OPEN ACCESS

Animal Science REPORT

The Innovation Life

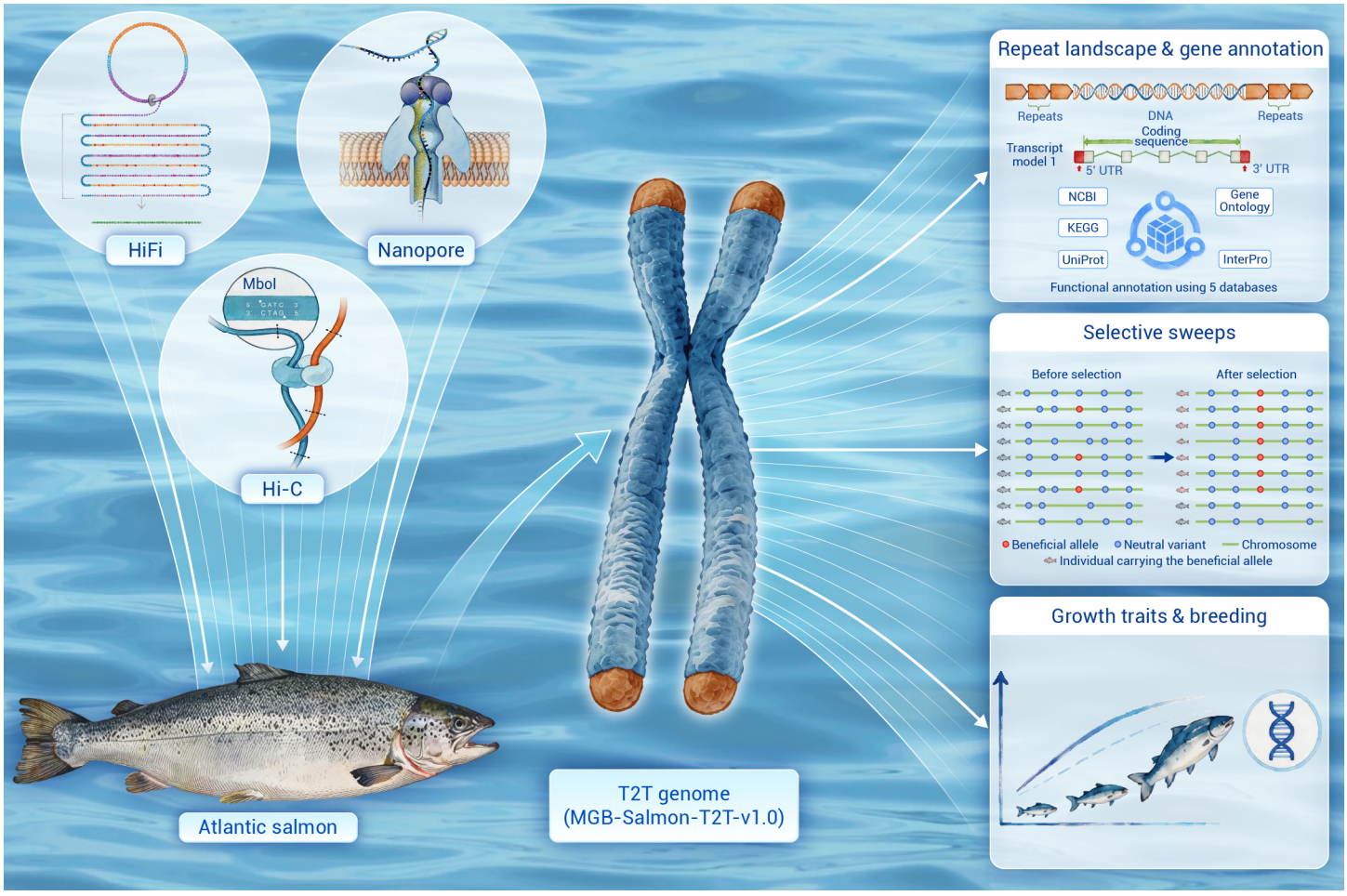
The first near-complete telomere-to-telomere genome of Atlantic salmon (*Salmo salar*) provides insights for genetic diversity

Xiaofei Yu,[1,2,6] Baojun Zhao,[1,6] Wentao Han,[1,4,6] Yuxiang Liu,[1,6] Ziyu Wang,[2] Xinghai Zhu,[1] Yangfan Wang,[1] Lingzhao Fang,[5] Shi Wang,[1,3,4] Jingjie Hu,[1,3,4,*] and Zhenmin Bao[1,3,4,*]

*Correspondence: hujingjie@ouc.edu.cn (J.H.); zmbao@ouc.edu.cn (Z.B.)

Received: December 12, 2025; Accepted: July 1, 2026; Published Online: July 7, 2026; https://doi.org/10.59717/j.xinn-life.2026.100236

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- We generated the first telomere-to-telomere (T2T) genome assembly of Atlantic salmon.
- The assembly showed high continuity and contained 15 complete T2T chromosomes.
- This Atlantic salmon genome assembly provides a key resource for selective breeding.

REPORT Animal Science

www.the-innovation.org/the-innovation-life

The first near-complete telomere-to-telomere genome of Atlantic salmon (*Salmo salar*) provides insights for genetic diversity

Xiaofei Yu,[1,2,6] Baojun Zhao,[1,6] Wentao Han,[1,4,6] Yuxiang Liu,[1,6] Ziyu Wang,[2] Xinghai Zhu,[1] Yangfan Wang,[1] Lingzhao Fang,[5] Shi Wang,[1,3,4] Jingjie Hu,[1,3,4,*] and Zhenmin Bao[1,3,4,*]

[1]MOE Key Laboratory of Marine Genetics and Qingdao Institute of Blue Seed Industry, Ocean University of China, Qingdao 266100, China

[2]Key Laboratory of Aquatic Resources and Utilization, School of Life Sciences, Nanchang University, Nanchang 330031, China

[3]Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao 266237, China

[4]Key Laboratory of Tropical Aquatic Germplasm of Hainan Province, Sanya Oceanographic Institution, Ocean University of China, Sanya 572024, China

[5]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus 8000, Denmark

[6]These authors contributed equally

*Correspondence: hujingjie@ouc.edu.cn (J.H.); zmbao@ouc.edu.cn (Z.B.)

