BMC Genomics

RESEARCH

Open Access

Genome refinement in Pacific bluefin tuna (*Thunnus orientalis*): chromosome-level assembly using hybrid linkage data

Xi Fu[1], Daiki Saka[1], Kazutoshi Yoshitake[2], Taiju Saito[3], Milos Havelka[3], Osamu Nishimiya[3], Takahiro Matsubara[3], Rie Goto[3], Toshinao Ineno[4] and Shuichi Asakawa[1*]

Abstract

Background High-quality reference genomes are fundamental for selective breeding and genetic engineering to improve aquaculture productivity. Here, we report an improved chromosome-level genome assembly of Pacific bluefin tuna (*Thunnus orientalis*), a commercially valuable species consumed worldwide. By integrating genetic linkage information from interspecific hybrids of *Euthynnus affinis* and *T. orientalis*, we anchored and oriented *T. orientalis* scaffolds into 24 pseudochromosomes.

Results The updated assembly achieved a scaffold N50 of 34.19 Mb, representing a notable improvement from 13.3 Mb, and contained 22,640 predicted protein-coding genes.

Conclusions This resource provides a robust genomic foundation for future genetic and genomic studies to advance the selective breeding of *T. orientalis* in aquaculture.

Keywords Pacific bluefin tuna, Aquaculture, Genome annotation, Linkage map, Genome improvement

Background

Tuna (*Thunnus* spp.) is among the most valuable fish for human consumption, with bluefin tuna being particularly prized for its high market demand and value [1]. Full-cycle aquaculture of bluefin tuna has been achieved through advances in larval rearing and juvenile grow-out technologies [2], yet production and sustainability challenges remain [3].

Genetic improvement through selective breeding and genetic engineering has emerged as a promising strategy for enhancing production efficiency in fish farming [4]. Thus, a high-quality species genome assembly is critical to support these genetic improvement efforts. Although recent progress in sequencing technologies has provided valuable genomic insights applicable to basic fish science [5] and aquaculture practices [4], genomic information on tuna remains limited. For example, as a model of tuna species, the current Pacific bluefin tuna, *Thunnus orientalis* (NCBI: txid8238) genome assembly (GenBank accession: GCA_021601225.2, hereafter referred to as the "TUNAREF2023") [6] remains fragmented (scaffold N50 = 13 Mb) and is not assembled to the chromosome level. High-resolution linkage maps are essential tools in

*Correspondence:
Shuichi Asakawa
asakawa@g.ecc.u-tokyo.ac.jp
[1]Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan
[2]School of Marine Biosciences, Kitasato University, Kanagawa 252-0373, Japan
[3]South Ehime Fisheries Research Center, Ehime University, Ainan 798- 4205, Japan
[4]Aquaculture Research Institute, Shingu Station, Kindai University, Wakayama 647-1101, Japan

genetics and genomics, and support applications such as comparative genome analysis and genome scaffolding [7]. One of the main challenges in linkage analysis is accurately distinguishing haplotypes, which typically requires high-depth sequencing (over 10× coverage) across numerous individuals to reliably resolve heterozygosity. In contrast, haploid individuals, such as those produced through gynogenesis or androgenesis, require only a single read per polymorphic site to determine the haplotype, thereby substantially reducing the sequencing effort required [8]. Similarly, in interspecific hybrids, sequence divergence between parental genomes enables the clear assignment of sequencing reads to either maternal or paternal origin. This haploid-equivalent resolution of single-nucleotide polymorphisms (SNPs) facilitates straightforward haplotype determination and allows the independent construction of linkage maps for both diploid parents [9, 10].

Here, we leveraged F1 interspecific hybrids between *T. orientalis* and *Euthynnus affinis* to anchor and orient scaffolds of the TUNAREF2023 assembly onto pseudochromosomes, producing a chromosome-level genome assembly. The final assembly comprised 192 scaffolds, including 24 chromosome-scale scaffolds, with a scaffold N50 of 34.19 Mb and 22,640 predicted protein-coding genes. Of the 580 scaffolds in the TUNAREF2023 assembly, 421 were successfully integrated into the pseudochromosomes using the SELDLA approach, whereas the remaining scaffolds, for which neither orientation nor chromosomal position could be reliably determined, were retained as unanchored sequences in our assembly. This improved genomic resource provides a robust foundation for future studies on the genomics, selective breeding, and conservation of *T. orientalis*.

Materials and methods

Generation and sequencing of hybrids

Interspecific hybrids of *E. affinis* and *T. orientalis* were generated to establish haploid linkage evidence. Specifically, eggs from a single *E. affinis* female were artificially fertilized with cryopreserved *T. orientalis* sperm. After 1–2 h, eggs undergoing cleavage were selected and transferred to a hatching tank maintained at 24 °C. Approximately 36–40 h post fertilization, 200 hatched individuals were collected and preserved in 100% ethanol [11]. The animal study protocol was reviewed and approved by the Animal Care Committee of Ehime University and the Animal Experimentation Committee of the Graduate School of Agricultural and Life Sciences, University of Tokyo.

