scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly and annotation of *Pampus argenteus* and *Pampus punctatissimus*

Fei Zhu[1,4], Sirong Jiang[2,4], Chaofeng Jia[1], DaFeng Xu[1], Yanyu Zhang[2], Shaoyang Bu[3], Shuran Du[1], Qian Meng[1], Ruijian Sun[1], Gao Bo[1] ✉ & Jian Luo[2,3] ✉

Pampus punctatissimus **and** ***Pampus argenteus*** **are species within the genus** ***Pampus*** **and are renowned for their fast growth and thick, flavorful meat. As highly valued commercial marine fishes of the order Perciformes, they are widely distributed across extensive shallow-water regions, including the Northwest Pacific (off China, Japan, and Korea) and the northern Indian Ocean. In this study, to facilitate further molecular evolutionary research on** ***Pampus*** **species, we constructed chromosome-level genome assemblies for both** ***P. punctatissimus*** **and** ***P. argenteus*****. Our assembly was 38.11 Mb longer than the reference genome (*P. argenteus*) in terms of total length, with greater scaffold N50 values. The BUSCO scores of 99.1 and 98.1 for the** ***P. punctatissimus*** **and** ***P. argenteus*** **genomes revealed the high completeness and accuracy, respectively, of our assembly. The** ***P. punctatissimus*** **and** ***P. argenteus*** **genomes contained 25,552 and 22,284 predicted protein-coding genes, respectively, and 24.34% and 24.88% of the assemblies were annotated as repetitive sequences, respectively. This study substantially advances the available genomic resources for** ***P. punctatissimus*** **and facilitates future research on the adaptation and evolution of this species.**

Background & Summary

P. punctatissimus and *P. argenteus* are commercially important marine fish widely distributed in shallow coastal waters of the Northwest Pacific, including the coasts of China, Japan, and South Korea, as well as the northern Indian Ocean[1]. These fish are highly prized for their rapid growth rate, palatable flesh, and high nutritional content, making them valuable species for commercial capture[2,3]. In China, *P. punctatissimus* and *P. argenteus* play essential roles in coastal fisheries, contributing substantially to local economies[4,5]. According to the China Fisheries Statistical Yearbook 2024, the annual catch of *Pampus* species, including *P. punctatissimus* and *P. argenteus*, reached 340,982 tons in 2023, highlighting its economic relevance and the urgent need for sustainable exploitation.

Despite their economic significance, *P. punctatissimus* and *P. argenteus* face growing ecological threats. Overfishing, habitat destruction, coastal development, and increasing environmental pollution have caused notable declines in wild populations in many regions[6]. These changes threaten the long-term viability of natural stocks and call for urgent conservation efforts[7]. Such pressures underscore the need for conservation, and technological progress over the past two decades has enabled the first successful, reliable large-scale breeding of *P. argenteus*[8]. In contrast, efforts with *P. punctatissimus* remain limited to artificial fertilization trials. Effective management and sustainable utilization of *P. punctatissimus* and *P. argenteus* resources require a clear understanding of the genetic background and population structure of these species. High-resolution genomic information enables the evaluation of intraspecific genetic diversity, supports the identification of adaptive traits, and informs breeding, restocking, and conservation strategies[9].

To provide genomic insights and support further molecular research on *P. punctatissimus and P. argenteus*, we generated high-quality, chromosome-scale genome assemblies using a combination of long-read sequencing,

[1]Jiangsu Marine Fisheries Research Institute, Nantong, 226000, China. [2]Hainan Seed Industry Laboratory, Sanya, 572000, China. [3]School of Marine Biology and Fisheries, Hainan University, Haikou, 570000, China. [4]These authors contributed equally: Fei Zhu, Sirong Jiang. ✉e-mail: 676699683@qq.com; luojian@hainanu.edu.cn

