scientific data

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of an Arctic fish species pale eelpout (*Lycodes pallidus*)

Ran Zhang, Rui Wang, Xing Miao, Hai Li, Puqing Song, Yuan Li ✉ & Longshan Lin ✉

Eelpouts (Zoarcidae) are known for their bipolar distributions and distinctive biogeographic histories. However, limited genomic data have hindered our understanding of their adaptive evolution. In this study, we present a thoroughly annotated chromosome-level genome assembly of pale eelpout (*Lycodes pallidus*) generated through the integration of Illumina, PacBio circular consensus, and Hi-C sequencing techniques. The final assembly spans 753.4 Mb, with its high quality confirmed by a scaffold N50 of 28.6 Mb and a Benchmarking Universal Single-Copy Ortholog (BUSCO) completeness of 99.3%. In comparison to other eelpouts and related fishes, the *L. pallidus* genome is larger and exhibits greater repetitive element content, accounting for approximately 45% of its total length. We annotated 21,419 protein-coding genes, a significant proportion of which are involved in signal transduction mechanisms and transcription. These findings provide valuable genetic resources for elucidating the evolutionary mechanisms underlying polar fish adaptation.

Background & Summary

The family Zoarcidae (eelpouts) demonstrates remarkable adaptive radiation and is abundant at both poles[1–3]. These teleosts originated in the Northern Hemisphere, evolved unique type III antifreeze proteins (AFPs) to survive subfreezing temperatures during the Miocene Ice Age, subsequently colonized the Southern Oceans via specific migratory routes, and are currently among the most rapidly speciating fishes[4–8], contributing to their distinct biogeographic patterns. Within this group, the genus *Lycodes* Reinhardt, 1831 comprises more than 60 species, making it the most extensive lineage[9]. Among them, the pale eelpout (*Lycodes pallidus* Collett, 1879) serves as an ideal model for evolutionary studies. It follows an Arctic zoogeographic pattern, commonly occurring in the Greenland, Barents, and Kara Seas, and has recently been observed in the Pacific Arctic regions by the Chinese National Arctic Research Expedition (CHINARE) (Fig. 1)[3,10,11]. Consequently, a high-quality chromosome-level genome assembly for *L. pallidus* is essential for gaining insights into eelpout sympatric speciation, adaptive radiation to polar extremes, and biogeographic hirtory.

Here, we report a chromosome-level genome assembly for *L. pallidus* constructed by combining Illumina short reads, PacBio high-fidelity (HiFi) long reads, and Hi-C chromatin interaction data. The genome covers 753.4 Mb, similar in size to the *Anisarchus medius* genome (739.1 Mb)[12] but larger than that of other eelpout species, such as *Zoarces viviparus* (663.0 Mb)[13], *Melanostigma gelatinosum* (662.0 Mb)[14], and *Lycodes pacificus* (646.4 Mb) (GenBank GCA_028022725.1). The assembly consists of 24 pseudo-chromosomes ranging from 15.4 Mb to 35.6 Mb in length, with a scaffold N50 of 28.6 Mb (Fig. 2A). Its integrity and accuracy are confirmed by over 99% genome coverage from short and long reads, alongside a benchmarking universal single-copy ortholog (BUSCO) analysis indicating 99.3% complete genes (Fig. 2B). Approximately 30.0% of the genome was identified as transposable elements (TEs). Our annotation pipeline predicted a total of 21,419 protein-coding genes, highlighting functional enrichment in signal transduction and transcriptional regulation pathways. This assembly establishes a robust foundation for investigating the molecular basis of adaptive evolution in polar fishes.

Laboratory of Marine Biodiversity Research, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, 361005, P.R. China. ✉e-mail: liyuan@tio.org.cn; linlongshan@tio.org.cn