Genomic DNA was extracted from each F1 hybrid using a NucleoSpin Tissue XS kit (Macherey-Nagel, Düren, Germany). Sequencing libraries were constructed using the Nextera DNA Library Preparation and Nextera Index kits (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Sequencing was performed on an Illumina HiSeq X Ten instrument to generate 150 bp paired-end reads. Quality-filtered reads were then aligned with BWA mem v.0.7.15 [12] using default parameters to a reference genome generated by combining the *E. affinis* genome (GenBank accession: GCA_019973915.1) with the "TUNAREF2023" reference. The aligned reads were sorted using SAMtools v.1.17 [13], and variant calling was conducted using BCFtools [13], using the *mpileup* and *call* commands with multiallelic-caller and variant-only flags (-mv).

Genome elongation using linkage analysis

Informative SNPs were extracted from specific regions by omitting regions where reads were mapped to male *T. orientalis* and female *E. affinis* genomes. Next, the SNPs were confirmed to be heterozygous in the *T. orientalis* genome. SNPs present in less than 30% of hybrids were removed, with those in which the minor allele was found in less than 30% of the hybrids (i.e., with the major allele appearing in more than 70% of the hybrids). Heterozygous SNPs in more than 10% of the individuals were suspected to be multi-copy genes and were excluded. The extracted SNPs were input into SELDLA v.2.3.0 for linkage analysis using the following options: DP = 0 --GQ = 0–r 1000 --cs = 2 --mode=crossbreed.

Repeat annotation

Homology-based and *de novo* strategies have been used to identify repetitive sequences in the SELDLA-extended *T. orientalis* genome. A *de novo* repeat library was generated using RepeatModeler v.2.0.5 [14] with default parameters. The SELDLA-extended *T. orientalis* assembly was subsequently screened for repeats using RepeatMasker v.4.1.7 in two runs using (i) a *de novo*-generated repeat library and (ii) the Dfam database (release 3.8) under the vertebrate category [15, 16]. The results of both runs were integrated into a repetitive sequence library. Using this repeat database, repetitive sequences were identified by homology searches using RepeatMasker and rmBLASTn.

Gene prediction and functional assignment

Gene predictions were conducted using the BRAKER3 pipeline with RNA-seq and protein data [17]. In detail, RNA sequencing datasets for *T. orientalis* (Supplementary Material 4 -Table S1) were obtained from the Sequence Read Archive database [18]. The RNA-Seq datasets were quality-controlled using Trim_Galore v.0.6.10 (https://github.com/FelixKrueger/TrimGalore). The trimmed reads were aligned to the SELDLA-extended *T. orientalis* genome using STAR v.2.7.11b (STAR, RRID: SCR_004463) [19]. For homology-based gene

prediction, the full genome-retrieved proteins of the four tuna species (Supplementary Material 4 - Table S2, upper part) were retrieved from the Ensembl database [20] or NCBI RefSeq [21]. The BRAKER3 pipeline was used, in which GeneMark-ETP integrates both transcriptional and protein-based hints to identify 'high-confidence' genes. These were subsequently used by AUGUSTUS v.3.5.0 for gene model prediction. Finally, the gene predictions from AUGUSTUS and GeneMark-ETP were merged and refined using TSEBRA [22]. BRAKER3 was run using the following additional parameters: -gff3-min_contig = 10000.

For functional annotation of the predicted genes, protein sequences generated by BRAKER were queried against the Swiss-Prot and NCBI non-redundant (nr) protein databases (downloaded November 2024) using BLASTP with a minimum sequence identity threshold of 30% and an E-value cutoff of 1e-6. Protein domains were predicted using InterProScan v.5.71 [23] with default parameters, and the associated Gene Ontology (GO) terms were retrieved from the corresponding InterPro entries. Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) annotations were assigned using KofamKOALA [24], applying an E-value threshold of 1e-6. To categorize GO terms into three major functional domains (biological processes, molecular functions, and cellular components), the ontology structure was provided in the go-basic.obo file was parsed using the GOATOOLS library [25].

Quality assessment

To assess the completeness of the SELDLA-extended *T. orientalis* assembly, two approaches were used: (1) evaluation based on complete Core Vertebrate Genes (CVGs) [26] and (2) BUSCO. First, 233 complete CVGs were confirmed using gVolante [27]. Second, the actinopterygii_obd12 lineage dataset (creation date: 2025-04-11, number of genomes: 75, number of BUSCOs: 7,207) was used to run BUSCO v.5.8.2 [28] in genome mode for the SELDLA-extended *T. orientalis* assembly. To assess the accuracy of scaffold arrangement in the 24 pseudochromosomes, the genomic locations of single-copy orthologs were compared between *T. orientalis* and Japanese medaka (*Oryzias latipes*) (GenBank accession: GCF_002234675.1). The comparison was visualized using Circos v0.69- 8 (Circos, RRID: SCR_011798) [29]. Only complete and unduplicated orthologs co-identified using BUSCO across the two species were selected for comparison.