Received: December 12, 2025; Accepted: July 1, 2026; Published Online: July 7, 2026; https://doi.org/10.59717/j.xinn-life.2026.100236

Citation: Yu X., Zhao B., Han W., et al. (2026). The first near-complete telomere-to-telomere genome of Atlantic salmon (*Salmo salar*) provides insights for genetic diversity. *The Innovation Life* **4**:100236.

Atlantic salmon is one of the world's most valuable aquaculture species, providing substantial economic and nutritional benefits. In China, the identification of the Yellow Sea Cold Water Mass has enabled offshore salmon farming and supported the development of innovative deep-sea aquaculture platforms such as Deep Blue No. 1. However, the lack of a gap-free telomere-to-telomere (T2T) reference genome has limited progress in population genetics, functional genomics, and genetic improvement programs, including ICSASG_v2 and Ssa_v3.1, contained unresolved gaps and collapsed duplicated regions. To overcome these constraints, we integrated PacBio HiFi long reads, ultra-long Oxford Nanopore reads, and Hi-C scaffolding to generate a T2T genome assembly (MGB-Salmon-T2T-v1.0) with a contig N50 of 101.69 Mb and scaffold N50 of 114.45 Mb, representing a substantial advance in genome continuity and completeness. This work provides the first T2T genome assembly of Atlantic salmon, marking a major milestone in aquaculture genomics and establishing a powerful foundation for selective breeding of this globally important species. Furthermore, constructing a species-wide catalogue of genetic variants represents a critical step toward comprehensive functional genome annotation and underpins the ongoing international FarmGTEx project, which aims to link genetic variation with regulatory activity, tissue-specific expression, and phenotypic outcomes.

INTRODUCTION

Aquaculture is increasingly important for ensuring global food security and meeting the growing demand for seafood, with aquaculture products accounting for 52% of global fish consumption in 2018 and projected to reach 59% by 2030.[1,2] Atlantic salmon (*Salmo salar*), the second most valuable farmed aquatic species after shrimp, has undergone decades of selective breeding for growth, disease resistance, and flesh quality.[3,4] Norway dominates global production, contributing approximately 55% of total output, while salmon farming has expanded to other high-latitude regions such as Canada and the United Kingdom.[5–7] In China, the discovery of the Yellow Sea Cold Water Mass and associated technological advances enabled the first successful open-ocean harvest from the "Deep Blue No.1" platform in 2021, overcoming previous constraints imposed by heat stress.[8–10] As a high-value source of protein and omega-3 fatty acids, Atlantic salmon represents an important species for climate-smart aquaculture, where continued genetic improvement is essential for enhancing resilience, reducing environmental impacts, and supporting sustainable industry growth.[11]

Selective breeding and genomic selection have substantially accelerated genetic gains in Atlantic salmon breeding programs.[12–14] However, the effectiveness of genomic selection remains constrained by incomplete characterization of regulatory variants, unresolved structural variation, and limited understanding of the molecular basis of complex traits. Successive improvements in reference genomes, from ICSASG_v2 to Ssa_v3.1 and additional lineage-specific chromosome-level assemblies, have markedly enhanced genome quality and revealed extensive genomic diversity within the species.[15] Nevertheless, a true telomere-to-telomere (T2T) Atlantic salmon genome is still lacking. This limitation hampers comprehensive identification of structural variants, accurate annotation of regulatory elements, and elucidation of the genetic architecture underlying economically important traits and disease susceptibility. Recent studies and the FarmGTEx white paper have highlighted the importance of T2T genomes for integrative functional genomics and precise regulatory variant annotation in livestock species, including cattle, pigs, chickens, and sheep.[16–20] Establishing a T2T genomic framework for Atlantic salmon is therefore critical for advancing genomic selection, genome editing, and sustainable breeding strategies.

In this study, we present the first T2T genome assembly for Atlantic salmon, providing a comprehensive and highly contiguous genomic reference for this economically important species. By resolving previously collapsed, repetitive, and duplicated regions, this assembly substantially improves genome completeness and enables more accurate characterization of complex genomic features that remained inaccessible in earlier references. This high-quality genomic resource provides a robust foundation for studies of genome structure, evolution, and functional variation, while facilitating the identification of genetic determinants underlying key aquaculture traits. Ultimately, the T2T genome offers an important resource for developing more precise and effective genomic-assisted breeding strategies in Atlantic salmon.

MATERIALS AND METHODS

Sample collection and sequencing

A 1.5-year-old female Atlantic salmon (*S. salar*) from a commercial farm in Yantai, Shandong Province, China, was used for genome assembly and annotation. Brain, liver, heart, spleen, skeletal muscle, and gonad tissues were collected, flash-frozen in liquid nitrogen, and stored at −80°C. High-molecular-weight DNA extracted from muscle tissue was used for PacBio HiFi, Oxford Nanopore ultra-long (ONT), Hi-C, and genome survey sequencing. ONT libraries were prepared using the SQK-LSK114 kit and sequenced on R10.4.1 flow cells.[21] Base calling was performed using Dorado v0.9.1 with the dna_r10.4.1_e8.2_400bps_sup@v5.0.0 model. PacBio libraries with 15–20 kb inserts were sequenced on the PacBio Revio platform in circular consensus sequencing (CCS) mode to generate highly accurate HiFi reads. Hi-C libraries were constructed following,[22] involving formaldehyde cross-linking, MboI digestion, biotin labeling, proximity ligation, and paired-end sequencing on the DNBSEQ platform. Total RNA was extracted from brain, liver, heart, spleen, skeletal muscle, and gonad tissues using TRIzol reagent (Invitrogen). Subsequently, paired-end short-read sequencing was performed on the NovaSeq 6000 platform, and Iso-Seq long-read sequencing was conducted on the PacBio Revio platform.