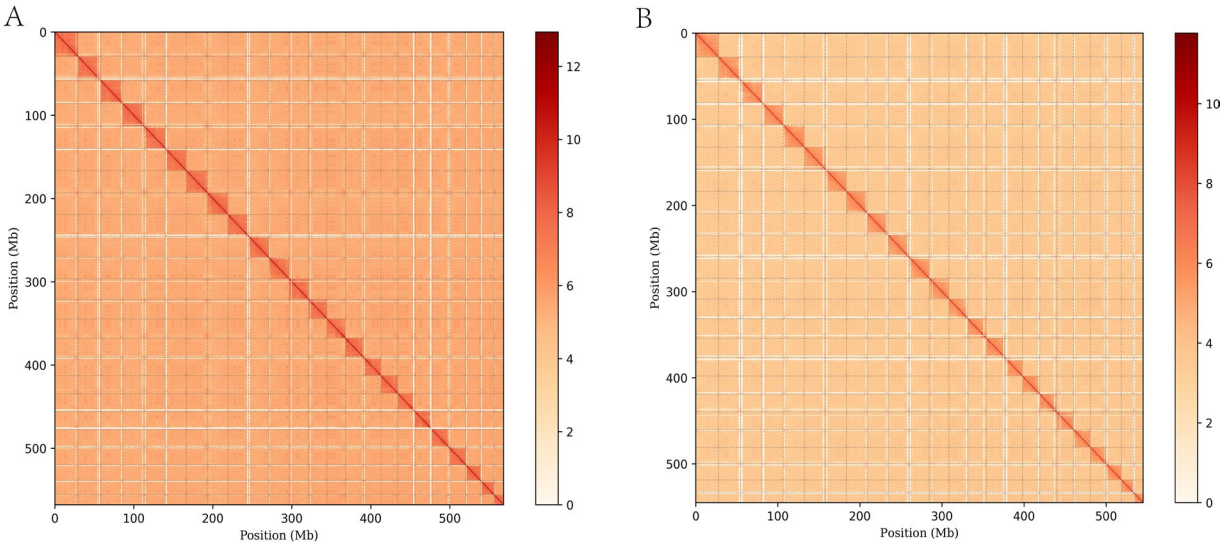

Fig. 1 Genome assemblies of *P. punctatissimus* and *P. argenteus*. (**A**) Hi-C interactive heatmap of the *P. punctatissimus* primary assembly. (**B**) Hi-C interactive heatmap of the *P. argenteus* primary assembly.

short-read polishing, and Hi-C scaffolding technologies. The final assemblies span 572.95 Mb and 556.18 Mb in total length, with scaffold N50 values of 25.56 Mb and 24.89 Mb for *P. punctatissimus and P. argenteus*, respectively, and successfully anchor 99.11% and 98.10% of sequences onto 24 pseudochromosomes, respectively, which is consistent with the karyotypes of related species. Compared with those of the previously published *P. argenteus* genome[10], our assemblies are 38.11 Mb longer and contain fewer scaffolds and higher scaffold N50 values, indicating improved continuity and completeness. Assessment with BUSCO analysis revealed high completeness scores of 99.11% and 98.10%. In total, 27,468 and 32,624 protein-coding genes were predicted, of which substantial proportions were functionally annotated. Repetitive elements accounted for 24.34% and 24.88% of the genome, revealing its structural complexity. These comprehensive genome assemblies represent a critical genomic resource for *P. punctatissimus and P. argenteus*. They will not only facilitate functional genomic and evolutionary studies but also support genetic conservation, molecular breeding, and the sustainable utilization of this ecologically and economically important species.

Methods

Sampling and sequencing. All experimental procedures complied with international animal welfare standards and were approved by the Institutional Animal Care and Use Committee (IACUC) of Jiangsu Provincial Marine Fisheries Research Institute (Protocol No. IACUC 2024-02). *P. punctatissimus* and *P. argenteus* (Fig. 1) were collected from the southern Yellow Sea at 2 months of age and transported to the Rudong Marine Research Station of Jiangsu Marine Fisheries Research Institute. Following a 7-day acclimation period, the fish were reared to 6 months of age under controlled conditions with the following water parameters: a temperature of 25.0 ± 5.0 °C, a pH of 8.4 ± 0.2, a salinity of 26 ± 0.5‰, and a 12 h light:12 h dark cycle (12 L:12 D). All the fish were fed a commercial diet twice daily (08:00 and 15:00) at 3% body weight. All experiments were performed under MS-222 (100 mg/L) anesthesia with efforts to minimize suffering.

Total genomic DNA was extracted from fresh muscle using an animal Genomic DNA Kit (TIANGEN, Beijing, China). The purified DNA was used as input material for construction of the DNA library, which was sequenced using the PacBio Sequel I and Illumina HiSeq. 2,500 platforms at Jiema (Shanghai, China). Total RNA was extracted from the gonads, muscle tissues and intestines of three individuals, and three biological replicates of each tissue were established. RNA sequencing (RNA-seq) was then performed according to the standard process at Jiema (Shanghai, China). Hi-C library construction was performed using the same fresh muscle for DNA extraction. The DNA was cross-linked *in situ*, extracted, and digested with the restriction enzyme DpnII. The resulting fragments were ligated to form chimeric junctions and subsequently purified, after which the Hi-C libraries were amplified using PCR. To enhance and validate the assemblies, we generated 29.56 Gb and 25.54 Gb of PacBio HiFi reads for *P. punctatissimus* and *P. argenteus*, respectively, resulting in sequencing depths of 51.59 × and 45.93 × , respectively. The resulting Hi-C libraries were sequenced using the Illumina NovaSeq X Plus sequencing system with a 150 bp read length. The quality of the Hi-C reads was evaluated using FastQC (v0.12.133)[11]. In total, we generated 121.46 Gb and 80.28 Gb of Hi-C reads for *P. punctatissimus* and *P. argenteus*, respectively[12] (Table s1, Fig. s1).