Comparative genomic analysis of orthologs

Orthologous relationships were inferred using OrthoFinder v.3.0.1 b1 [30] by comparing the predicted genes in *T. orientalis* with those from the following eight representative fish species listed in Table S2 (Supplementary Material 4): zebrafish (*Danio rerio*), climbing perch (*Anabas testudineus*), tiger tail seahorse (*Hippocampus comes*), Japanese medaka (*Oryzias latipes*), Yellowtail amberjack (*Seriola lalandi*), three-spined stickleback (*Gasterosteus aculeatus*), southern bluefin tuna (*Thunnus maccoyii*), and yellowfin tuna (*Thunnus albacares*). The results were visualized using ComplexUpset [31].

Results

Sequencing summary

Sequencing of interspecific hybrid individuals generated a total of 390 Gb of reads. When the reads derived from hybrid individuals were mapped separately to each parental genome, the mean mapping rates were 90.28% for the *E. affinis* genome and 90.52% for the *T. orientalis* reference genome (TUNAREF2023), indicating no strong mapping bias toward either parent. When mapped to the combined *E. affinis* and TUNAREF2023 reference genome, the overall mapping rate increased to 94.2%,reflecting the mixed genomic origin of the hybrid reads. Reads derived from each parental genome preferentially align to their corresponding homologous sequences, reducing the impact of interspecific divergence on read mapping.

Chromosome-level assembly using linkage map data

A total of 4,048,459 SNPs specific to *T. orientalis* were used for linkage analysis using SELDLA. The final SELDLA-extended *T. orientalis* assembly (GenBank accession No. GCA_055472215.1) comprised 192 scaffolds, including 24 chromosome-scale scaffolds larger than 10 Mb, corresponding to the expected karyotype of *T. orientalis* (n = 24) [32] (Supplementary Material 4 - Table S3). The remaining 168 scaffolds were shorter sequences that could not be anchored to the chromosome-scale scaffolds and are therefore retained as unplaced scaffolds in the current assembly. The SELDLA-extended assembly had a total N50 of 34.19 Mb, with the longest scaffold measuring 42.02 Mb. The total assembly size was 795.65 Mb, corresponding to 99.3% of the estimated genome size (~ 800 Mb) [33]. The chromosome-level scaffolds cover 791.84 Mb (99.55%) of the total sequences in the new assembly. This assembly yielded a 3-fold increase in average scaffold length and a 2.6-fold increase in the N50 scaffold compared to the "TUNAREF2023."

To assess the distribution of scaffolds within the pseudochromosomes, each sub-scaffold inherited from the TUNAREF2023 assembly was categorized as either "Located only" or "Located and Oriented" according to our linkage information. In cases where the scaffolds were subdivided into fragments and assigned to different pseudochromosomes, their contributions were normalized

against the length of the parent scaffold. The placement of scaffolds from the TUNAREF2023 assembly across the 24 pseudochromosomes is summarized (Table 1; Supplementary File S1). For each chromosome, the number of scaffolds assigned, the total sequence length of scaffolds that were both located and oriented versus those that could only be localized, and the proportion of sequences in each category are reported. This provides a genome-wide overview of the anchoring process, highlighting that most of the sequences were not only placed but also confidently oriented by linkage information.

Repeat and gene set annotations

Repetitive sequences in the *T. orientalis* genome were annotated using a combination of ab initio– and homology-based approaches. In total, repetitive elements (REs) accounted for 30.41% (~242 Mb) of the SELDLA-extended *T. orientalis* assembly (Table 2; Supplementary Material 4 - Table S4). The most abundant categories were long interspersed nuclear elements (LINEs; 25.8 Mb, 3.24%) and DNA transposons (17.0 Mb, 2.14%). Repetitive elements were unevenly distributed along the chromosomes, and the largest fraction (21.9%) remained classified as "unclassified," indicating the presence of lineage-specific or highly diverged repeats requiring further characterization. A clear negative correlation was observed between repeat density and gene density across non-overlapping 100-kb windows in chromosome-scale scaffolds (Fig. 1, tracks d and e).

To further characterize the unplaced scaffolds, we examined their repeat content using RepeatMasker annotations. Repetitive elements accounted for approximately 30.0% of the total length of the chromosome-scale scaffolds (237.2 Mb of 791.8 Mb), whereas unplaced scaffolds were strongly enriched in repetitive sequences, with repeats comprising approximately 87.4% of their total length (3.32 Mb of 3.80 Mb). These results indicate that the unplaced scaffolds largely consist of highly repetitive and likely non-coding regions, which are intrinsically difficult to anchor and orient at the chromosome scale. This enrichment of repetitive elements in unplaced scaffolds provides a plausible explanation for the high completeness of the predicted proteome despite the presence of residual unanchored sequences.