Genome assembly and annotation

Ultra-long ONT reads were filtered using SeqKit v2.9.0 to retain reads ≥ 70 kb with Q-scores ≥ 15 and assembled using Hifiasm v0.24.0 with ONT support enabled.[23,24] Hi-C reads were mapped to the draft assembly using Juicer v1.5.7,[25] followed by scaffolding and misassembly correction using 3D-DNA and manual inspection in Juicebox. To improve assembly continuity, additional assemblies generated from more stringent ONT datasets (> 80

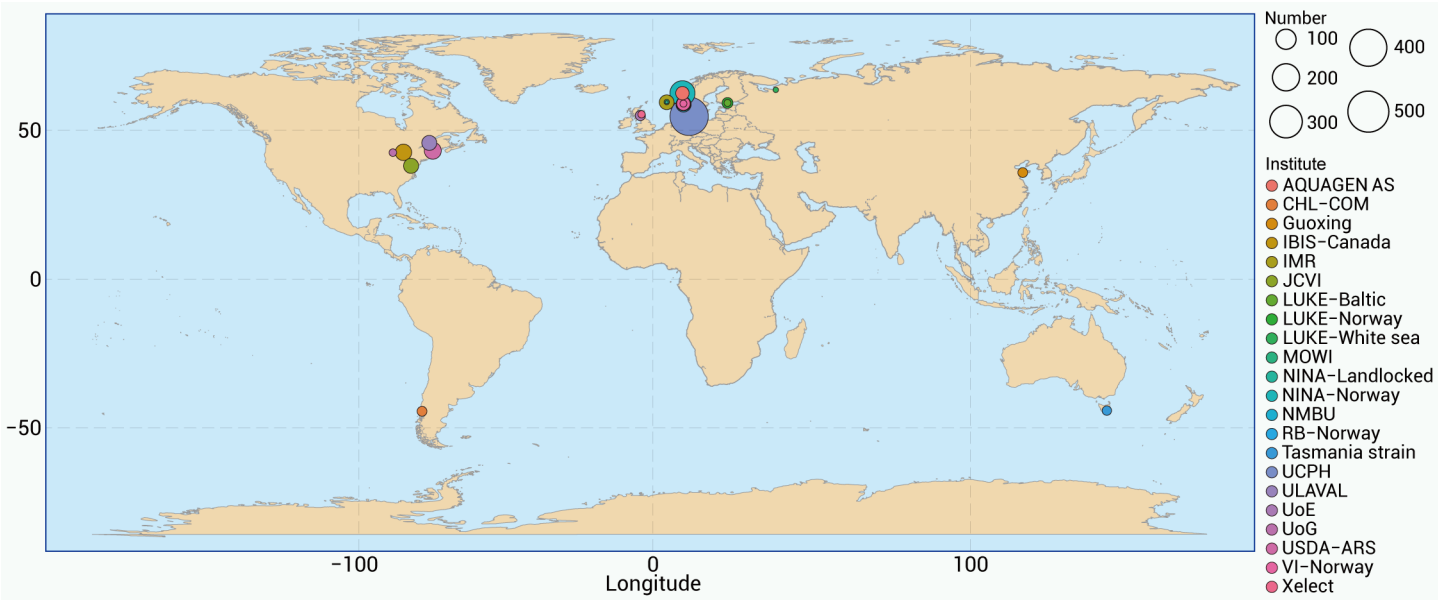


Figure 1. Geographic distribution map of all collected sample data.

kb/Q18 and > 90 kb/Q18) were incorporated for error correction and gap filling using Quartet v1.2.5.[26] Chromosome-scale scaffolds from alternative assemblies were used to replace corresponding regions when necessary. Final polishing was performed with NextPolish v0.2.1 using PacBio HiFi and Illumina short reads.[27] Telomeric sequences were identified using TeloExplorer implemented in Quartet.

Repetitive sequences were annotated using combined de novo and structure-based approaches. Long terminal repeat retrotransposons (LTR-RTs) were identified using LTRfinder and refined using LTR_retriever. Additional repeat families were generated using RepeatModeler v2.0.5, and redundant sequences were removed using CD-HIT v4.8.1. The resulting non-redundant repeat library was used by RepeatMasker v4.1.7 to annotate genome-wide repetitive elements.

Genome annotation integrated ab initio prediction, homology-based evidence, and transcriptomic support. Initial gene prediction was conducted using BRAKER v3.0.8,[28] incorporating protein homology from *Danio rerio*, *Oncorhynchus kisutch*, *Epinephelus lanceolatus*, and *Centropristis striata*, together with RNA-seq data. To maintain compatibility with previous genome resources, annotations from the Ssa_v3.1 reference genome were transferred using Liftoff.[29] Low-confidence gene models, including unsupported single-exon genes and genes containing internal stop codons, were removed using GFFread.[30] To further improve annotation quality, full-length Iso-Seq transcripts were processed using the IsoSeq3 pipeline. Following primer removal using lima and transcript refinement, high-quality isoforms were aligned to the genome using pbmm2 and integrated into PASA v2.5.3 to refine exon-intron structures and annotate untranslated regions (UTRs).