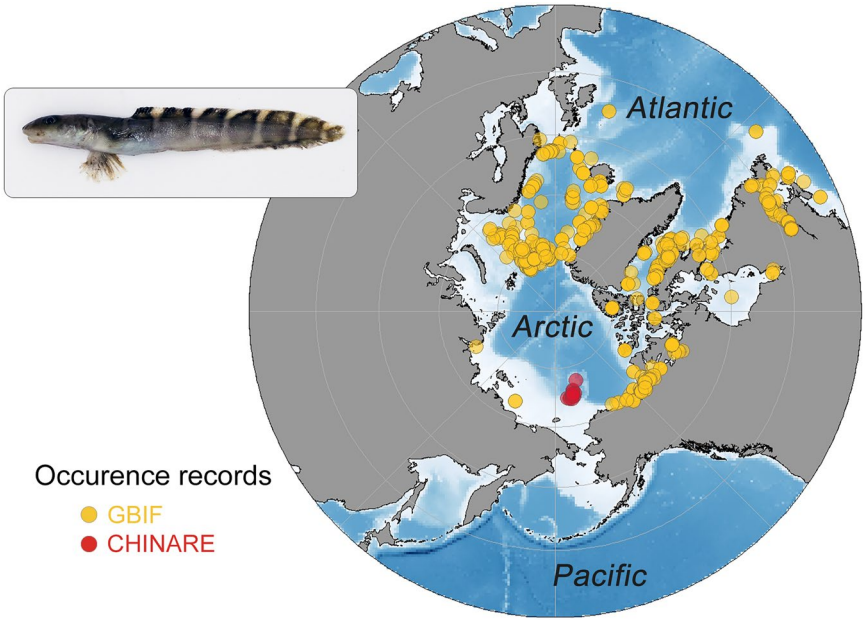

Fig. 1 A sample image (top left) of *Lycodes pallidus* with its occurrence records from the Global Biodiversity Information Facility (GBIF) (yellow dots) and the Chinese National Arctic Research Expedition (CHINARE) (red dots).

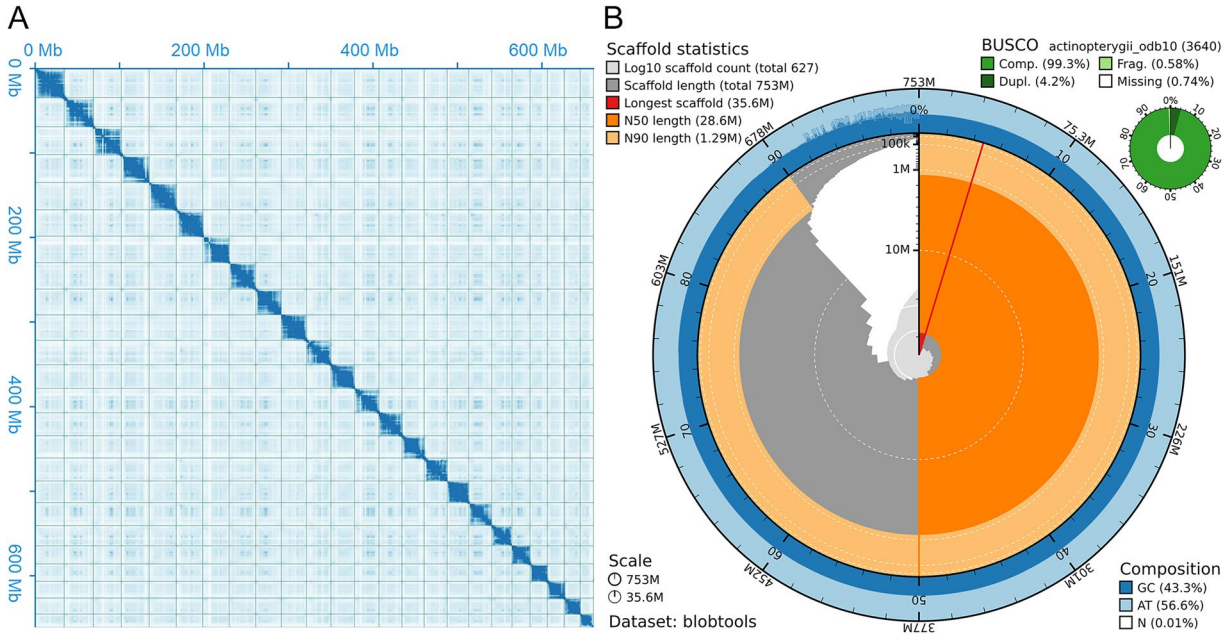

Fig. 2 Integrity assessment of *Lycodes pallidus* genome. (**A**) The contig contact density map constructed using Hi-C data represents the 24 pseudo-chromosomes. (**B**) Snail plot of the assembly, integrating scaffold statistics, base composition, and BUSCO results.