A total of 22,640 putative protein-coding genes were predicted (producing 30,446 transcript isoforms due to alternative splicing), covering approximately 35% of the *T. orientalis* genome (Table 2). Functional annotation

Table 1 Summary of scaffold placement and orientation using linkage analysis

Chromosome	#TUNAREF 2023 Scaffolds	Mb Oriented	Mb Located-only	Total Mb	% Oriented	% Located-only
LG01	22	38.32	0.76	39.08	98.1	1.9
LG02	11	20.9	0.52	21.42	97.6	2.4
LG03	25	39.61	1.78	41.39	95.7	4.3
LG04	22	35.1	0.79	35.89	97.8	2.2
LG05	14	38.24	0.31	38.55	99.2	0.8
LG06	15	34.59	1.11	35.71	96.9	3.1
LG07	28	34.19	1.79	35.98	95	5
LG08	9	29.38	1.22	30.6	96	4
LG09	45	33.81	2.9	36.71	92.1	7.9
LG10	14	34.94	0.46	35.4	98.7	1.3
LG11	10	30.16	0.54	30.7	98.2	1.8
LG12	12	31.94	0.5	32.44	98.5	1.5
LG13	17	30.71	5.62	36.33	84.5	15.5
LG14	17	32.05	0.51	32.56	98.4	1.6
LG15	5	31.59	0.17	31.76	99.5	0.5
LG16	27	33.62	1.1	34.71	96.8	3.2
LG17	18	31.45	1.39	32.84	95.8	4.2
LG18	30	25.26	2.07	27.33	92.4	7.6
LG19	18	27.89	0.88	28.77	97	3
LG20	16	27.99	0.19	28.18	99.3	0.7
LG21	8	33.82	0.24	34.06	99.3	0.7
LG22	4	30.37	0.13	30.51	99.6	0.4
LG23	17	26.01	0.67	26.68	97.5	2.5
LG24	17	29.68	0.6	30.27	98	2
Total	421	761.62	26.25	787.87	96.7	3.3

For ease of presentation, the chromosome names are shown as LG01, LG02, …. These names correspond to the official chromosome identifiers in the NCBI-submitted assembly (e.g., C0000000001_linkage_scaffold_1, C0000000002_linkage_scaffold_2) as well as those listed in Supplementary File S1. Genome-wide totals are provided in the last row. Detailed per-scaffold information is available in Supplementary File S1

Table 2 Genome assembly statistics and annotation of the *Thunnus orientalis* genome

Genome assembly	**Value**
Number of sequences	192
Total length (bp)	795,647,880
N50 (bp)	34,195,991
Max scaffold length (bp)	42,023,264
GC content (%)	39.67
Number of gaps	665
BUSCO (%)	C = 98.3; S = 97.4; D = 0.9; F = 0.8; M = 0.9
Repeats annotation	**Value**
SINEs (bp)	1,533,091 (0.19%)
LINEs (bp)	25,797,628 (3.24%)
LTR elements (bp)	3,249,944 (0.41%)
DNA transposons (bp)	17,004,590 (2.41%)
Simple repeats (bp)	17,103,504 (2.15%)
Low complexity (bp)	2,638,066 (0.33%)
Small RNA (bp)	1,411,561 (0.18%)
Unclassified (bp)	174,264,846 (21.9%)
Total (bp)	241,985,022 (30.41%)
Gene annotation	
Number of predicted genes	22,640
Average gene length (bp)	12,400
Average CDS length (bp)	1,807
Average exon per transcript	11
Percent of the genome covered by genes	35
BUSCO	C = 88.5; S = 69; D = 19.5; F = 2.2; M = 9.3
Functional annotation total	**22,410 (98.98%)**
Swissport	19,942 (88.07%)
TrEMBL	22,144 (97.81%)
NCBI nr	21,207 (93.67%)
InterPro	20,374 (89.99%)

BUSCO benchmarking universal single-copy orthologs, *C* complete, *S* complete and single copy, *D* complete and duplicated, *F* fragmented, *M* missing, *LINEs* long interspersed nuclear elements, *LTR* long terminal repeat, *SINEs* short interspersed nuclear elements, *CDS* coding DNA sequence

assigned at least one database hit to 22,410 genes (98.98%), based on searches against multiple protein databases (Supplementary Material 4 - Table S5). Specifically, 13,587 genes were annotated with KEGG Orthology (KO) terms, corresponding to 8,509 unique KO entries (Supplementary File S2), and 18,137 genes were associated with at least one Gene Ontology (GO) term, representing 6,901 unique GO terms (Supplementary File S3).

The number of predicted protein-coding genes in *T. orientalis* was comparable to those reported for closely related scombrid species, including yellowfin tuna (*T. albacares*; 24,623 genes), eastern little tuna (*E. affinis*; 23,059 genes), and southern bluefin tuna (*T. maccoyii*; 24,659 genes) (Supplementary Material 4 - Table S2). Although the predicted gene number was lower than that of a previously reported *T. orientalis* annotation (26,433 genes) [34], the current assembly represents a substantial improvement in quality (N50 increased from 8.2 kb to 34.19 Mb).

