbled flounder.

	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	29,396	13,873	1,616	9.09	177.74	1,515
	AUGUSTUS	30,204	8,873	1,313	7.33	179.17	1,195
Homolog	<i>Hippoglossus hippoglossus</i>	50,947	18,947	2,239	13.10	170.92	1,381
	<i>Platichthys flesus</i>	42,022	14,282	1,908	11.00	173.52	1,238
	<i>Limanda limanda</i>	34,961	13,246	1,784	9.91	179.88	1,286
	<i>Hippoglossus stenolepis</i>	45,454	18,530	2,251	13.08	172.12	1,348
	<i>Pleuronectes platessa</i>	42,610	13,994	1,927	10.59	181.87	1,258
Liftoff	<i>Platichthys flesus</i>	21,814	12,650	1,817	10.51	172.89	1,139
	<i>Pleuronectes platessa</i>	22,098	12,238	1,797	10.38	173.17	1,114
Trans ORF	RNAseq	12,671	15,241	1,883	12.09	291.07	1,057
MAKER		22,778	14,042	1,704	10.53	267.22	1,178

Table 5. Statistics of predicted protein-coding genes in marbled flounder.

		Number	Percent (%)
Total		22,778	
Annotated		22,386	98.28
	NR	22,996	98.14
	SwissProt	20,426	87.16
	TrEMBL	22,723	96.97
	KOG	18,826	80.34
	TF	5,585	23.83
	InterPro	21,922	93.55
	GO	16,755	71.50
	KEGG_ALL	22,635	96.59
	KEGG_KO	16,354	69.79
	Pfam	20,732	88.47
Unannotated		392	1.72

Table 6. Statistics of functional annotation of protein-coding genes in marbled flounder.

<https://github.com/TransDecoder/TransDecoder>) for protein-coding gene prediction. Finally, the above gene prediction results were integrated by MAKER2 (2.31.10)⁴⁸, producing a non-redundant protein-coding gene set composed of 22,778 genes (Table 5). Comparisons of the statistical characteristics of the gene set elements between marbled flounder and other six related species exhibited a similar distribution patterns (Fig. 3).

Functional annotation of protein-coding genes were performed by matching them with multiple public protein databases, such as GO⁴⁹, KEGG⁵⁰, Pfam⁵¹, UniProt⁵², NR⁵³, and InterPro⁵⁴, using DIAMOND (v2.1.8)⁵⁵ (E-value = 1e−5) or the corresponding inbuilt software. As a result, 22,386 genes, accounting for 98.28% of the predicted genes in marbled flounder, were annotated in at least one functional database (Table 6). In addition, the non-coding RNAs, including 2,703 tRNAs and 860 rRNAs, were annotated using tRNAscan-SE (v2.0.12)⁵⁶ and BLASTN, respectively. 1,278 miRNAs and 1,318 snRNAs were annotated using Infernal (v1.1.4)⁵⁷ based on Rfam database (Table 7).