Variant detection and population genomic analyses

Whole-genome resequencing data from 1,399 Atlantic salmon individuals (Figure 1 & Table S1) were processed using FastQC (https://github.com/s-andrews/fastqc) and Trimmomatic v0.39.[31] Clean reads were aligned to the Ssa_v3.1 reference genome using BWA-MEM,[32] and duplicate reads were removed using Picard implemented in GATK v4.2.3.[33] Variants were called using GATK HaplotypeCaller and jointly genotyped using GenotypeGVCFs. Variant quality score recalibration (VQSR) was applied, and SNPs were further filtered using the following criteria: QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, and SOR > 3.0. Biallelic SNPs with sequencing depth ≥ 5 were retained using VCFtools,[34] followed by additional filtering using PLINK v1.90 to exclude variants with minor allele frequency < 0.01 or call rates < 0.8.[35] After quality control, approximately 42 million high-confidence SNPs were retained for downstream analyses.

Selective sweeps within populations were identified using SweeD based on composite likelihood ratio (CLR) tests.[36] CLR values were calculated in 10-kb windows across each chromosome, and candidate regions were defined as the top 1% of the empirical distribution. Population differentiation was assessed using Weir and Cockerham's Fst statistic implemented in VCFtools with 40-kb sliding windows and 10-kb step sizes. Cross-population selective sweeps between the Guoxing strain and Norwegian wild populations were further evaluated using XP-CLR with 40-kb windows, 10-kb steps, and a minimum of five SNPs per window.[37]

Gene enrichment analysis

According to genome annotation, candidate genes under selection were defined as those located within or overlapping regions showing selection signatures. To understand the biological significance of selection signatures, we performed gene ontology (GO) enrichment analysis on differentially expressed genes using the GOstats tool from the Bioconductor package.[38] GO terms with a corrected P-value < 0.05 were considered significantly enriched. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using Clusterprofiler 4.0.[39] KEGG pathways with a Benjamini-Hochberg FDR corrected P-value < 0.05 were considered enriched for differential expression.

RESULTS AND DISCUSSIONS

Genome assembly and telomeres identification

In this study, we integrated 150.22 G of PacBio HiFi long reads (≥ 5 kb), 236.15 G of ultra-long Oxford Nanopore reads (≥ 70 kb, ≥ Q15), and 524.85 G of Hi-C data for T2T genome assembly (MGB-Salmon-T2T-v1.0). The total genome size of the T2T is 2,981.11 Mb, representing an expansion relative to the 2,756.58 Mb observed in Ssa_v3.1 (Figure 2A). A total of 15 T2T chromosomes were assembled (Table S2), with detailed annotation indicating which chromosomes are complete T2T assemblies, which are incomplete T2T assemblies, and the locations of telomeric sequences. This increase is accompanied by significantly enhanced contiguity: the contig N50 and N90 of the T2T assembly reached 101.69 Mb and 45.84 Mb, respectively, compared to only 28.05 Mb and 1.04 Mb in Ssa_v3.1. The scaffold N50 also improved from 96.06 Mb in Ssa_v3.1 to 114.45 Mb in the T2T assembly, further reflecting the higher assembly quality (Table 1). BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis of genome completeness indicates a slight improvement in the T2T version, with a genome BUSCO score of 98.8%, compared to 98.5% in Ssa_v3.1. The proportions of single-copy (S) duplicated (D), fragmented (F), and missing (M) orthologs are largely consistent between assemblies, with the T2T version showing slightly higher single-copy (39.2%) and duplicated (59.6%) BUSCOs, and marginally lower fragmented (0.7%) and missing (0.6%) ones. Gene annotation revealed a higher