Genome assembly and quality assessment. Two preliminary contig sequences were generated using Hifiasm (v0.25)[13] by combining PacBio HiFi reads (hifiasm–primary -o p*.asm -t32 HiFi-reads.fq.gz). *P. punctatissimus* and *P. argenteus* genome assemblies of 572.95 Mb and 556.18 Mb, respectively, were obtained. The numbers of contigs were 56 and 68, respectively. The contig N50 values were 24.65 kb and 23.11 kb, respectively (Table 1). High-quality Hi-C data were used to cluster and orient the preliminary contig genome onto the pseudochromosome genome using a 3D-DNA pipeline (180922)[14], and ambiguous fragments were adjusted

	Preliminary version	
	P. punctatissimus	*P. argenteus*
Total length (bp)	572,945,019	556,180,442
GC content (%)	37	36.8
Contig number	56	68
Contig N50 (bp)	24,648,676	23,114,274
Contig N90 (bp)	17,229,324	16,318,311
	Final version	
	P. punctatissimus	*P. argenteus*
Number of chromosome-level scaffolds	24	24
Number of contigs	28	67
Assembly length(bp)	567,850,933	556,145,282
Contig N50(bp)	24,648,676	23,114,274
Scaffold N50 (bp)	25,557,465	24,888,855
Number of Scaffold	24	60
Scaffold anchor ratio(%)	99.11	98.1

Table 1. Major indicators of the *P. punctatissimus* and *P. argenteus* subgenome.

Term	*Pampus punctatissimus*		*Pampus argenteus*	
	BUSCO number	Percentage (%)	BUSCO numbers	Percentage (%)
Total BUSCO groups searched	3,640	100.00%	3,640	100.00%
Complete BUSCOs	3,606	99.07%	3,571	98.10%
Complete and single-copy BUSCOs	3,582	98.41%	3,552	97.60%
Complete and duplicated BUSCOs	24	0.66%	18	0.50%
Fragmented BUSCOs	12	0.33%	24	0.70%
Missing BUSCOs	22	0.60%	46	1.30%

Table 2. BUSCO result of *P. punctatissimus* and *P. argenteus* Genome.

Pampus punctatissimus								
Type	**RepeatMasker TEs Len.(bp)**	**RepeatMasker TEs mean**	**RepeatProteinMask TE Len.(bp)**	**RepeatProteinMask TE mean**	**De novo Len.(bp)**	**De novo Len. Mean**	**Combined TEs Len.(bp)**	**Combined TEs Len. mean**
DNA	12065743	2.11	1106135	0.19	58386137	10.19	66396595	11.59
LINE	7456421	1.3	7384625	1.29	20428599	3.57	23533714	4.11
SINE	72594	0.01	0	0	626050	0.11	674647	0.12
LTR	7286646	1.27	4004102	0.7	25313744	4.42	29149531	5.09
Other	0	0	0	0	0	0	0	0
Unknown	74411	0.01	0	0	23156255	4.04	23230494	4.05
Total TE	24744667	4.32	12490999	2.18	119970586	20.94	132191773	23.07
Pampus argenteus								
Type	**RepeatMasker TEs Len. (bp)**	**RepeatMasker TEs % in genome**	**RepeatProteinMask TEs Len. (bp)**	**RepeatProteinMask TEs % in genome**	**De novo Len. (bp)**	**De novo % in genome**	**Combined TEs Len. (bp)**	**Combined TEs % in genome**
DNA	23,366,966	4.2	1,998,187	0.36	21,734,899	3.91	38,624,412	6.94
LINE	16,305,262	2.93	8,988,738	1.62	15,659,472	2.82	22,063,536	3.97
SINE	493,499	0.09	0	0	1,205,890	0.22	1,593,935	0.29
LTR	9,645,314	1.73	4,077,776	0.73	6,014,718	1.08	12,672,298	2.28
Other	1,876	0	0	0	0	0	1,876	0
Unknown	232,790	0.04	3,819	0	56,329,678	10.13	56,564,591	10.17
Total TE	45,458,527	8.17	15,060,397	2.71	98,842,385	17.77	124,093,139	22.31

Table 3. Repeat sequence annotation information.

manually with Juicebox (v1.1330)[15]. Finally, 24 chromosome-scale scaffolds were obtained for each genome, where 99.11% (567.85 Mb) and 98.10% (556.15 Mb) of the contigs were anchored to the chromosome-scale scaffolds[16,17] (Fig. 1). The scaffold N50 values were 25.56 Mb and 24.89 Mb for each genome (Table 1).