Data validation

The quality and completeness of the SELDLA-extended *T. orientalis* genome assembly were evaluated using multiple independent approaches. First, 230 of 233 (98.7%) CVGs [26] were identified, indicating high completeness at the assembly level. Consistently, BUSCO analysis of the genome assembly recovered 7,085 complete BUSCOs (98.3%), including 7,020 single-copy (97.4%) and 65 duplicated (0.9%) BUSCOs (Table 2), supporting the overall contiguity and completeness of the assembled genome.

Comparative synteny analysis further demonstrated structural consistency between *T. orientalis* and *Oryzias latipes*, with 6,727 of 6,773 single-copy orthologs localized to corresponding chromosomes in both species (Fig. 2). Accordingly, 24 scaffolds longer than 10 Mb in the *T. orientalis* assembly were concordantly numbered with *O. latipes* chromosomes (Figure S1).

The completeness of the predicted gene set was separately assessed using BUSCO, which identified 6,378 complete BUSCOs (88.5%) out of 7,207 actinopterygian single-copy orthologs (Table 2). This analysis revealed a notable proportion of duplicated (19.5%) and fragmented (2.2%) BUSCOs, together with 9.3% missing BUSCOs, indicating incomplete recovery of conserved coding genes at the gene model level despite the high completeness of the underlying genome assembly.

Orthogroup clustering across nine teleost species grouped 218,379 genes into 21,668 orthogroups, of which 11,437 orthogroups were conserved across all examined species (Fig. 3). Among the 22,640 predicted genes in *T. orientalis*, 21,428 genes (94.6%) were assigned to 17,807 orthogroups, and 323 orthogroups were unique to scombrid species (*T. albacares*, *T. maccoyii*, and *T. orientalis*). In contrast, 761 orthogroups conserved in the other teleost species were not detected in *T. orientalis*, likely reflecting a combination of incomplete gene recovery and potential lineage-specific gene loss.

Although 98.98% of the predicted genes were assigned at least one functional annotation, the BUSCO and orthogroup analyses indicate that the recovered gene set does not fully represent the expected conserved actinopterygian coding repertoire. Therefore, the high annotation hit rate reflects comprehensive annotation of the recovered gene models, rather than complete coverage of all conserved genes. This limitation should be taken into consideration when interpreting functional and comparative genomic analyses based on the current gene annotation.