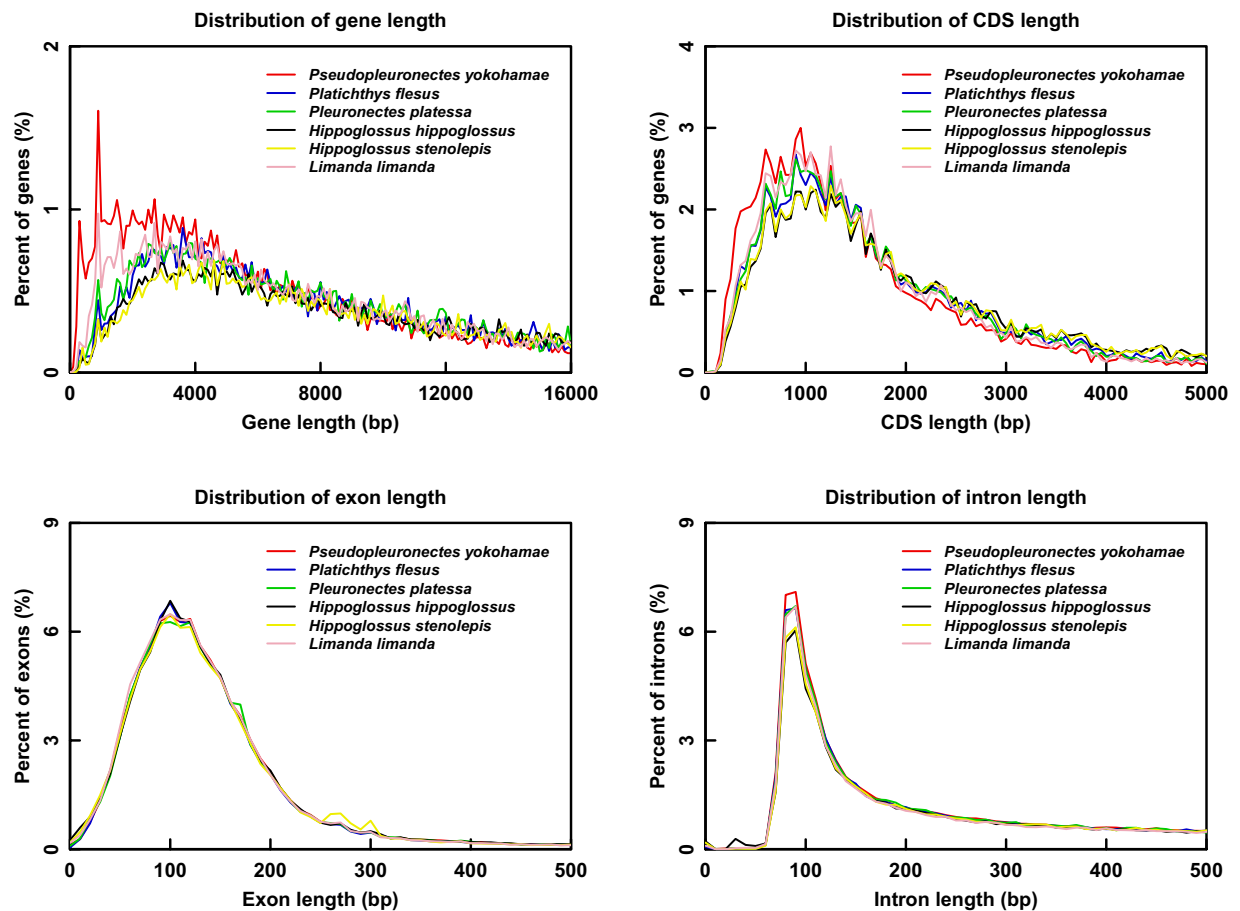


Fig. 3 Comparisons of the statistical characteristics of the gene set elements between marbled flounder, *P. flesus*, *P. platessa*, *H. hippoglossus*, *L. limanda*, and *H. stenolepis*. (a) Gene length distributions. (b) CDS length distributions. (c) Exon length distributions. (d) Intron length distributions.

Gene family and phylogenetic analysis. Gene families of marbled flounder and the other seven species in the Pleuronectidae family, including *H. hippoglossus*, *P. flesus*, *H. stenolepis*, *L. limanda*, *P. platessa*, *Scophthalmus maximus*, and *Paralichthys olivaceus*, were identified and clustered by Orthofinder (v2.5.5)⁵⁸. As a result, a total of 17,403 common families were obtained, which include 196 species-specific gene families (590 genes) identified in marbled flounder (Fig. 4B,C). Subsequently, 12,370 single copy genes were used for multiple sequence alignment using MAFFT (v7.525)⁵⁹. Then, a maximum likelihood (ML) phylogenetic tree was constructed using RAxML (v8.2.12)⁶⁰ and performed 1000 bootstrap replications. Additionally, MCMCtree program implemented in PAML (v4.9)⁶¹ was conducted to estimate divergence times among species. The result showed that marbled flounder began to diverge around 11.5 million years ago (Fig. 5).