Methods

Sample collection and genome sequencing. The fish specimen was collected aboard the RV *Xue Long 2* at station *R13* (75.46°N, 167.89°W) in the Chukchi Borderland, Central Arctic Ocean, on July 31, 2023 using a 6.5 m triangular net (a triangular structure made of steel with 2.2 m wide and 0.65 m tall opening and a 20 mm mesh codend liner). The sample was frozen in liquid nitrogen for 12 hours prior to being stored at −80 °C until DNA extraction. Fish surveys and monitoring are integral to the CHINARE, with the field collection program approved by the Chinese Arctic and Antarctic Administration of the State Oceanic Administration (CAA) and the Polar Research Institute of China (PRIC). The Institutional Committee on the Care and Use of Animals at the

Length range (bp)	Total number	Total length (bp)	Mean length (bp)
1~2000	199	204,450	1027.39
2000~4000	964	3,144,827	3262.27
4000~6000	2,194	11,111,566	5064.52
6000~8000	2,969	20,938,192	7052.27
8000~10000	3,428	30,940,504	9025.82
10000~12000	9,863	112,268,215	11382.77
12000~14000	333,033	4,467,035,139	13413.19
14000~16000	809,440	12,089,160,512	14935.22
16000~18000	343,366	5,757,048,107	16766.51
18000~	48,243	910,704,268	18877.44

Table 1. Statistic summary of HiFi reads obtained through PacBio circular consensus sequencing.

Third Institute of Oceanography, Ministry of Natural Resources, confirmed that no experimental animals were involved in the collection or analysis of these Arctic fish samples. The occurrence records were derived from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/).

High-quality genomic DNA was extracted from muscle tissue using the sodium dodecyl sulfate (SDS) protocol, and three sequencing libraries were prepared for genome survey and assembly. First, the short fragment library was built using the VAHTS Universal Plus DNA Library Prep Kit for Illumina ND617 (Vazyme Biotech Co., Ltd., China) following the manufacturer's instructions. After verifying the size distribution and concentration using a Qseq. 400 (BIOptic Inc., China) and a Qubit 3.0 Fluorometer (Life Technologies, CA, USA), the library was sequenced on the Illumina NovaSeq X Plus Platform (Illumina, Inc., USA) to generate 150-bp paired-end (PE150) reads. Second, the long read library was built using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, USA) through magnetic bead purification, DNA damage and end repair, and adapter ligation, following the standard PacBio protocol (Pacific Biosciences, USA). After examining the concentration and integrity using a Nanodrop 2000 (Thermo Fisher Scientific, USA) and a Qubit 3.0 Fluorometer, the library was subjected to the PacBio Revio System (Pacific Biosciences, USA) on circular consensus sequencing (CCS) model for obtaining HiFi reads (Table 1). Third, the Hi-C library was built using the Frasergen Hi-C Kit (Frasergen, China) following the protocol instructions, with muscle tissue processed via cell cross-linking, DNA digestion and fragmentation (DpnII), biotin labelling of repaired sticky ends, ligation, enrichment of biotin-labelled fragments, and purification. The concentration of 300–700 bp insert reads was assessed using an Agilent 2100 bioanalyzer (Agilent, USA) and a Qubit 3.0 fluorometer, and the library was sequenced on the Illumina HiSeq sequencing platform (Illumina, Inc., USA) to produce PE150 reads.

Total RNA was isolated from pooled tissues (skin, muscle, and liver) using the TRIzol reagent kit (Thermo Fisher Scientific Inc., USA) to assist in gene prediction. The mRNA was collected using oligo (dT) coupled to magnetic beads and randomly fragmented with RNA Fragmentation Reagents (Thermo Fisher Scientific Inc., USA). The first-strand cDNA was synthesized with fragmented mRNA as the template, followed by second-strand synthesis. The cDNA was purified with AMPure XP beads (Beckman Coulter Inc., USA). Double-strand cDNA was subjected to end repair, adenylated, and ligated to adapters. After quality and concentration examination of the extracted RNA, the final cDNA libraries were sequenced on the Illumina HiSeq platform (Illumina Inc., USA) to generate PE150 reads. The isolation and sequencing processes were performed by the Biomarker Technologies Co., Ltd. (Beijing, China).

Reads preprocessing and genome assembly. The raw Illumina, HiFi, and Hi-C sequencing data were subjected to adapter trimming and quality filtering using Fastp v0.24.0[15] with the parameters "-q 10 -u 50 -y -g -Y 10 -e 20 -l 100", yielding 34.26 Gb of high-quality Illumina short reads with a GC content of 42.47% and Q20 and Q30 base percentages of 96.47% and 90.69%, respectively. Alignments of 10,000 randomly selected clean short reads to the nucleotide sequence (NT) database with BLAST + v2.9.0[16] suggested the absence of contamination in the DNA sample. A *k*-mer ($k = 21$) distribution analysis performed with Jellyfish v2.2.10[17] and GenomeScope 2.0 v2.0.1[18] revealed a genome repeat content of 28.49% and heterozygosity of 1.21% (Fig. 3).