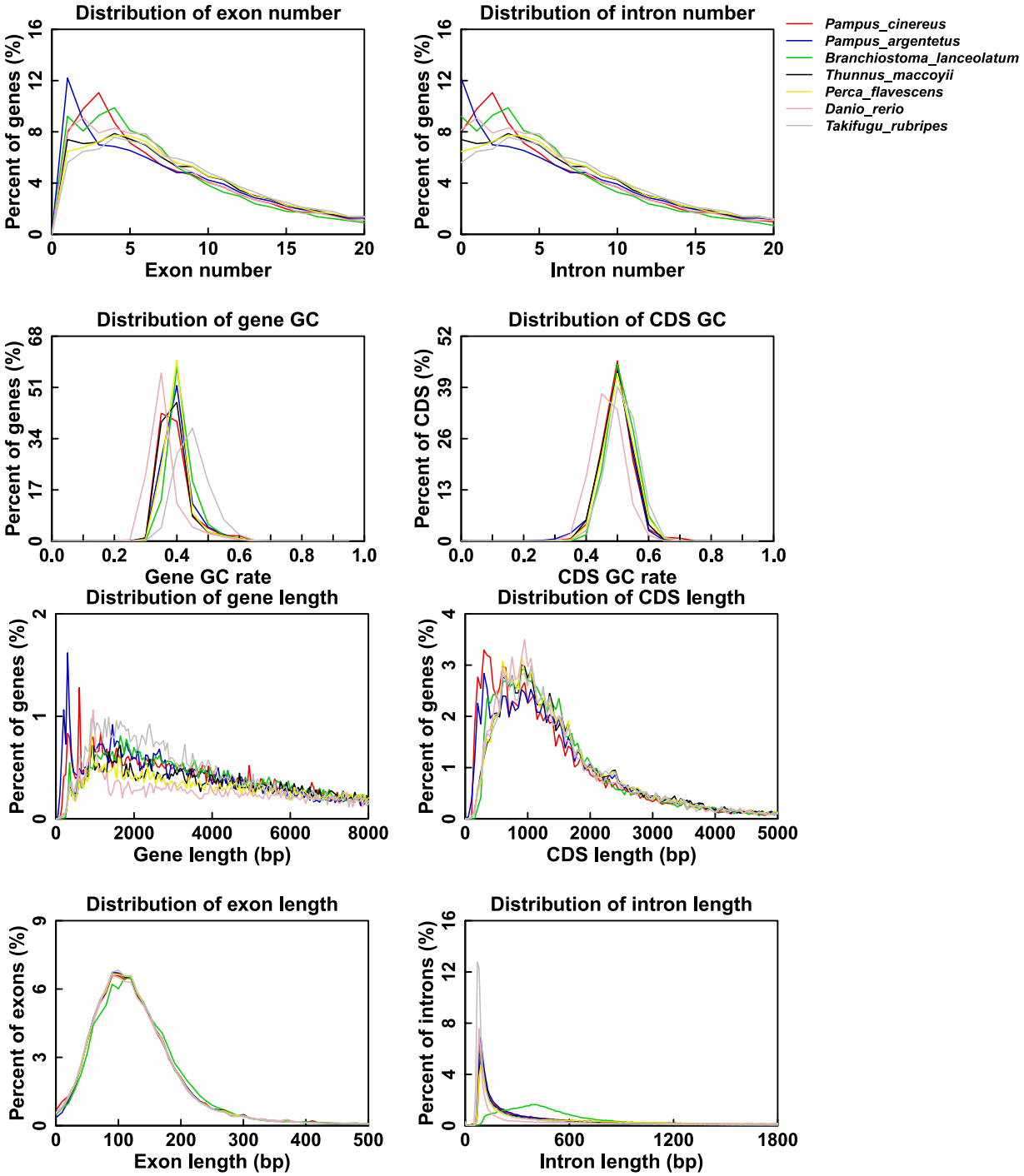

Fig. 2 Comparison of the distribution of gene structure and sequence characteristics among six closely related species (Pampus punctatissimus, Pampus argenteus, Branchiostoma lanceolatum, Thunnus maccoyii, Perca flavescens, Danio rerio, and Takifugu rubripes), including: Distribution of exon number, showing the proportion of the number of gene exons in different species. Distribution of intron number, showing the distribution proportion of the number of gene introns in each species. Distribution of gene GC, reflecting the distribution characteristics of the overall GC content of genes in different species. Distribution of CDS GC, showing the distribution of CDS GC content in each species.Distribution of gene length, showing the proportion distribution of gene lengths in different species. Distribution of CDS length, reflecting the distribution characteristics of CDS lengths in each species. Distribution of exon length, showing the proportion of exon lengths in different species. Distribution of intron length, showing the distribution proportion of intron lengths in each species.

We used two methods to evaluate genome completeness and accuracy: (1) BUSCO[18]. The BUSCO analysis results revealed that 3,606 (99.07%) and 3,571 (98.1%) of the 3,640 complete gene elements in the BUSCO sets were covered by the *P. punctatissimus* and *P. argenteus* genomes, respectively (Table 2). (2) Calculation of the mapping rate

	Pampus punctatissimus		*Pampus argenteus*	
	Annotation No.	**Annotation ratio(%)**	**Annotation No.**	**Annotation ratio(%)**
Complete BUSCOs	3491	95.9	3549	97.5
Complete and single-copy BUSCOs	3452	94.8	3523.52	96.8
Complete and duplicated BUSCOs	39	1.1	25.48	0.7
Fragmented BUSCOs	61	1.7	29.12	0.8
Missing BUSCOs	88	2.4	61.88	1.7
Total BUSCO groups searched	3640	100	3640	100

Table 4. Gene structure annotation busco result.

of the HiFi reads by mapping them to the *P. punctatissimus* and *P. argenteus* genomes through BWA (0.7.17)[19]. The mapping rates of the HiFi data were 99.54% and 99.96%, respectively. The above results revealed that we obtained high-quality reference genomes of *P. punctatissimus* and *P. argenteus* with high completeness and accuracy.