Data Records

The WGS short-read sequencing data have been uploaded to NCBI Sequence Read Archive (SRA) database with the accession number SRR33323076⁶². The PacBio HiFi sequencing data have been deposited at NCBI SRA database with the accession number SRR33323077⁶³. The Hi-C sequencing data have been submitted to NCBI SRA database with the accession number SRR33323078⁶⁴. The ONT sequencing data have been deposited into NCBI SRA database with the accession number SRR33323079⁶⁵. RNA-seq data can be obtained under the accession number SRR33264204⁶⁶. The genome assembly has been submitted to the NCBI Genbank with accession number GCA_051314325.1⁶⁷. The genome annotation file, gene CDS, and protein data have been deposited at Figshare⁶⁸.

Technical Validation

Compared to the previously published marbled flounder genome, the contig N50 length of our T2T genome assembly increased from 2.00 kb to 26.29 Mb, exhibiting a substantial enhancement in contiguity of the genome assembly. The completeness of the marbled flounder genome assembly was assessed using BUSCO (v5.1.2) with the actinopterygii_odb10 database containing 3,640 BUSCOs. The results showed that 3,614 (99.29%) complete BUSCOs were identified, which is higher than that of closely related species, such as *Platichthys flesus* (98.2%)⁶⁹ and *Pleuronectes platessa* (97.9%)⁷⁰, indicating a high level of integrity. Moreover, the base-level accuracy and the consensus quality value (QV) was evaluated using Merquy⁷¹, and the average per-base accuracy rates and QV were 0.99986 and 38.52, respectively,

Type		Copy	Average length(bp)	Total length(bp)	% of genome
miRNA		1,278	87	110,867	0.019026
tRNA		2,703	75	202,733	0.034790
rRNA	rRNA	860	121	104,267	0.017893
	18S	2	1,123	2,246	0.000385
	28S	0	0	0	0
	5.8S	1	154	154	0.000026
	5S	857	119	101,867	0.017475
snRNA	snRNA	1,318	147	193,404	0.033177
	CD-box	189	126	23,903	0.0041
	HACA-box	77	153	11,755	0.002016
	splicing	1,042	149	155,764	0.02672
	scaRNA	10	198	1,982	0.00034

Table 7. Statistics of non-coding RNA in marbled flounder.

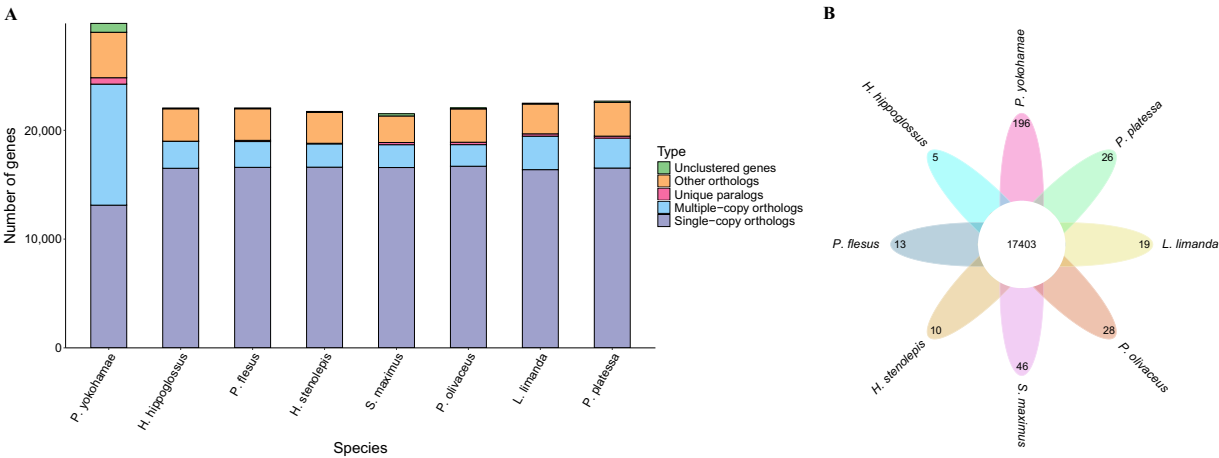


Fig. 4 Comparison of gene families between marbled flounder and *H. hippoglossus*, *P. flesus*, *H. stenolepis*, *L. limanda*, *P. platessa*, *S. maximus*, and *P. olivaceus*. **(A)** Statistics of homologous gene numbers across 8 species. **(B)** Gene family clustering across 8 species.