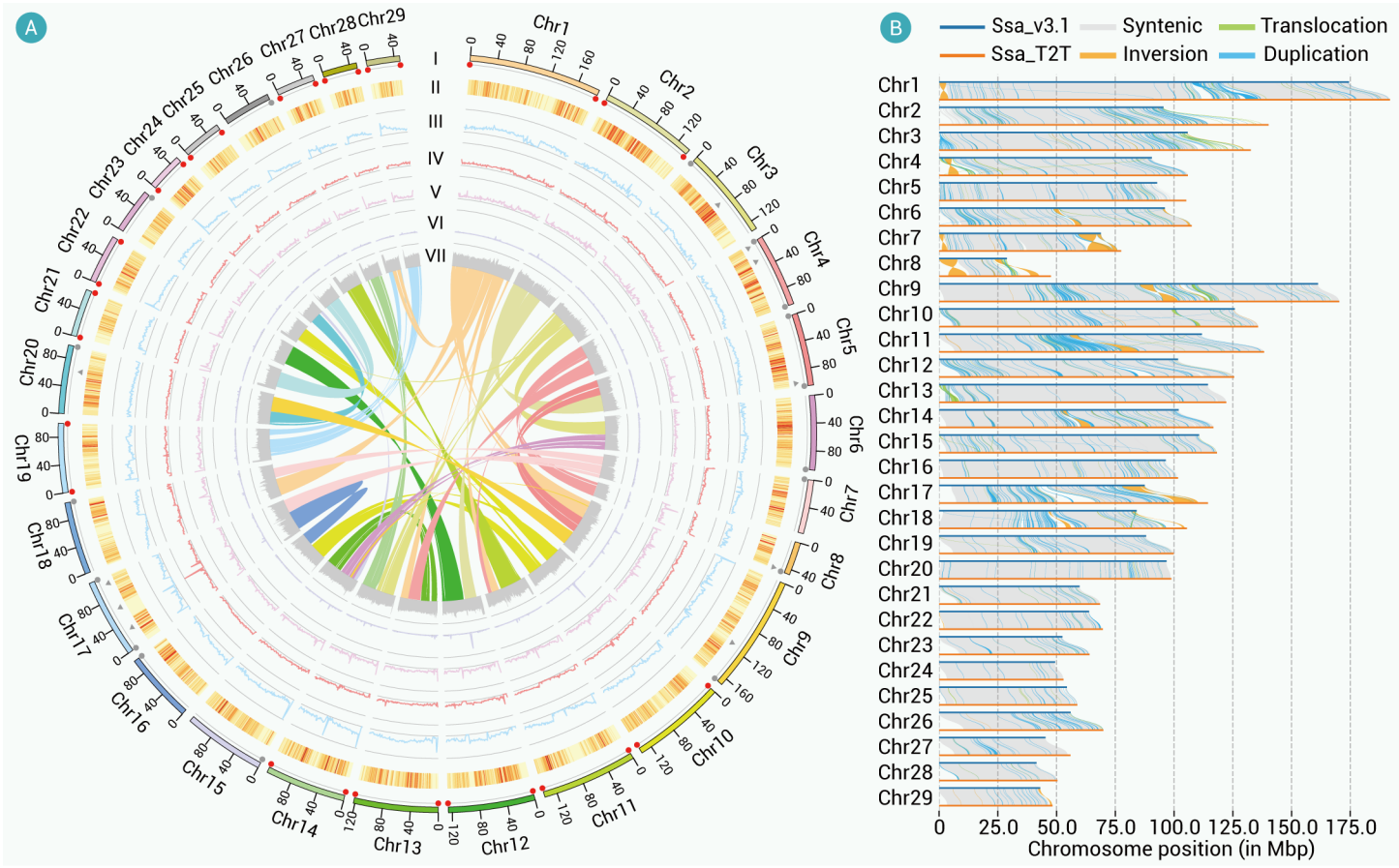


Figure 2. (A) Genomic features and homology distribution map; (I) Telomeres: red dots indicate telomeres in T2T chromosomes, while gray dots represent telomeres in non-T2T chromosomes; gray blocks denote gaps, with each gap represented by N = 500; (II) Gene density; (III) DNA transposable elements; (IV) LINE elements; (V) LTR elements; (VI) SINE elements; (VII) GC content (range: min = 0.30, max = 0.55); The inner links represent collinear blocks across the 29 chromosomes; (B) Comparison of different genome versions.

Table 1. The summary of interspersed repeat contents in the genome assembly.

Type	MGB-Salmon-T2T-v1.0		Ssa_v3.1	
	Length (bp)	% in genome	Length (bp)	% in genome
SINE	51438910	1.73%	46535559	1.69%
LINE	318373011	10.68%	313285627	11.36%
LTR	389650547	13.07%	359718497	13.05%
DNA transposons	624897878	20.96%	616268556	22.36%
Unclassified	112473621	3.77%	109057239	3.96%
Rolling-circles	1353991	0.05%	1331029	0.05%
Small RNA	32754602	1.10%	22424581	0.81%
Satellites	28140261	0.94%	20346581	0.74%
Simple repeats	171542547	5.75%	78047592	2.83%
Low complexity	12228775	0.41%	12388787	0.45%

number of predicted protein-coding genes in the T2T assembly (45,639) compared to Ssa_v3.1 (43,000). Annotated completeness remained unchanged at 98.8% for both assemblies. The distribution of annotated genes among categories was also similar, with the T2T assembly having slightly more single-copy (41.4%) and slightly fewer duplicated (57.4%) genes relative to Ssa_v3.1 (Table S3).

Despite an overall high level of synteny across all 29 chromosomes, we identified substantial structural rearrangements when the Atlantic salmon T2T genome was compared to Ssa_v3.1. We identified numerous inversions, translocations, and duplications across multiple chromosomes. The largest 10.6 Mb inversion on Chr7 (Chr7: 63,142,492–73,829,407 in the new assembly, corresponding to Chr7: 59,240,931–66,982,771 in Ssa_v3.1), are highly likely to reflect corrections of previous assembly or scaffolding errors rather than definitive biological variation. Notably, in the corresponding Ssa_v3.1 region, two adjacent 100-N gaps (Chr7:59,240,932–59,241,031 and Chr7:67,025,904–67,026,003) are located near the inversion boundaries, strongly suggesting potential mis-scaffolding orientation in the previous reference. A similar pattern was also observed for the Chr14 inversion breakpoint (breakpoint in Ssa_v3.1: 51,656,445; Gap region in Ssa_v3.1: 51,656,334–51,656,433) (Figure 2B).