After removing low-quality reads and adapters, the retained 23.40 Gb clean HiFi long reads had Q20 and Q30 scores of 97.57% and 94.15%, respectively, and the 91.24 Gb clean Hi-C reads had Q20 and Q30 scores of 97.51% and 93.79%. A primary *de novo* assembly was generated using hifiasm v0.24.0-r702[19] by integrating HiFi and Hi-C reads, followed by polishing using NextPolish v1.4.1[20] with HiFi and Illumina reads, and contamination removal using FCS-GX v0.5.4[21]. The Hi-C data were mapped to the genome draft and indexed using BWA v2.2.1[22], SAMtools v1.21[23], and Picard Toolkit (https://broadinstitute.github.io/picard/) following Arima-HiC Mapping Pipeline A160156 v03 (https://github.com/ArimaGenomics/mapping_pipeline). Chromosome-scale scaffolds were subsequently constructed using Juicer v1.19.02[24] and YaHS v1.2.2[25]. The gaps were filled using Minimap2 v2.28[26] and TGS-GapCloser v.1.2.1[27], and a second round of Hi-C data scaffolding finalized the chromosome-level genome assembly (Table 2).

Repetitive sequence analysis. TEs were *de novo* identified using EDTA v2.2.2[28], an automated whole-genome annotation pipeline that incorporates the LTRharvest, LTR_retriever, GRF, TIR-Learner, HelitronScanner, and RepeatModeler programs, along with customized filtering scripts, for recognition of TE

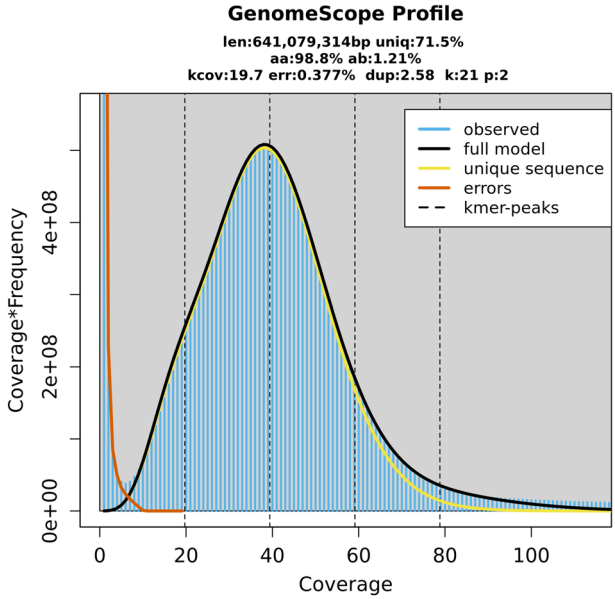

Fig. 3 *K*-mer distribution map ($k = 21$) based on the Illumina short reads.

Assembly statistics	Long reads (PacBio) + Hi-C reads
Contig N50	4,682,479 bp
Scaffold N50	28,576,443 bp
Scaffold L50	12
Scaffold N90	1,288,722 bp
Scaffold L90	32
Total scaffolds	627
GC content	43.35%
Total length	753,345,134 bp
Number of annotated protein-coding genes	21,419
BUSCO single complete	95.1%
BUSCO duplicated complete	4.2%

Table 2. *Lycodes pallidus* genome assembly statistics.

subclass and compilation of a comprehensive non-redundant TE library. The TEs constituted 28.13% of the *L. pallidus* genome, including 57.7 Mb long terminal repeat retrotransposons (LTRs), 20.3 Mb long interspersed nuclear elements (LINEs), and 69.2 Mb terminal inverted repeats (TIRs), representing 7.66%, 2.70%, and 9.19% of the genome, respectively (Table 3), suggesting recent transposition activities (Fig. 4). The interspersed repeats were also screened and masked using RepeatModeler2 v2.0.6[29] and RepeatMasker v4.1.7-p1 (https://repeatmasker.org/), accounting for 29.61% of the assembly. Moreover, the tandem repeats (TRs) were estimated to be 118.9 Mb using TRF v4.10.0[30] with the parameters "2 7 7 80 10 50 500 (for matching weight, mismatching penalty, indel penalty, match probability, indel probability, and minimum alignment score and maximum period size to report)", based on the retention of the longest sequence after deduplication.