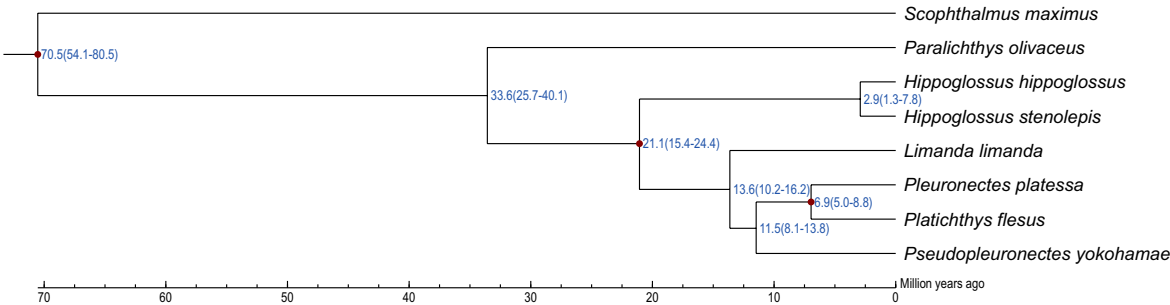


Fig. 5 Phylogenetic relationships and divergence time estimation for marbled flounder and 7 other species in the Pleuronectidae family. *S. maximus* was chosen as an outgroup. All nodes were supported by 100 cycles of bootstrap resampling. Blue numbers represent the estimated divergence times with a 95% confidence interval. Divergences used for the recalibration of time estimation are marked by red dots.

highlighting a high-quality assembly. Furthermore, WGS short reads, PacBio HiFi reads, and ONT ultra-long reads were aligned to the genome assembly using bwa (v0.7.17)²⁸ and minimap2 (v2.17)³³. The average mapping rates and coverage were more than 99.89% and 99.70%, respectively, highlighting a high sequences consistency.

Data availability

The genome sequencing data, including WGS short-read, PacBio HiFi, Hi-C, and ONT ultra-long, can be publicly available from the NCBI SRA database with the accession number [SRR33323076⁶²](#), [SRR33323077⁶³](#), [SRR33323078⁶⁴](#), and [SRR33323079⁶⁵](#). RNA-seq data used for genome annotation also can be publicly obtained from the NCBI SRA database with the accession number [SRR33264204⁶⁶](#). The genome assembly can be publicly available from the NCBI Genbank with accession number [GCA_051314325.1⁶⁷](#). The genome annotation file, gene CDS, and protein data can be publicly downloaded at Figshare⁶⁸ (<https://doi.org/10.6084/m9.figshare.28823984.v2>).

Code availability

All software and tools were used in this study following the instructions and protocols provided by the respective software developers. The software versions and corresponding parameters applied have been described in the Methods section, and default parameters were used if no parameter was described. No custom code was used in this work.

Received: 30 April 2025; Accepted: 8 September 2025;