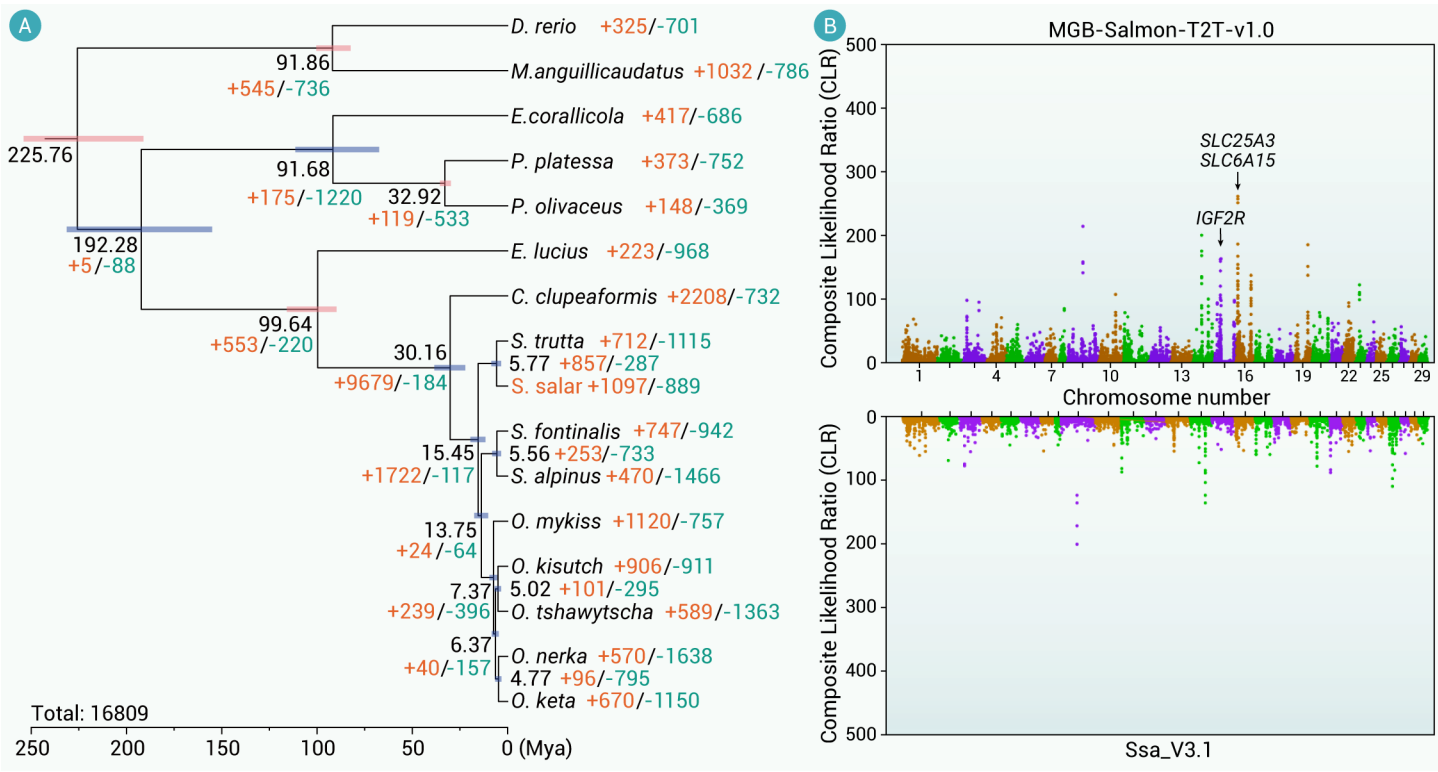


Figure 3. (A) Evolutionary and phylogenetic analysis of Atlantic salmon; (B) Noval selection signatures identified using the T2T genome in the Guoxing population.

Repetitive sequence annotation

To evaluate improvements in the repeat landscape of our newly assembled Atlantic salmon genome MGB-Salmon-T2T-v1.0, we compared its repetitive element content and subclass composition with those of the reference genome Ssa_v3.1 (Table 1 & Table S4).

The total length of annotated repeat sequences in the T2T assembly (1716.95 Mb) was substantially greater than that of the Ssa_v3.1 version (1558.40 Mb), reflecting an increase of 158.55 Mb. Among transposable elements (TEs), SINEs, LINEs, rolling-circle (RC), and unclassified elements exhibited modest increases in length in the T2T assembly relative to Ssa_v3.1.

Meanwhile, DNA transposons were slightly longer in the T2T assembly by approximately 8.63 Mb (Ssa_v3.1: 616.27 Mb; MGB-Salmon-T2T-v1.0: 624.90 Mb). LTR retrotransposons showed a more pronounced increase by approximately 29.93 Mb (Ssa_v3.1: 359.72 Mb; MGB-Salmon-T2T-v1.0: 389.65 Mb). Notably, the detection of simple repeats showed significant improvement, with their annotated length increasing from 78.05 Mb in Ssa_v3.1 to 171.54 Mb in the T2T assembly, with an increase of 93.49 Mb.

Gene structure and functional annotation

The final annotation comprised 45,639 genes, of which 33,934 were supported by full-length Iso-Seq data, and 27,742 genes were structurally refined, demonstrating strong transcript-level support. Comparative analysis with Ssa_v3.1 identified 4,693 genes unique to the T2T assembly; among these, 2,093 were supported by Iso-Seq evidence, suggesting that a substantial proportion of these newly identified genes are biologically meaningful rather than annotation artifacts.

To further assess the quality and completeness of functional annotation, we compared the annotated gene sets of the Atlantic salmon T2T genome with those of the reference genome Ssa_v3.1 using five major databases: NR, Swiss-Prot, KEGG, GO, and InterPro. Overall, 44,972 protein-coding genes in the T2T assembly were annotated by at least one database. Notably, 24,431 genes were supported by all five databases, compared with 24,293 genes in Ssa_v3.1, indicating improved annotation consistency and higher confidence in gene functional assignment. Consistent with these results, the T2T genome showed increased numbers of annotated genes across all individual databases, including GO (30,826 vs. 29,997), InterPro (44,046 vs. 42,300), KEGG (30,995 vs. 30,366), NR (44,427 vs. 42,517), and Swiss-Prot (41,694 vs. 40,790). Together, these improvements reflect enhanced gene structure accuracy, more comprehensive functional characterization, and improved resolution of protein-coding regions in the T2T assembly.

Evolution and phylogeny of Atlantic salmon

To infer the evolutionary relationships among *S. salar* and other representative salmonid species, we selected single-copy orthologous sequences identified by Orthofinder v3.0.1b1 and constructed a super alignment matrix (979,140 aa) by merging the multiple sequence alignments of these genes. Using this matrix, a maximum likelihood phylogenetic tree was constructed with RAxML v8.2.12 under the JTT+I+G4 model with 1,000 bootstrap replications to ensure robust support for internal nodes. Subsequently, divergence time estimation was performed using the mcmctree module of the PAML package v4.9, calibrated with divergence times from TimeTree and fossil record (MRCA of the clade comprising Pleuronectidae and Paralichthyidae was constrained with a minimum age of 29.62 Mya and a 95% soft upper boundary at 35.3 Mya, based on †*Oligofluorenes germanicus*). Gene family expansion and contraction were assessed using CAFE5 with default parameters,[40] based on gene family classifications from OrthoFinder and the inferred phylogenetic relationships. The estimated divergence times were 30.16 Mya between *S. salar* and *Coregonus clupeaformis*, 15.45 Mya between *S. salar* and the (*Salvelinus* + *Oncorhynchus*) lineage, and 5.77 Mya between *S. salar* and *S. trutta*. These divergence events were accompanied by extensive expansions and contractions of gene families, reflecting a complex and dynamic evolutionary trajectory for Salmonidae (Figure 3A).