Genome annotation and quality assessment. The repetitive sequences in the *P. punctatissimus* and *P. argenteus* genomes were identified using a combined strategy based on homology prediction and de novo searching methods. First, we used RepeatModeler (v2.0.1)[20] and LTR_Finder (v2.9.0)[21] to construct de novo repeat sequence libraries. Afterward, the de novo libraries and the known Repbase library were merged. Finally, RepeatMasker (4.1.2-p1)[22] was used to search for repetitive sequences in the *P. punctatissimus* and *P. argenteus* genomes.

A total of 139.48 Mb (24.34%) and 138.36 Mb (24.88%) of repetitive sequences were identified in the *P. punctatissimus* and *P. argenteus* genomes, respectively. DNA transposons were the main type, accounting for 48.76% and 51.4%, respectively. LTR (long terminal repeat) retrotransposons accounted for 29.45% and 21.22%, respectively (Table 3).

Ab initio predictions, homology searches and RNA-seq data were used to predict protein-coding genes of the *P. punctatissimus* and *P. argenteus* genomes using the MAKER (v3.01.03)[23] pipeline. Ab initio predictions were performed using Augustus (v3.0.3)[24] and SNAP (v2006-07-28)[25] software. In homology searches, the protein sequences of *Danio rerio*[26], *Trachinotus ovatus*[27], *Seriola dumerili*[28], *Decapterus maruadsi*[29] and *P. argenteus* were provided as protein evidence. Trinity (v2.1.1)[30] was used to assemble the clean RNA-seq reads from the NCBI database into inchworm contigs.

In total, 25,552 and 22,284 protein-coding genes were predicted in the *P. punctatissimus* and *P. argenteus* genomes, respectively, with average gene lengths of 10,935.41 bp and 12,728.71 bp, respectively. The average numbers of exons per gene were 9.21 and 10.5 (Table s2). These results are consistent with those of closely related species (Fig. 2). The predicted gene models showed excellent BUSCO completeness (actinopterygii_odb10; n = 3,640), with 94.6% (94.8% single-copy, 1.1% duplicated and only 2.4% missing) and 97.5% (96.8% single-copy, 0.7% duplicated and only 1.7% missing) of conserved genes present, demonstrating high annotation quality (Table 4). We identified syntenic blocks containing a minimum of five shared genes using MCScanX (v1.4)[31], and the resulting syntenic blocks between species were visualized using Circos (0.69–9)[32]. The presence of syntenic blocks between the two genomes further supported their quality as chromosome-scale assemblies (Fig. 3).

Functional annotations. Gene functions were inferred according to the best match of the alignments to the National Center for Biotechnology Information (NCBI) Nonredundant (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), TrEMBL and Swiss-Prot[33] protein databases using Diamond BLASTP (v2.0.7)[34] with an E-value threshold of 1E-5. The protein domains were annotated using InterProScan (v5.50-84.0)[35] based on the InterPro protein databases. A total of 23,240 and 21,623 genes were annotated with functions in at least one database, accounting for 91.31% and 97.97% of the predicted proteins of the *P. punctatissimus* and *P. argenteus* genomes, respectively (Table 5).

Annotation of noncoding RNA genes. We used tRNAscan-SE (v2.0.9)[36] algorithms with default parameters to identify the genes associated with tRNA, which is an adaptor molecule composed of RNA that bridges the three-letter genetic code in messenger RNA (mRNA) with the twenty-letter code of amino acids in proteins. RNAmmer (v1.2)[37] was used to predict rRNA sequences. snoRNAs are a class of small RNA molecules that guide chemical modifications of other RNAs, mainly ribosomal RNAs, transfer RNAs and small nuclear RNAs. MiRNAs and snRNAs were identified using Infernal (v1.1.2)[38] software against the Rfam (v14.6)[39] database with default parameters. The analysis revealed a total of 1076 and 1243 miRNAs, 2470 and 4913 tRNAs, 914 and 1322 snRNAs, and 2083 and 7812 rRNAs in the *P. punctatissimus* and *P. argenteus* genomes, respectively (Table 6).