Published online: 17 October 2025

References

- Kim, H., Huh, S., Kim, S.-T. & Seo, Y.-I. Age and Growth of Marbled sole, *Pleuronectes yokohamae* (Gunther) in the southern waters of Korea. *Journal of the Korean Society of Fisheries and Ocean Technology* **51**, 35–41 (2015).
- Minami, T. & Tanaka, M. Life History cycles in flatfish from the Northwestern Pacific, with particular reference to their early life histories. *Netherlands Journal of Sea Research* **29**, 35–48 (1992).
- Hata, M., Sugimoto, R., Hori, M., Tomiyama, T. & Shoji, J. Occurrence, distribution and prey items of juvenile marbled sole *Pseudopleuronectes yokohamae* around a submarine groundwater seepage on a tidal flat in southwestern Japan. *Journal of Sea Research* **111**, 47–53 (2016).
- Kume, G. *et al.* Seasonal distribution, age, growth, and reproductive biology of marbled sole *Pleuronectes yokohamae* in Tokyo Bay, Japan. *Fisheries Science* **72**, 289–298 (2006).
- Lee, J.-H. *et al.* Changes in growth of marbled sole *Pseudopleuronectes yokohamae* between high and low stock-size periods in Tokyo Bay, Japan. *Fisheries Science* **75**, 929–935 (2009).
- Lee, J.-H. *et al.* Comparison between surface-reading and cross-section methods using sagittal otolith for age determination of the marbled sole *Pseudopleuronectes yokohamae*. *Fisheries Science* **75**, 379–385 (2009).
- Yamamoto, S., Mitamura, H., Obitsu, N. & Sato, M. Residence and movement patterns of the marbled flounder *Pseudopleuronectes yokohamae* around a spawning ground in southern Suo Nada, Seto Inland Sea, Japan. *Fisheries Science* **87**, 161–171 (2021).
- Cho, J.-H. *et al.* A preliminary study of dietary protein requirement of juvenile marbled flounder (*Pseudopleuronectes yokohamae*). *Animal Nutrition* **7**, 548–555 (2021).
- Zhang, H. *et al.* Analysis of genetic diversity and population structure of *Pleuronectes yokohamae* indicated by AFLP markers. *Biochem Syst Ecol* **44**, 102–108 (2012).
- Higashitani, T., Takatsu, T., Nakaya, M., Joh, M. & Takahashi, T. Maternal effects and larval survival of marbled sole *Pseudopleuronectes yokohamae*. *Journal of Sea Research* **58**, 78–89 (2007).
- Zheng, F. *et al.* Transcriptome profiling of the hypothalamus and pituitary at gonadal maturation and regression phases in marbled flounder (*Pseudopleuronectes yokohamae*). *Genes & Genomics* **39**, 713–722 (2017).
- Pan, T., Zhang, Y., Gao, T. & Li, F. Genetic diversity of *Pleuronectes yokohamae* population revealed by fluorescence microsatellite labeled. *Biochem Syst Ecol* **55**, 118–124 (2014).
- Sato, M. *et al.* Genetic structure and demographic connectivity of marbled flounder (*Pseudopleuronectes yokohamae*) populations of Tokyo Bay. *Journal of Sea Research* **142**, 79–90 (2018).
- Tsukagoshi, H. *et al.* Genetic variation and population structure of marbled sole *Pleuronectes yokohamae* and crested flounder *P. schrenki* in Japan inferred from mitochondrial DNA analysis. *Biochem Syst Ecol* **58**, 274–280 (2015).
- Hossen, S., Kim, S. C., Cho, Y. & Kho, K. H. Vital Analysis of Cryopreserved Sperm of Marbled Flounder, *Pseudopleuronectes yokohamae*. *Frontiers in Physiology* **12** (2021).
- Cui, Q. *et al.* Transcriptome profiling of the low-salinity stress responses in the gills of the juvenile *Pseudopleuronectes yokohamae*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **32**, 100612 (2019).
- Nishizawa, H., Kono, Y., Arai, N., Shoji, J. & Mitamura, H. Ventilatory and behavioural responses of the marbled sole *Pseudopleuronectes yokohamae* to progressive hypoxia. *Journal of Fish Biology* **90**, 2363–2374 (2017).
- Genomic Resources Development, C. *et al.* Genomic Resources Notes accepted 1 February 2015 – 31 March 2015. *Molecular Ecology Resources* **15**, 1014–1015 (2015).
- Sun, Z. *et al.* Telomere-to-telomere gapless genome assembly of the Chinese sea bass (*Lateolabrax maculatus*). *Scientific Data* **11**, 175 (2024).
- Zhou, Q. *et al.* Telomere-to-telomere gapless genome assembly of the giant grouper (*Epinephelus lanceolatus*). *Scientific Data* **11**, 1342 (2024).
- Lu, S. *et al.* Gap-free telomere-to-telomere haplotype assembly of the tomato hind (*Cephalopholis sonnerati*). *Scientific Data* **11**, 1268 (2024).
- Nguinkal, J. A., Zoclanclounon, Y. A. B., Brunner, R. M., Chen, Y. & Goldammer, T. Haplotype-resolved and near-T2T genome assembly of the African catfish (*Clarias gariepinus*). *Scientific Data* **11**, 1095 (2024).
- Li, C. *et al.* Near complete genome assembly of Yadong trout (*Salmo trutta*). *Scientific Data* **12**, 74 (2025).
- Zhou, Y. *et al.* Telomere-to-telomere genome and resequencing of 254 individuals reveal evolution, genomic footprints in Asian icefish, *Protosalanx chinensis*. *GigaScience* **14**, giae115 (2025).
- Xue, L. *et al.* Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. *Genome Biology* **22**, 203 (2021).
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170–175 (2021).
- Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology* **16**, 259 (2015).
- Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems* **3**, 95–98 (2016).
- Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

32. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99–101 (2016).
33. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
34. Xu, M. *et al.* TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **9**, g1aa094 (2020).
35. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *Plos One* **9**, e112963 (2014).
36. Lin, Y. *et al.* quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic Res-England* **10**, uhad127 (2023).
37. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.11–14.10.14 (2009).
38. 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).
39. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* **35**, W265–W268 (2007).
40. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
41. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573–580 (1999).
42. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* **268**, 78–94 (1997).
43. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**, W435–W439 (2006).
44. Li, H. Protein-to-genome alignment with minimap. *Bioinformatics* **39**, btad014 (2023).
45. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
46. 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).
47. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11**, 1650–1667 (2016).
48. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
49. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000).
50. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30 (2000).
51. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2020).
52. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45**, D158–D169 (2017).
53. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **35**, D61–D65 (2007).
54. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research* **49**, D344–D354 (2021).
55. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**, 59–60 (2015).
56. Chan, P. P. & Lowe, T. M. J. o. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods in Molecular Biology* **1962**, 1–14 (2019).
57. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
58. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
59. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).
60. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
61. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**, 1586–1591 (2007).
62. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33323076> (2025).
63. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33323077> (2025).
64. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33323078> (2025).
65. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33323079> (2025).
66. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33264204> (2025).
67. Zheng, W. W. & Xu, W. T. NCBI GenBank https://identifiers.org/ncbi/insdc.gca:GCA_051314325.1 (2025).
68. Zheng, W. W. Gap-free telomere-to-telomere genome of marbled flounder (*Pseudopleuronectes yokohamae*). *figshare* <https://doi.org/10.6084/m9.figshare.28823984.v2> (2025).
69. Fletcher, C., Alexander, D. & Reed, B. The genome sequence of the European flounder, *Platichthys flesus* (Linnaeus, 1758). *Wellcome Open Res* **8**, 381 (2023).
70. Brittain, R., Brenen, M., McTierney, S., Adkins, P. & Mrowicki, R. The genome sequence of the European plaice, *Pleuronectes platessa* (Linnaeus, 1758). *Wellcome Open Res* **8**, 367 (2023).
71. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).

Acknowledgements

This work was supported by the National Science and Technology Innovation 2030-Major Program (2023ZD0405505), the project by the National Marine Genetic Resource Center, Ministry of Agriculture and Rural Affairs Special Project - Collection and Preservation of Marine Fishery Biological Resources.

Author contributions

Y.X. and W.X. conceived and designed the project. A.C., Z.M., N.Z. and R.Z. collected the samples for this study. W.Z., A.C. and Y.J. conducted the genome assembly and bioinformatics analysis. W.Z. and W.X. drafted the manuscript. Y.X., A.C., Z.M., Y.J. and J.Z. provided suggestions for manuscript improvement and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.X. or Y.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025