Gene prediction and annotation. We employed BRAKER3 v3.0.7.6[31], an automated pipeline, to integrate evidence from the RNA-seq reads, which were mapped to the indexed soft-masked assembly using HISAT2 v2.2.1[32], and protein data from the vertebrate OrthoDB[33] to predict reliable genes with GeneMark-ETP[34] and AUGUSTUS[35,36]. As a result, a total of 21,419 protein-coding genes and 26,554 protein sequences were predicted, with average lengths of 8,637.71 bp and 568.15 amino acids (aa), respectively. Genome-wide gene function was inferred via orthology assignment using eggNOG-mapper v2.1.12 combined with the eggNOG database[37,38], and annotations were extracted and converted from the KOG databases with TBtools-II v2.210[39] (Fig. 5). Additionally, 28,238 transfer RNAs were identified with tRNAscan-SE v2.0.12[40], and 1,114 ribosomal RNAs, 1,155 microRNAs, and 2,794 small nuclear RNA genes were identified using Infernal v1.1.5[41] based on the Rfam database[42].

Class	Subclass	Count	Length (bp)	Proportion
LINE	CR1	238	103971	0.01%
	I	2160	1178836	0.16%
	Jockey	89	39306	0.01%
	L1	3827	1853487	0.25%
	L2	18733	8995444	1.19%
	Proto2	795	293213	0.04%
	R2	301	273527	0.04%
	RTE	4002	1326255	0.18%
	Rex	10372	4483516	0.60%
	unknown	4756	1794515	0.24%
LTR	Bel_Pao	1275	711834	0.09%
	Copia	3106	1453325	0.19%
	Endogenous_Retrovirus	1080	262129	0.03%
	Gypsy	14235	10242252	1.36%
	Retrovirus	863	327661	0.04%
	unknown	120999	44713252	5.94%
SINE	Alu	194	94631	0.01%
	tRNA	3737	921797	0.12%
TIR	CACTA	69084	15553584	2.06%
	Dada	172	154284	0.02%
	IS3EU	863	463976	0.06%
	Mutator	43442	16130850	2.14%
	PIF_Harbinger	8776	2018564	0.27%
	Tc1_Mariner	2172	671638	0.09%
	hAT	83140	34228845	4.54%
	piggyBac	55	26598	0.00%
nonLTR	DIRS_YR	274	38352	0.01%
	Ngaro_YR	251	79293	0.01%
	Penelope	1442	1533591	0.20%
nonTIR	Crypton_YR_transposon	124	23915	0.00%
	helitron	58827	12767649	1.69%
Low complexity	/	34	6244	0.00%
Repeat fragment	/	69489	49160412	6.53%
Total	/	528907	211926746	28.12%

Table 3. Summary of whole-genome transposable elements (TEs) annotation of *Lycodes pallidus* genome.

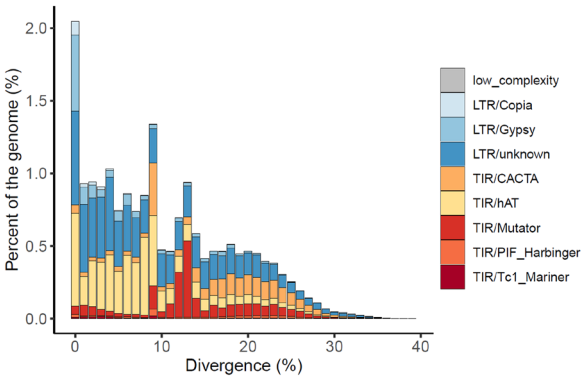

Fig. 4 Repeat landscape plots showing the expansion of transposable elements (TEs) in the *Lycodes pallidus* genome. The distribution of TE copies as percentage of divergence from consensus repeat models versus genome coverage.

Data Records

The reference genome assembly and all sequencing data, including Illumina short reads, PacBio long reads, and Hi-C data, were deposited in the NCBI database under BioProject PRJNA1231400 (BioSample SAMN47210740, GenBank GCA_049309625.1[43], and SRA SRR32607915[44], SRR32607916[45], and SRR32607917[46]). The reference genome, masked assembly, repetitive sequences, and gene annotation files are available on Figshare[47,48].