Data Records

All the raw sequencing data underlying this article will be shared upon request to the corresponding author. All the raw sequencing data generated during this study have been deposited at the National Genomics Data Center as a BioProject under accession number PRJCA040983. The PacBio HiFi reads of the genome were deposited under the accession numbers CRR1956151 and CRR1956152. The Hi-C reads were archived under the accession numbers CRR1956153 and CRR1956154. The whole-genome sequence data reported in this paper have been deposited in the Genome Warehouse at the National Genomics Data Center under accession numbers

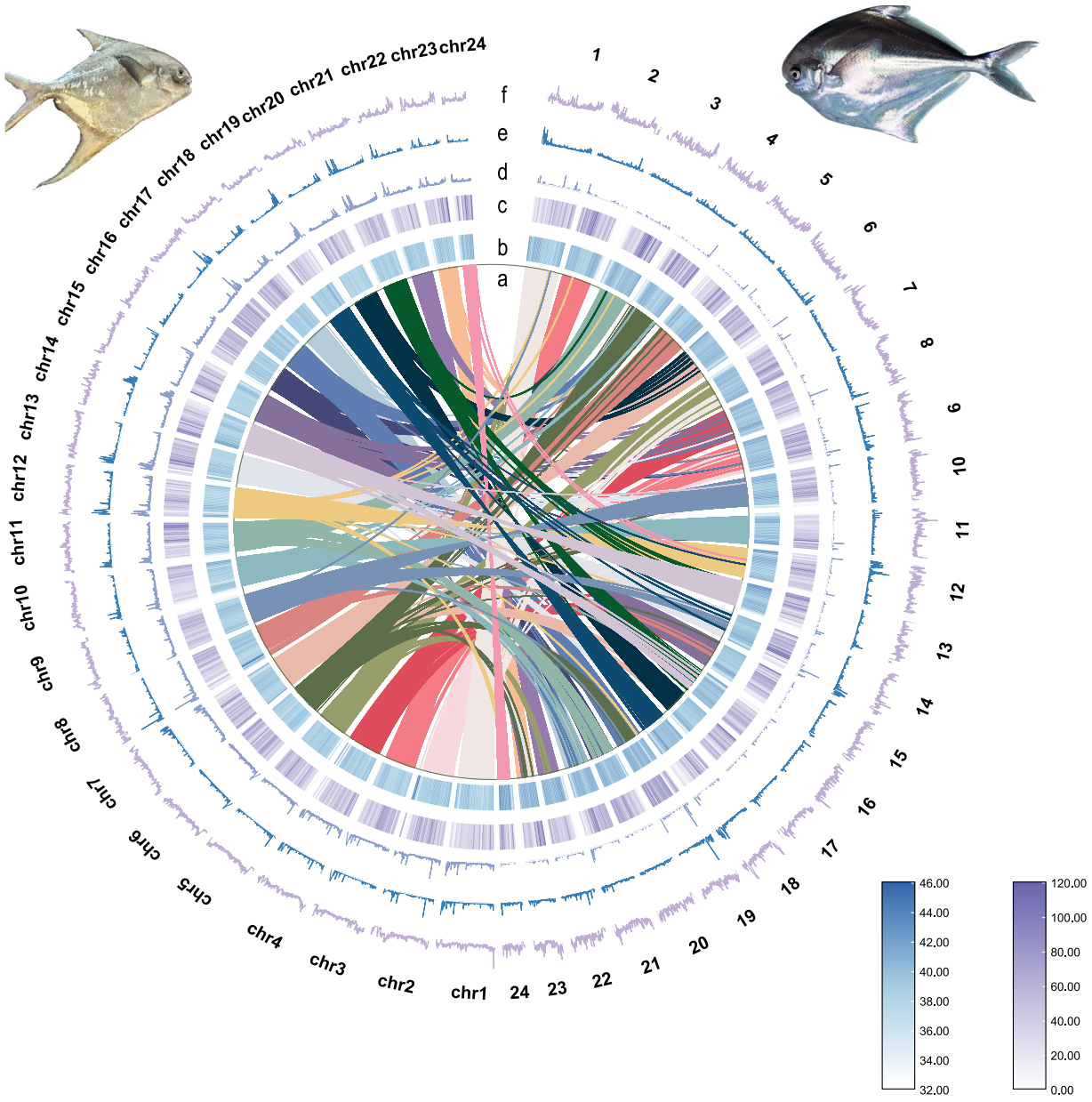

Fig. 3 Genomic features of the *P. punctatissimus* (left) and *P. argenteus* (right) genomes. Circles from inner to outer separately represent a. interchromosomal synteny between *P. punctatissimus* and *P. argenteus*; b. gene density; c. GC density; d. Copia density; e. Gypsy density; and f. DNA transposon density.

	Pampus punctatissimus		*Pampus argenteus*	
	Number	**Percentage(%)**	**Number**	**Percentage(%)**
Total	25451		22,072	
InterPro	19837	77.94	19,393	87.86
GO	9435	37.07	14,737	66.77
KEGG_ALL	22916	90.04	21,476	97.3
KEGG_KO	16372	64.33	15,762	71.41
Swissprot	21117	82.97	20,227	91.64
TrEMBL	22871	89.86	21,441	97.14
NR	23211	91.2	21,612	97.92
Annotated	23240	91.31	21,623	97.97
Unannotated	2211	8.69	449	2.03

Table 5. Gene function annotation.