Selection signatures

Using the T2T genome assembly, we identified a greater number of candidate selective sweep windows compared to the Ssa_v3.1 reference. Specifically, 2,829 windows in the T2T assembly exhibited SweeD composite likelihood ratio (CLR) values above log 7.65, corresponding to the top 1% threshold, whereas 2,753 windows exceeded the top 1% threshold (log CLR > 5.51) in Ssa_v3.1. Among these, 1,114 windows were shared between the two assemblies, while 1,715 and 1,639 windows were uniquely identified in the MGB-Salmon-T2T-v1.0 and Ssa_v3.1 assemblies, respectively, indicating

improved resolution of selective signals in the T2T genome. Gene Ontology (GO) enrichment analysis of genes located within the candidate selective sweep regions identified nine significantly enriched GO terms across the biological process, cellular component, and molecular function categories. These terms included regulation of proteolysis (GO:0030162), extracellular matrix (GO:0031012), and protein sumoylation (GO:0016925), highlighting biological processes and cellular functions that may have been targets of selection. KEGG pathway analysis further identified the insulin signaling pathway (KEGG ID:sasa04910) as the only significantly enriched pathway, involving five key genes: *insr, ptp1b, ampk, foxo1,* and *mek1/2* (Figure S2)

To minimize potential bias associated with single-method detection, we further validated these selective sweep regions using complementary approaches, including XP-CLR and Fst analyses. A total of 2,500 candidate windows were identified by XP-CLR and Fst based on the top 1% thresholds for each method. Among these, 1,330 windows exhibited at least 20 kb overlap between the two approaches, and 828 windows exhibited overlap between the three approaches (Table S5). The consistency among these independent methods supports the robustness of the detected selection signals (Figure S1). Notably, the T2T assembly enabled the identification of additional candidate genes within selective regions, including *IGF2R* and *SLC25A3 and SLC6A15* (Figure 3B), which exhibit high expression in brain (Table S6) and may play roles in growth regulation and domestication in China.

CONCLUSION

Using PacBio HiFi, ultra-long ONT, and Hi-C sequencing, we generated a T2T genome with a contig N50 of 101.69 Mb and scaffold N50 of 114.45 Mb, A total of 15 T2T chromosomes were assembled, representing a major advance in genome completeness and continuity for Atlantic salmon. This T2T genome resolves previously collapsed repeat-rich, duplicated regions, and telomeric regions. We identified 45,639 protein-coding genes (BUSCO 98.8%), and population-level analyses uncovered previously unrecognized selective sweep regions and key candidate genes associated with growth traits that are central to breeding improvement. The high-resolution T2T assembly establishing a species-wide catalogue of genetic variants is an essential first step toward functional genome annotation and forms the foundation for the ongoing international FarmGTEx project, which aims to link genetic variation with regulatory activity, tissue-specific expression, and phenotypic outcomes.