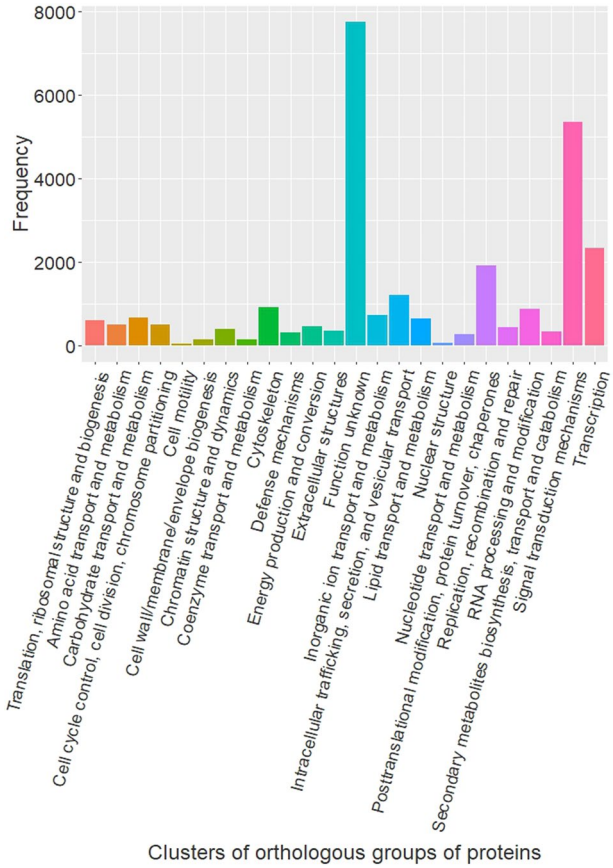

Fig. 5 Cluster of orthologous of protein-coding gene in *Lycodes pallidus* genome.

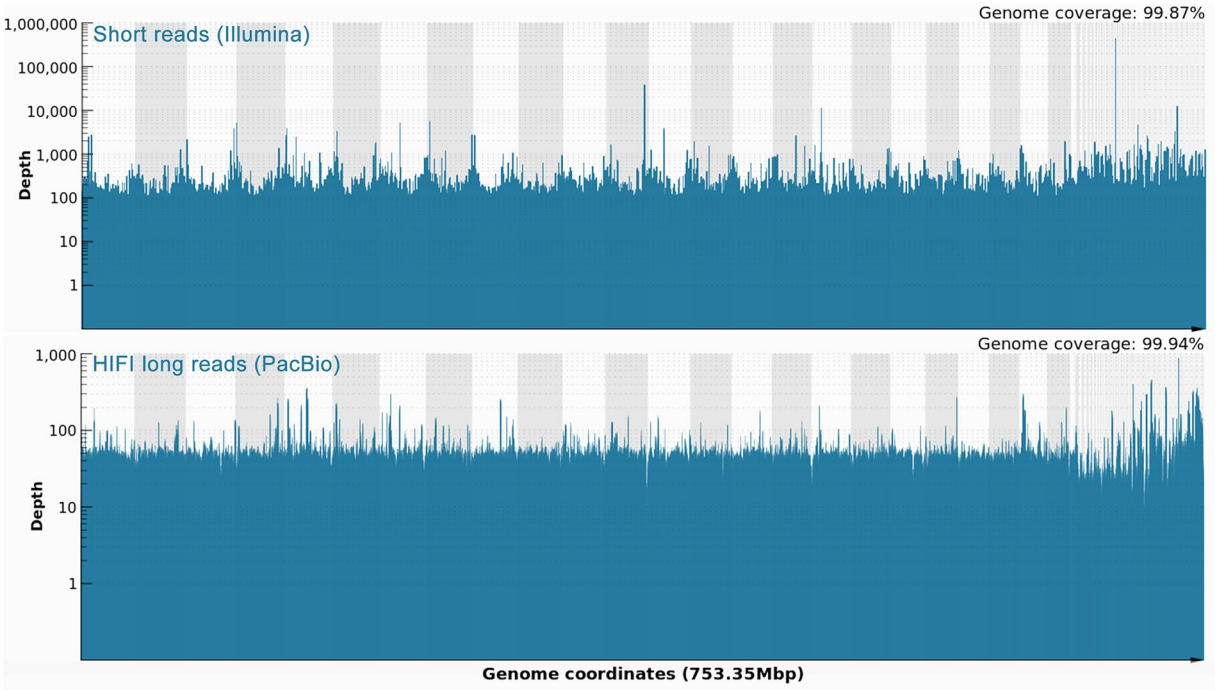

Fig. 6 The genome coverage and sequence depth of short reads and long reads to the final assembly indicate a highly complete assembly.