Pampus punctatissimus				
Type	**Copy**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	1,076	87.93	94,611	0.0165
tRNA	2,470	76	187,713	0.0328
rRNA	2,083	714.69	1,488,707	0.2598
snRNA	914	146.89	134,255	0.0234
Pampus argenteus				
Type	**Copy**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	1,243	87.48	108,732	0.02
tRNA	4,913	75.3	369,928	0.07
rRNA	7,812	261.29	2,041,195	0.37
snRNA	1,322	147.44	194,918	0.04

Table 6. Non-RNA annotation information.

GWHGEMB00000000.1 and GWHGEMC00000000.1, which are publicly available at https://ngdc.cncb.ac.cn/gwh. The assembly data are additionally available at Genbank[16,17].

Technical Validation

DNA quality assessment was performed for both sequencing platforms. Electrophoretic analysis revealed high-molecular-weight DNA fragments (>30 kb) suitable for PacBio sequencing. Nucleic acid purity was verified by spectrophotometry (NanoDrop ND-1000, Labtech), with A260/A280 ratios meeting high-quality DNA standards. For Illumina sequencing, the DNA concentrations quantified by a fluorimeter (Qubit 4.0; Thermo Fisher) exceeded the minimum requirements for library construction. Genome characterization included two key analyses: (1) 21-mer frequency distribution estimation using Illumina short reads to predict genome size and (2) assembly evaluation with BUSCO (v5), which revealed >98% complete ortholog recovery. Comparative genomics revealed 24 conserved syntenic blocks between *P. punctatissimus and P. argenteus*, validating the chromosomal-level accuracy of our assembly.

Code availability

All commands and pipelines used in data processing were executed according to the manual and protocols of the corresponding bioinformatics software.

Data availability

The dataset has been deposited to National Genomics Data Center as a BioProject under accession number PRJCA040983.The chromosome-level genome assembly of *P. punctatissimus* and *P. argenteus* has been deposited in the National Center for Biotechnology Information (NCBI) GenBank database under the accession number JBTAOU000000000 and JBTAOV000000000. Genome assembly data of *P. punctatissimus* and *P. argenteus* have been deposited in the Genome Warehouse (GWH) in National Genomics Data Center (NGDC) under accession number GWHGEMB00000000.1 (*P. punctatissimus*) and GWHGEMC00000000.1 (*P. argenteus*).

Received: 22 September 2025; Accepted: 27 January 2026;
Published online: 25 February 2026

References

1. Liu, J., Li, C. & Ning, P. A redescription of grey pomfret Pampus punctatissimus (Bloch, 1795) with the designation of a neotype (*Teleostei: Stromateidae*). *Chin. J. Ocean. Limnol.* **31**, 140–145, https://doi.org/10.1007/s00343-013-2039-9 (2013).
2. Wei, J. H. *et al.* Species diversity and distribution of genus Pampus (*Pelagiaria: Stromateidae*) based on global mitochondrial data. *Front. Mar. Sci* **9**, 1050386, https://doi.org/10.3389/fmars.2022.1050386 (2022).
3. Zhu, F. *et al.* Transcriptome analysis unveils adaptation strategies in silver pomfret (*Pampus argenteus*) gills under hypoxic conditions. *Fish Physiol. Biochem.* **51**(2), 79, https://doi.org/10.1007/s10695-025-01492-0 (2025).
4. Cui, Z. *et al.* Molecular identification of Pampus fishes (*Perciformes, Stromateidae). Ichthyol. Res.* **57**, 32–39, https://doi.org/10.1007/s10228-009-0119-9 (2010).
5. Huang, X. *et al.* Analyses of growth performance and realized heritability of the second generation of Indo-Pacific *Pampus argenteus*. *J. Fish Biol.* **102**(3), 596–604, https://doi.org/10.1111/jfb.15291 (2023).
6. Kang, B., Wang, L. & Liu, M. Species traits determined different responses to zero-growth policy in China's marine fisheries. *Sci. Rep.* **12**(1), 20410, https://doi.org/10.1038/s41598-022-24897-w (2022).
7. Zhang, C. *et al.* Genomics and transcriptomics identify quantitative trait loci affecting growth-related traits in silver pomfret (*Pampus argenteus*). *Comp. Biochem. Physiol. D Genomics Proteomics* **54**, 101414, https://doi.org/10.1016/j.cbd.2025.101414 (2025).
8. Hu, J. *et al.* Two high quality chromosome-scale genome assemblies of female and male silver pomfret (*Pampus argenteus*). *Sci. Data* **11**(1), 1100, https://doi.org/10.1038/s41597-024-03914-9 (2024).
9. Yue, G. H. & Wang, L. Current status of genome sequencing and its applications in aquaculture. *Aquaculture* **468**, 337–347, https://doi.org/10.1016/j.aquaculture.2016.10.036 (2017).
10. Wei, J. *et al.* Chromosome-level genome assembly of the silver pomfret *Pampus argenteus*. *Sci. Data* **11**(1), 234, https://doi.org/10.1038/s41597-024-03070-0 (2024).
11. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* **7**, 1338, https://doi.org/10.12688/f1000research.15931.2 (2018).
12. *National Genomics Data Center* https://ngdc.cncb.ac.cn/gsa/browse/CRA027317 (2025).