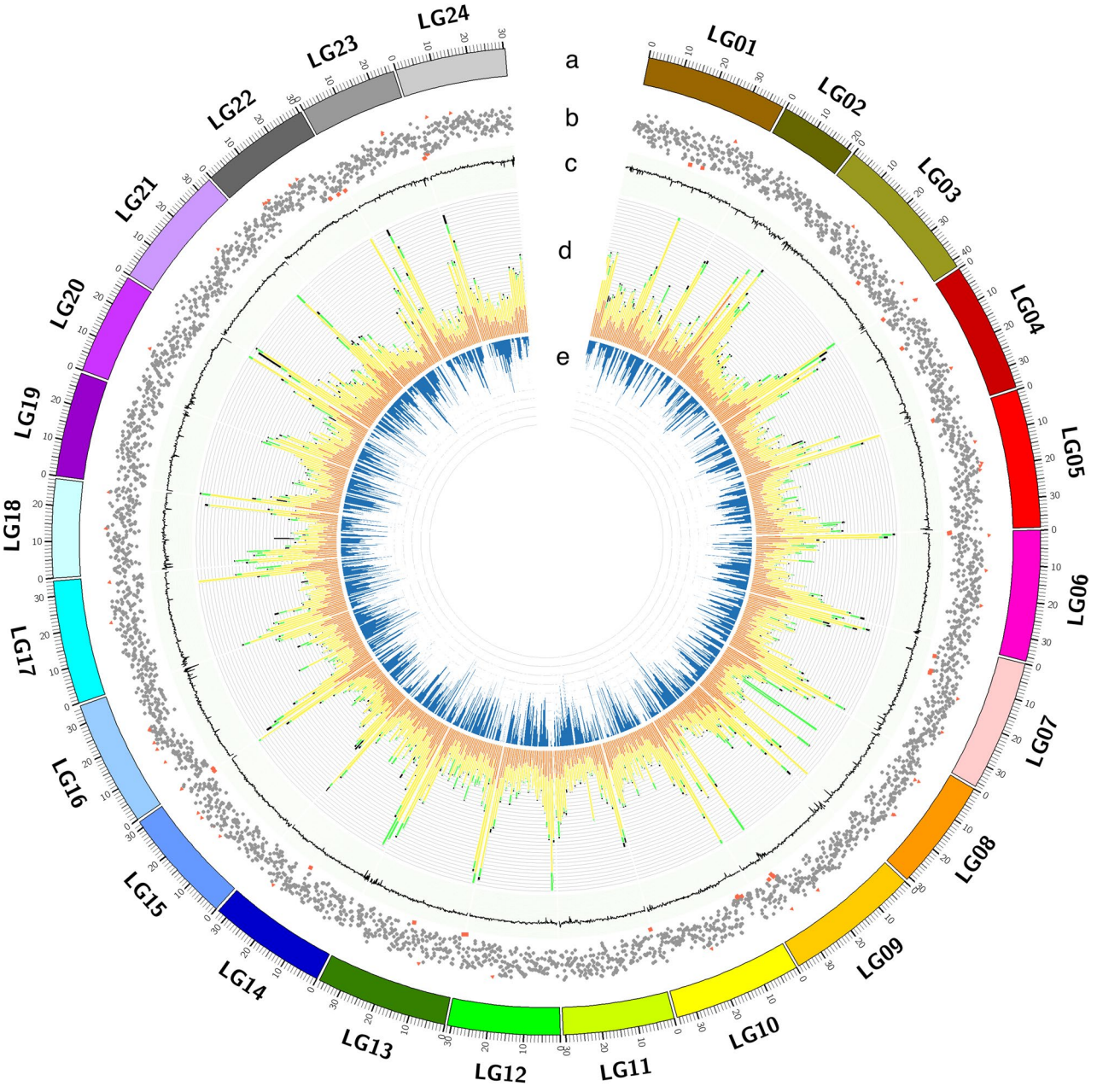

Fig. 1 Circos plot depicting genome features across the 24 Thunnus orientalis chromosomes. **a** T. orientalis chromosomes (LG01–LG24 on an Mbscale). **b** Short-read coverage plot. Coverage within 2 SD of the mean is shown as grey circles. Abnormal sequence coverage (± 2 SD from the mean) isindicated with a red square or triangle, respectively. **c** GC content (percentage). **d** Distribution of repeat elements: DNA transposons (light orange bar),LINEs (yellow bar), LTRs (green bar), and SINEs (black bar). **e** Gene density. A window size of 1 Mb was used for tracks (**d**) and (**e**), 0.1 Mb for track (**c**), and 0.2 Mb for track (**b**)

Conclusions

The new chromosome-level *T. orientalis* genome assembly presented herein has a scaffold N50 of 34.19 Mb, 98.3% BUSCO completeness, and 22,640 annotated genes, representing a step forward compared to the previous assembly. Of the 22,640 protein-coding genes, 4,677 (20.65%) contained putative, alternative, spliced transcripts. The availability of chromosome-level genome assemblies is expected to support applied and fundamental research, particularly the genetic improvement of this commercially valuable fish species.