Technical Validation

The mapping rates of the short and long reads to the final assembly were analysed using BWA-MEM2 v2.2.1[49] and Minimap2 v2.28[26], and the genome coverage and depth of the qualified bam files of the reads to the final assembly were assessed and visualized using bamdep (https://github.com/slw287r/dep). Both Illumina reads and HiFi reads achieved over 99% genome coverage and mapping rates (Fig. 6).

The final assembly quality was assessed using BUSCO v5.8.3 on the basis of the "actinopterygii_odb10 (v5/2024-01-08)" dataset[50], and a snail plot was generated using BlobToolKit v4.4.5[51]. The results revealed that 99.3% of the BUSCOs were classified as "complete" ("single-copy" and "duplicated"), with only 0.6% being "fragmented" and 0.7% being "missing", indicating highly complete assembly (Fig. 2; Table 2).

Code availability

All software and tools used are publicly accessible, with the versions specified in the *Methods* section. The default parameters were applied unless otherwise stated, per the manuals and protocols of the respective bioinformatics software.

Received: 7 April 2025; Accepted: 10 June 2025;
Published online: 10 July 2025

References

1. Gon, O. & Heemstra, P. C. *Fishes of the South Ocean* (J. L. B. Smith Institute of Ichthyology, 1990).
2. Lannoo, M. J. & Eastman, J. T. Brain and sensory organ morphology in Antarctic eelpouts (Perciformes: Zoarcidae: Lycodinae). *J. Morphol.* **267**, 115–127, https://doi.org/10.1002/jmor.10391 (2006).
3. Mecklenburg, C. W. *et al. Marine Fishes of the Arctic Region* (Conservation of Arctic Flora and Fauna, 2018).
4. Rabosky, D. L. *et al*. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395, https://doi.org/10.1038/s41586-018-0273-1 (2018).
5. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, e89543, https://doi.org/10.1371/journal.pone.0089543 (2014).
6. Radchenko, O. A. The system of the suborder Zoarcoidei (Pisces, Perciformes) as inferred from molecular genetic data. *Russ. J. Genet.* **51**, 1096–1112, https://doi.org/10.1134/S1022795415100130 (2015).
7. Hobbs, R. S., Hall, J. R., Graham, L. A., Davies, P. L. & Fletcher, G. L. Antifreeze protein dispersion in eelpouts and related fishes reveals migration and climate alteration within the last 20 Ma. *PLoS ONE* **15**, e0243273, https://doi.org/10.1371/journal.pone.0243273 (2020).
8. Hotaling, S., Borowiec, M. L., Lins, L. S. F., Desvignes, T. & Kelley, J. L. The biogeographic history of eelpouts and related fishes: linking phylogeny, environmental change, and patterns of dispersal in a globally distributed fish group. *Mol. Phylogenetics Evol.* **162**, 107211, https://doi.org/10.1016/j.ympev.2021.107211 (2021).
9. Møller, P. R. & Gravlund, P. Phylogeny of the eelpout genus *Lycodes* (Pisces, Zoarcidae) as inferred from mitochondrial cytochrome *b* and 12S rDNA. *Mol. Phylogenetics Evol.* **26**, 369–388, https://doi.org/10.1016/S1055-7903(02)00362-7 (2003).
10. Mecklenburg, C. W., Møller, P. R. & Steinke, D. Biodiversity of arctic marine fishes: taxonomy and zoogeography. *Mar. Biodiv.* **41**, 109–140, https://doi.org/10.1007/s12526-010-0070-z (2011).
11. Zhang, R. Fish biodiversity in the Central Arctic Ocean and the biogeography of Zoarcidae (Xiamen University, 2022).
12. Liu, R. *et al*. Chromosome-level reference genome and annotation of the Arctic fish *Anisarchus medius*. *Sci. Data* **12**, 68, https://doi.org/10.1038/s41597-025-04419-9 (2025).
13. Fuhrmann, N. *et al*. Chromosome-Level Genome Assembly of the Viviparous Eelpout *Zoarces viviparus*. *Genome Biol. Evol.* **16**, vae155, https://doi.org/10.1093/gbe/evae155 (2024).
14. Bista, I. & Collins, M. Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory Team. The genome sequence of the limp eelpout, *Melanostigma gelatinosum* Günther, 1881. *Wellcome Open Res.* **9**, 192, https://doi.org/10.12688/wellcomeopenres.21268.1 (2024).
15. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884−−i890, https://doi.org/10.1093/bioinformatics/bty560 (2018).
16. Camacho, C. *et al*. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421, https://doi.org/10.1186/1471-2105-10-421 (2009).
17. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).
18. 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, https://doi.org/10.1038/s41467-020-14998-3 (2020).
19. 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, https://doi.org/10.1038/s41592-020-01056-5 (2021).
20. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics* **36**, 2253–2255, https://doi.org/10.1093/bioinformatics/btz891 (2019).
21. Astashyn, A. *et al*. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol.* **25**, 60, https://doi.org/10.1186/s13059-024-03198-7 (2024).
22. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
23. Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008, https://doi.org/10.1093/gigascience/giab008 (2021).
24. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98, https://doi.org/10.1016/j.cels.2016.07.002 (2016).
25. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808, https://doi.org/10.1093/bioinformatics/btac808 (2023).
26. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574, https://doi.org/10.1093/bioinformatics/btab705 (2021).
27. 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**, giaa094, https://doi.org/10.1093/gigascience/giaa094 (2020).
28. Ou, S. *et al*. Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline. *Genome Biol.* **20**, 275, https://doi.org/10.1186/s13059-019-1905-y (2019).
29. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA.* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
30. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580, https://doi.org/10.1093/nar/27.2.573 (1999).