13. Cheng, H. *et al*. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**(2), 170–175, https://doi.org/10.1038/s41592-020-01056-5 (2021).
14. Dudchenko, O. *et al*. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**(6333), 92–95, https://doi.org/10.1126/science.aal3327 (2017).
15. Robinson, J. T. *et al*. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst.* **6**(2), 256–258.e1, https://doi.org/10.1016/j.cels.2018.01.001 (2018).
16. *NCBI GenBank* https://www.ncbi.nlm.nih.gov/nuccore/JBTAOV000000000 (2025).
17. *NCBI GenBank* https://www.ncbi.nlm.nih.gov/nuccore/JBTAOU000000000 (2025).
18. Simão, F. A. *et al*. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**(19), 3210–3212, https://doi.org/10.1093/bioinformatics/btv351 (2015).
19. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14), 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
20. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA.* **117**(17), 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
21. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**(Web Server issue), W265–W268, https://doi.org/10.1093/nar/gkm286 (2007).
22. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **4**, 4.10.1–4.10.14, https://doi.org/10.1002/0471250953.bi0410s25 (2009).
23. Cantarel, B. L. *et al*. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**(1), 188–196, https://doi.org/10.1101/gr.6743907 (2008).
24. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**(Web Server issue), W309–W312, https://doi.org/10.1093/nar/gkh379 (2004).
25. Korf, I. Gene finding in novel genomes. *BMC Bioinf* **5**, 59, https://doi.org/10.1186/1471-2105-5-59 (2004).
26. Deng, Y. *et al*. Extensive sequence divergence between the reference genomes of two zebrafish strains, Tuebingen and AB. *Mol. Ecol. Resour.* **22**(6), 2148–2157, https://doi.org/10.1111/1755-0998.13602 (2022).
27. Zhang, D. C. *et al*. Chromosome-level genome assembly of golden pompano (*Trachinotus ovatus*) in the family Carangidae. *Sci. Data* **6**(1), 216, https://doi.org/10.1038/s41597-019-0238-8 (2019).
28. Nishiki, I., Yoshida, T. & Fujiwara, A. Complete genome sequence and characterization of virulence genes in Lancefield group C Streptococcus dysgalactiae isolated from farmed amberjack (*Seriola dumerili*). *Microbiol. Immunol.* **63**(7), 243–250, https://doi.org/10.1111/1348-0421.12716 (2019).
29. Chen, L. *et al*. Chromosome-level assembly and gene annotation of *Decapterus maruadsi* genome using Nanopore and Hi-C technologies. *Sci. Data* **11**(1), 69, https://doi.org/10.1038/s41597-024-02912-1 (2024).
30. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**(7), 644–652, https://doi.org/10.1038/nbt.1883 (2011).
31. Wang, Y. *et al*. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**(7), e49, https://doi.org/10.1093/nar/gkr1293 (2012).
32. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**(9), 1639–1645, https://doi.org/10.1101/gr.092759.109 (2009).
33. Boeckmann, B. *et al*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**(1), 365–370, https://doi.org/10.1093/nar/gkg095 (2003).
34. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**(Web Server issue), W20–W25, https://doi.org/10.1093/nar/gkh435 (2004).
35. Quevillon, E. *et al*. InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**(Web Server issue), W116–W120, https://doi.org/10.1093/nar/gki442 (2005).
36. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**(16), 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
37. Lagesen, K. *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**(9), 3100–3108, https://doi.org/10.1093/nar/gkm160 (2007).
38. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**(22), 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
39. Kalvari, I. *et al*. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**(D1), D192–D200, https://doi.org/10.1093/nar/gkaa1047 (2021).

Acknowledgements

This work was supported by the Jiangsu Province Frontier Technology R&D Program in Modern Agriculture (Grant No. BF2025324), 2025 Nantong Seed Industry Revitalization Project, Hainan Seed Industry Laboratory Postdoctoral Sustentation Fund (B24C1000101), and the High-Performance Computing Platform of YZBSTCACC.

Author contributions

Fei Zhu: Conceptualization, Methodology, Formal analysis, Writing-Original Draft, Writing-Review & Editing, Project administration, Funding acquisition; Sirong Jiang: Writing-Original Draft, Formal analysis & uploaded data; Yanyu Zhang: Validation, Investigation; Shaoyang Bo: Investigation; Ting Ye: Investigation; Bo Gao: Resources; Ruijian Sun: Investigation; Dafeng Xu: Investigation; Shuran Du: Methodology; Chaofeng Jia: Investigation; Qian Meng: Investigation; Jian Luo: Conceptualization, Writing-Review & Editing, Supervision.

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/s41597-026-06739-w.

Correspondence and requests for materials should be addressed to G.B. or J.L.