REFERENCES

1. FAO (2020). The state of world fisheries and aquaculture 2020. Sustainability in action (FAO). DOI:10.4060/ca9229en
2. Kobayashi M., Msangi S., Batka M., et al. (2015). Fish to 2030: The role and opportunity for aquaculture. *Aquacult. Econ. Manage.* **19**:282–300. DOI:10.1080/13657305.2015.994240
3. Gjedrem T. and Robinson N. (2014). Advances by selective breeding for aquatic species: A review. *Agric. Sci.* **5**:1152–1158. DOI:10.4236/as.2014.512125
4. Afewerki S., Asche F., Misund B., et al. (2023). Innovation in the Norwegian aquaculture industry. *Rev. Aquac.* **15**:759–771. DOI:10.1111/raq.12755
5. Garlock T., Asche F., Anderson J., et al. (2020). A global blue revolution: Aquaculture growth across regions, species, and countries. *Rev. Fish. Sci. Aquac.* **28**:107–116. DOI:10.1080/23308249.2019.1678111
6. Iversen A., Asche F., Hermansen O., et al. (2020). Production cost and competitiveness in major salmon farming countries 2003-2018. *Aquaculture* **522**:735089. DOI:10.1016/j.aquaculture.2020.735089
7. Osmundsen T.C., Olsen M.S., Gauteplass A., et al. (2022). Aquaculture policy: Designing licenses for environmental regulation. *Mar. Policy* **138**:104978. DOI:10.1016/j.marpol.2022.104978
8. Li A., Yu F., Si G., et al. (2017). Long-term temperature variation of the Southern Yellow Sea Cold Water Mass from 1976 to 2006. *Chin. J. Oceanol. Limnol.* **35**:1032–1044. DOI:10.1007/s00343-017-6037-1
9. Guo J., Yuan H., Song J., et al. (2020). Hypoxia, acidification and nutrient accumulation in the Yellow Sea Cold Water of the South Yellow Sea. *Sci. Total Environ.* **745**:141050. DOI:10.1016/j.scitotenv.2020.141050
10. Dong S.L. (2019). Research progress and prospects of large salmonid aquaculture in the Yellow Sea Cold Water Mass. *Period. Ocean Univ. China* **49**:1–9. DOI:10.16441/j.cnki.hdxb.20180303
11. Lutfi E., Berge G., Baeverfjord G., et al. (2022). Increasing dietary levels of the omega-3 long-chain polyunsaturated fatty acids, EPA and DHA, improves the growth, welfare, robustness, and fillet quality of Atlantic salmon in sea cages. *Br. J. Nutr.* **129**:10–28. DOI:10.1017/S0007114522000642
12. Houston R.D., Bean T.P., Macqueen D.J., et al. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nat. Rev. Genet.* **21**:389–409. DOI:10.1038/s41576-020-0227-y
13. Song H., Dong T., Yan X., et al. (2023). Genomic selection and its research progress in aquaculture breeding. *Rev. Aquac.* **15**:274–291. DOI:10.1111/raq.12716
14. Yanez J.M., Barria A., Lopez M.E., et al. (2023). Genome-wide association and genomic selection in aquaculture. *Rev. Aquac.* **15**:645–675. DOI:10.1111/raq.12750
15. Lien S., Koop B.F., Sandve S.R., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**:200–205. DOI:10.1038/nature17164
16. Fang L., Teng J., Lin Q., et al. (2025). The farm animal genotype-tissue expression (FarmGTEx) project. *Nat. Genet.* **57**:786–796. DOI:10.1038/s41588-025-02121-5
17. Pineda P., Macphillamy C., Ren Y., et al. (2025). Insights into natural neocentromere evolution from a cattle T2T X chromosome. *Nat. Commun.* **16**:10745. DOI:10.1038/s41467-025-65778-w
18. Zong W., Chen L., Zhang D., et al. (2025). Two telomere-to-telomere pig genome assemblies and pan-genome analyses provide insights into genomic structural landscape and genetic adaptations. *iMeta* **4**:e70013. DOI:10.1002/imt2.70013
19. Liu R., Yang X., Zhang Y., et al. (2025). A telomere-to-telomere pangenome reveals structural variations as key drivers of complex traits in chickens. DOI:10.21203/rs.3.rs-7302767/v1
20. Luo L.-Y., Wu H., Zhao L.-M., et al. (2025). Telomere-to-telomere sheep genome assembly identifies variants associated with wool fineness. *Nat. Genet.* **57**:218–230. DOI:10.1038/s41588-024-02037-6
21. Sereika M., Kirkegaard R.H., Karst S.M., et al. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat. Methods* **19**: 823-826. DOI:10.1038/s41592-022-01539-7
22. Han W., Wu S., Ding H., et al. (2023). Improved chromosomal-level genome assembly and re-annotation of leopard coral grouper. *Sci. Data* **10**:156. DOI:10.1038/s41597-023-02051-z
23. Shen W., Le S., Li Y., et al. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**:e0163962. DOI:10.1371/journal.pone.0163962
24. Cheng H., Concepcion G.T., Feng X., et al. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**:170–175. DOI:10.1038/s41592-020-01056-5
25. Durand N.C., Shamim M.S., Machol I., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**:95–98. DOI:10.1016/j.cels.2016.07.002
26. Lin Y., Ye C., Li X., et al. (2023). quarTeT: A telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic. Res.* **10**:uhad127. DOI:10.1093/hr/uhad127
27. Hu J., Wang Z., Liang F., et al. (2024). NextPolish2: A repeat-aware polishing tool for genomes assembled using HiFi long reads. *Genomics Proteomics Bioinformatics* **22**:qzad009. DOI:10.1093/gpbjnl/qzad009
28. Gabriel L., Bruna T., Hoff K.J., et al. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**:769–777. DOI:10.1101/gr.278090.123
29. Shumate A. and Salzberg S.L. (2021). Liftoff: Accurate mapping of gene annotations. *Bioinformatics* **37**:1639–1643. DOI:10.1093/bioinformatics/btaa1016
30. Pertea G. and Pertea M. (2020). GFF Utilities: GffRead and GffCompare. DOI:10.12688/f1000research.23297.1
31. Bolger A.M., Lohse M. and Usadel B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. DOI:10.1093/bioinformatics/btu170
32. Li H. and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI:10.1093/bioinformatics/btp324
33. McKenna A., Hanna M., Banks E., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**:1297–1303. DOI:10.1101/gr.107524.110
34. Danecek P., Auton A., Abecasis G., et al. (2011). The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. DOI:10.1093/bioinformatics/btr330
35. Purcell S., Neale B., Todd-Brown K., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**:559–575. DOI:10.1086/519795
36. Pavlidis P., Zivkovic D., Stamatakis A., et al. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**:2224–2234. DOI:10.1093/molbev/mst112
37. Chen H., Patterson N. and Reich D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* **20**:393–402. DOI:10.1101/gr.100545.109
38. Falcon S. and Gentleman R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**:257–258. DOI:10.1093/bioinformatics/btl567
39. Wu T., Hu E., Xu S., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**: 100141. DOI:10.1016/j.xinn.2021.100141
40. De Bie T., Cristianini N., Demuth J.P., et al. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271. DOI:10.1093/bioinformatics/btl097

FUNDING AND ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (2024YFD2400902), the Key R&D Program of Shandong Province, China (2025SFGC0401), the Blue Seed Industry Science and Technology Innovation Project (QDLYY-2024011). We acknowledge the support of the High-performance Biological Supercomputing Center at Ocean University of China for this research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Z.M Bao and J.J Hu designed the study. X.F Yu and Y.X Liu collected the data. X.F Yu, B.J Zhao, Z.Y Wang, W.T Han performed the analysis. X.F Yu, B.J Zhao, W.T Han wrote the manuscript. S Wang, L.Z Fang, Y.F Wang and X.H Zhu supervised the project. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

All procedures complied with institutional ethical guidelines and were approved by the Institutional Animal Care and Use Committee of Ocean University of China (Approval No. OUC-AE-2025-172).

DATA AND CODE AVAILABILITY

The raw data generated in this study are available in the National Center for Biotechnology Information (Bioproject: PRJNA1423218). Genome annotation files have also been deposited on Figshare (https://figshare.com/s/d92a9acf2608eeeae77d).