31. Gabriel, L. *et al*. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *Genome Res.* **34**, 769–777, https://doi.org/10.1101/gr.278090.123 (2024).
32. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915, https://doi.org/10.1038/s41587-019-0201-4 (2019).
33. Kuznetsov, D. *et al*. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **51**, D445–D451, https://doi.org/10.1093/nar/gkac998 (2023).
34. Bruna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757–768, https://doi.org/10.1101/gr.278373.123 (2024).
35. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644, https://doi.org/10.1093/bioinformatics/btn013 (2008).
36. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **62**, https://doi.org/10.1186/1471-2105-7-62 (2006).
37. Huerta-Cepas, J. *et al*. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314, https://doi.org/10.1093/nar/gky1085 (2019).
38. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829, https://doi.org/10.1093/molbev/msab293 (2021).
39. Chen, C. *et al*. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* **16**, 17331742, https://doi.org/10.1016/j.molp.2023.09.010 (2023).
40. 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**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
41. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
42. Griffiths-Jones, S. *et al*. Rfam: annotating non-codin RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124, https://doi.org/10.1093/nar/gki081 (2005).
43. *NCBI GenBank. Whole Genome Assembly.* https://identifiers.org/ncbi/insdc.gca:GCA_049309625.1 (2025).
44. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32607915 (2025).
45. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32607916 (2025).
46. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32607917 (2025).
47. Zhang, R. Assembly and annotation of pale eelpout (*Lycodes pallidus*) genome. *figshare* https://doi.org/10.6084/m9.figshare.28644908 (2025).
48. Zhang, R. Repetitive sequences of pale eelpout (*Lycodes pallidus*) genome. *figshare* https://doi.org/10.6084/m9.figshare.28646843 (2025).
49. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE Trans. Parallel Distrib. Syst.* 314–324, https://doi.org/10.1109/IPDPS.2019.00041 (2019).
50. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **38**, 4647–4654, https://doi.org/10.1093/molbev/msab199 (2021).
51. Challis, R., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. BlobToolKit – Interactive quality assessment of genome assemblies. *G3: Genes Genomes Genet.* **10**(4), 1361–1374, https://doi.org/10.1534/g3.119.400908 (2020).

Acknowledgements

This work was funded by the National Natural Science Foundation of China [42276247 and 42406243]. We are grateful to the Chinese Arctic and Antarctic Administration of the State Oceanic Administration (CAA) and the Polar Research Institute of China (PRIC) for their organizational and logistical support during the expeditions. We appreciate the scientists and crews on Research Vessels *Xue Long 2* for their assistance in operating the bottom trawls during the surveys. We acknowledge Musheng Lan, Kun Liu, and Shigang Liu for their generous help in fish sampling.

Author contributions

Conceptualization, L.L. and Y.L.; Methodology, R.Z., R.W., H.L. and Y.L.; Investigation, P.S. and X.M.; Visualization, R.Z., R.W. and X.M.; Supervision, L.L. and Y.L.; Writing-original draft, R.Z.; All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

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