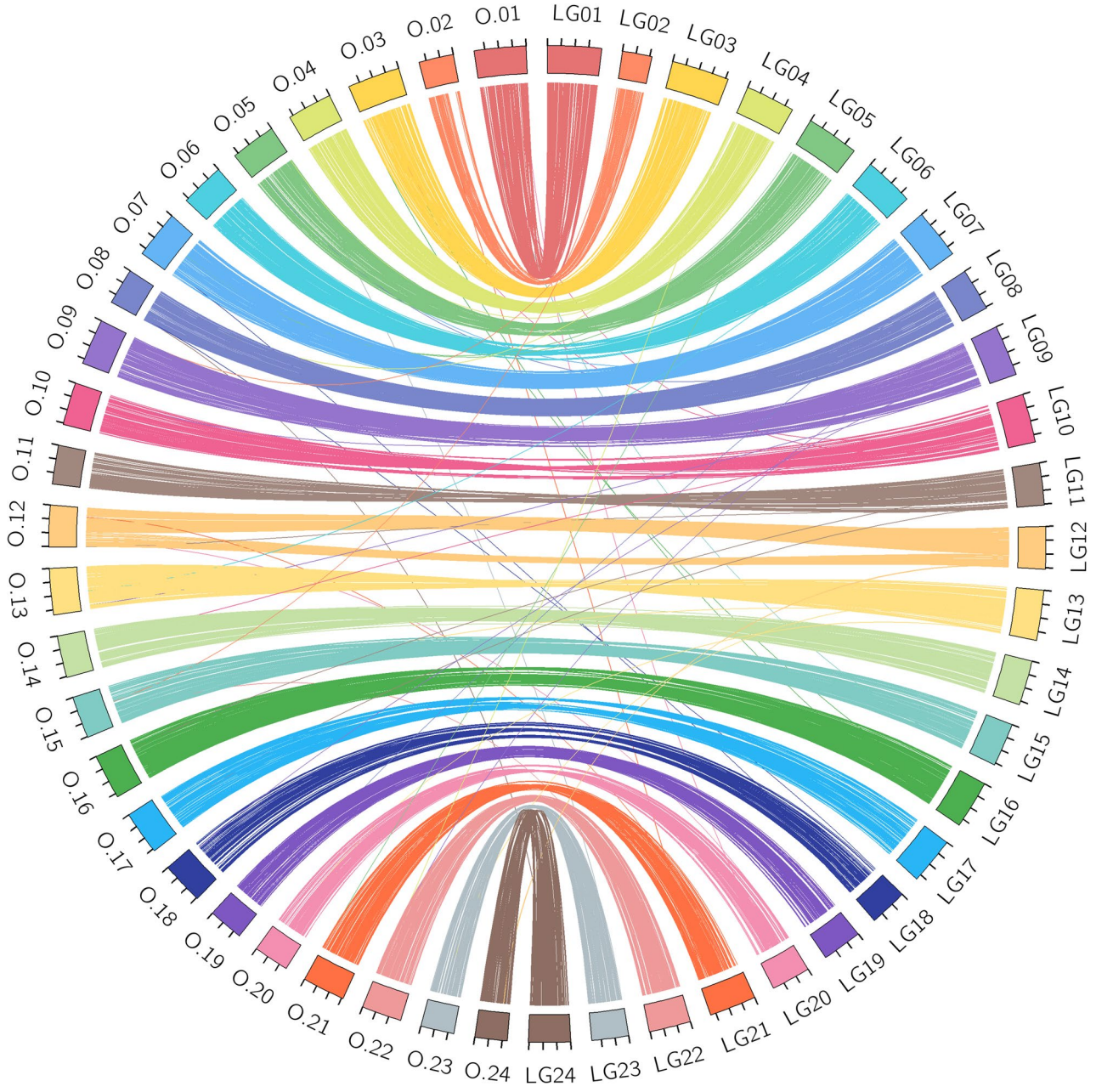

Fig. 2 Chromosome-level synteny between Thunnus orientalis (right, LG01–LG24) and Japanese medaka, Oryzias latipes (left, O.01–O.24) based on 6,773complete single-copy orthologs

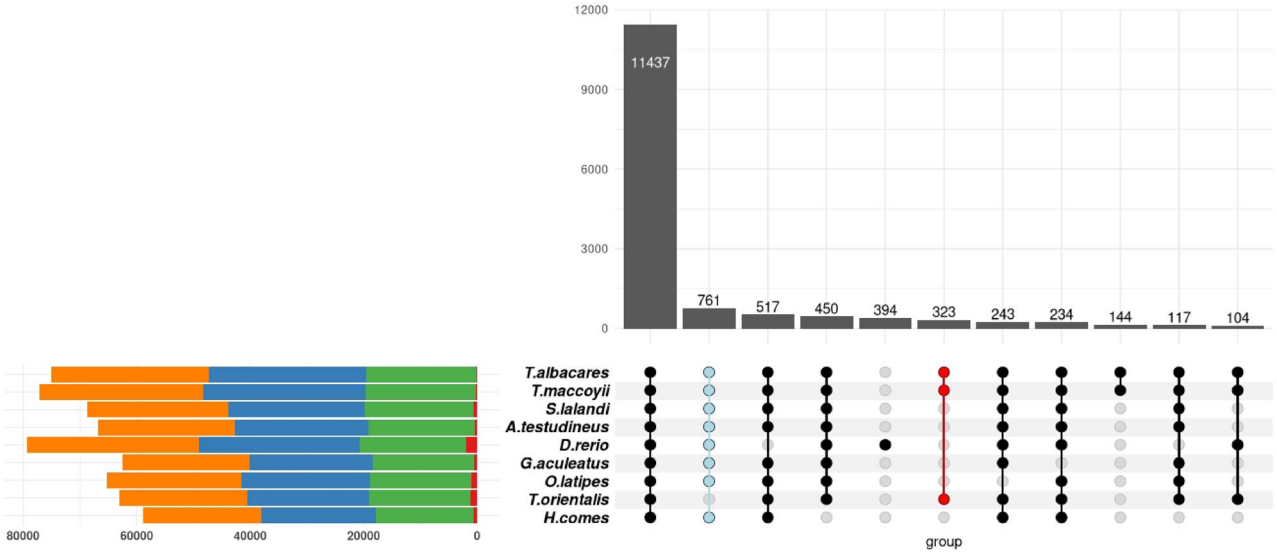

Fig. 3 Orthogroup size and proportion of genes assigned to orthogroups per species. Left: Bar plot showing orthogroup size metrics for each species. Yellow indicates the total number of genes; blue represents the number of genes assigned to orthogroups; green denotes the number of orthogroup-scontaining each species; red shows the number of unassigned genes. Right: UpSet plot showing the intersections of orthogroups among eight fish species.Bars represent the number of orthogroups shared by the species marked by connected dots below the x-axis. A total of 11,437 orthogroups were-conserved across all species. Orthogroups shared exclusively by the three tuna species Thunnus albacares, T. maccoyii, and T. orientalis are highlightedin red (n = 323). Notably, T. orientalis lacks 761 orthogroups that are present in all other species, highlighted in light blue. Intersections comprising fewerthan 100 orthogroups are omitted for visual clarity

Abbreviations

NCBI	National Center for Biotechnology Information
BUSCO	Benchmarking universal single-copy orthologs
LINEs	Long interspersed nuclear elements
LTR	Long terminal repeat
SINEs	Short interspersed nuclear elements
CDS	Coding DNA sequence

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12856-y.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements
We thank Director Tokihiko Okada and the staff of the Aquaculture Technology and Production Center of Kindai University for providing tuna testes.

Authors' contributions
SA and TI designed this study. DS conducted the experiments. The Ehime members, TS, ON, TM, RG, MH (**E. affinis**), and TI (**T. orientalis**) provided resources and contributed to the generation of interspecific hybrid fish. KY developed SELDLA software. XF and KY analyzed the data. XF and SA drafted the manuscript. All the authors have read and approved the final version of the manuscript.

Funding
This study was supported in part by the Japan Society for the Promotion of Science KAKENHI (Grant nos. 20H00429, 23H00342).

Data availability
The genome assembly, gene annotations, and supporting files generated in this study are available from Zenodo (DOI: https://doi.org/10.5281/zenodo.18751919). The chromosome-level genome assembly has also been deposited in GenBank under accession number GCA_055472215.1. The short-read sequencing data derived from interspecific hybrid individuals used for genome assembly have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA1277492.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 17 October 2025 / Accepted: 9 April 2026
Published online: 20 April 2026

References

1. Benetti D, Partridge G, Buentello A. Advances in Tuna Aquaculture: From Hatchery to Market. Academic; 2015.
2. Higuchi K, Takashi T, Okita K, Hayashida T, Nyuji M, Gen K. Effect of a long photoperiod on the timing of spawning in the Pacific bluefin tuna, Thunnus orientalis. Aquac Rep. 2024;36:102062.
3. Ahmad A, Sheikh Abdullah SR, Hasan HA, Othman AR. Ismail N 'izzati. Aquaculture industry: Supply and demand, best practices, effluent and its current issues and treatment technology. J Environ Manage. 2021;287:112271.
4. Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, et al. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020;21:389–409.

5. Benestan L. Population Genomics Applied to Fishery Management and Conservation. Population Genomics: Marine Organisms. 2019:399–421.
6. Hayashida T, Soma S, Nakamura Y, Higuchi K, Kazeto Y, Gen K. Transcriptome characterization of gonadal sex differentiation in Pacific bluefin tuna, Thunnus orientalis (Temminck et Schlegel). Sci Rep. 2023;13:13867.
7. Jiao W, Fu X, Dou J, Li H, Su H, Mao J, et al. High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. DNA Res. 2014;21:85–101.
8. Zhang X, Mizukoshi M, Zhang H, Tan E, Igarashi Y, Suzuki Y, et al. Ultrahigh-density linkage map construction using low-coverage whole-genome sequencing of a doubled haploid population: case study of torafugu (takifugu rubripes). Genes (Basel). 2018;9(3):120.
9. Yoshitake K, Igarashi Y, Mizukoshi M, Kinoshita S, Mitsuyama S, Suzuki Y, et al. Artificially designed hybrids facilitate efficient generation of high-resolution linkage maps. Sci Rep. 2018;8:16104.
10. Yoshitake K, Ishikawa A, Yonezawa R, Kinoshita S, Kitano J, Asakawa S. Construction of a chromosome-level Japanese stickleback species genome using ultra-dense linkage analysis with single-cell sperm sequencing. NAR Genom Bioinform. 2022;4:lqac026.
11. Havelka M, Sawayama E, Saito T, Yoshitake K, Saka D, Ineno T, et al. Chromosome-Scale Genome Assembly and Transcriptome Assembly of Kawakawa; A Tuna-Like Species. Front Genet. 2021;12:739781.
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
13. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
14. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
15. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2016;44:D81–9.
16. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA. 2021;12:2.
17. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 2024;34:769–77.
18. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. The Sequence Read Archive: a decade more of explosive growth. Nucleic Acids Res. 2022;50:D387–90.
19. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
20. Dyer SC, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Barrera-Enriquez VP, et al. Ensembl 2025. Nucleic Acids Res. 2024;53:D948–57.
21. Goldfarb T, Kodali VK, Pujar S, Brover V, Robbertse B, Farrell CM, et al. NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. Nucleic Acids Res. 2025;53:D243–57.
22. Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. BMC Bioinformatics. 2021;22:566.
23. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
24. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2.
25. Klopfenstein D, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 2018;8.
26. Hara Y, Tatsumi K, Yoshida M, Kajikawa E, Kiyonari H, Kuraku S. Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation. BMC Genomics. 2015;16:977. https://doi.org/10.1186/s12864-015-2007-1.
27. Nishimura O, Hara Y, Kuraku S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics. 2017;33:3635–7.
28. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 2021;38:4647–54.
29. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.
30. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
31. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph. 2014;20:1983–92.
32. Lee Y-H, Yen T-B, Chen C-F, Tseng M-C. Variation in the Karyotype, Cytochrome Gene, and 5S rDNA of Four (Perciformes, Scombridae) Tunas. Zool Stud. 2018;57:e34.
33. Ida H, Oka N, Terashima H, Hayashizaki K-I. Karyotypes and cellular DNA contents of three species of the family Scombridae from japan. Nippon Suisan Gakkai Shi. 1993;59:1319–23.
34. Nakamura Y, Mori K, Saitoh K, Oshima K, Mekuchi M, Sugaya T, et al. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. Proc Natl Acad Sci U S A. 2013;110:11061–6.

Publisher's Note