



OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of *Nemipterus japonicus* based on PacBio sequencing and Hi-C technology

Feng-Ling Lei, Ren-Xie Wu , Fan-Li Long, Fang-Yuan Han & Zhen-Bang Liang

Nemipterus japonicus is an economically important demersal fish species inhabiting warm coastal waters of the Indo-West Pacific region, yet its genomic resources remain scarce. Here, we successfully assembled a high-quality chromosome-level reference genome of *N. japonicus* by integrating Illumina, PacBio, and Hi-C sequencing technologies. The genome assembly spans 723.27 Mb, with a contig N50 of 30.83 Mb, and 97.04% of the contigs were anchored onto 22 pseudochromosomes, yielding a scaffold N50 of 30.96 Mb. A total of 23,514 protein-coding genes and 6,532 non-coding RNAs were predicted, with repetitive sequences accounting for 21.84% of the genome. Comparative genomic analysis revealed that *N. japonicus* has 218 expanded and 1,076 contracted gene families, with the divergence times of 56.9 to 85.9 million years ago from three Sparidae species. Positive selection analysis further identified 163 positively selected genes potentially involved in growth and development, reproduction, immune response, and DNA repair of *N. japonicus*. This study provides fundamental data for investigating the genetic evolution of *N. japonicus* and the genomes of other Nemipterid fish.

Background & Summary

Japanese threadfin bream (*Nemipterus japonicus*) is a representative species of the family Nemipteridae (Perciformes), a commercially important demersal fish widely distributed in the warm coastal waters of the Indo-West Pacific region¹. In the southeastern coastal waters of China, *N. japonicus* is among the key target species for bottom trawling and longline fisheries, with particularly high abundance in the northern South China Sea (NSCS)². Its annual catch (approximately 50,000 tons) accounts for about 15–20% of the total yield of *Nemipterus* species in the NSCS. *N. japonicus* has a high nutritional value, a delicate flavor, and contains substantial proteins, omega-3 fatty acids, and various vitamins and minerals in its muscle tissue³. Ecologically, it occupies a mid-to-high trophic level in marine food webs, feeding on small invertebrates and fishes^{4,5}, while also serving as a crucial prey for higher-level carnivorous fish⁶. These attributes underscore its pivotal role in fisheries resources and ecosystems across the Indo-West Pacific region.

Current researches on *N. japonicus* encompass multiple disciplines, including morphological characteristics^{1,7}, growth traits^{8,9}, feeding ecology^{10,11}, reproductive biology^{12,13}, population dynamics^{14,15}, and phylogenetics^{16–19}. Previous studies indicated that *N. japonicus* exhibits typical r-selected life history traits characterized by rapid growth, high fecundity, prolonged spawning periods with extensive spawning ground distribution, and variable feeding behaviors, highlighting its strong ecological plasticity. Molecular phylogenetic analyses revealed two distinct geographic lineages of *N. japonicus*: Indian Ocean and Northwest Pacific populations, with clear phenotypic differentiation in abdominal coloration (yellow vs. silver-white), suggesting potential genetic differentiation associated with cryptic species^{20,21}. Along the Indian coast, the eastern population shows pronounced sexual growth dimorphism, a pattern less marked in the western population⁸. The northern population displayed allometric growth patterns, and the southern population followed an isometric growth pattern²². Additionally, the primary spawning season for *N. japonicus* in the Bay of Bengal (November to March) significantly differs from other Indo-West Pacific populations (May to October)²¹. The diversity and endemism observed in reproductive biology, ecological adaptability, and population genetic structure of *N. japonicus* suggest that it may possess unique and complex genomic features. However, only the mitochondrial genome data

College of Fisheries, Guangdong Ocean University, Zhanjiang, 524088, China. ✉e-mail: wurenxie@163.com

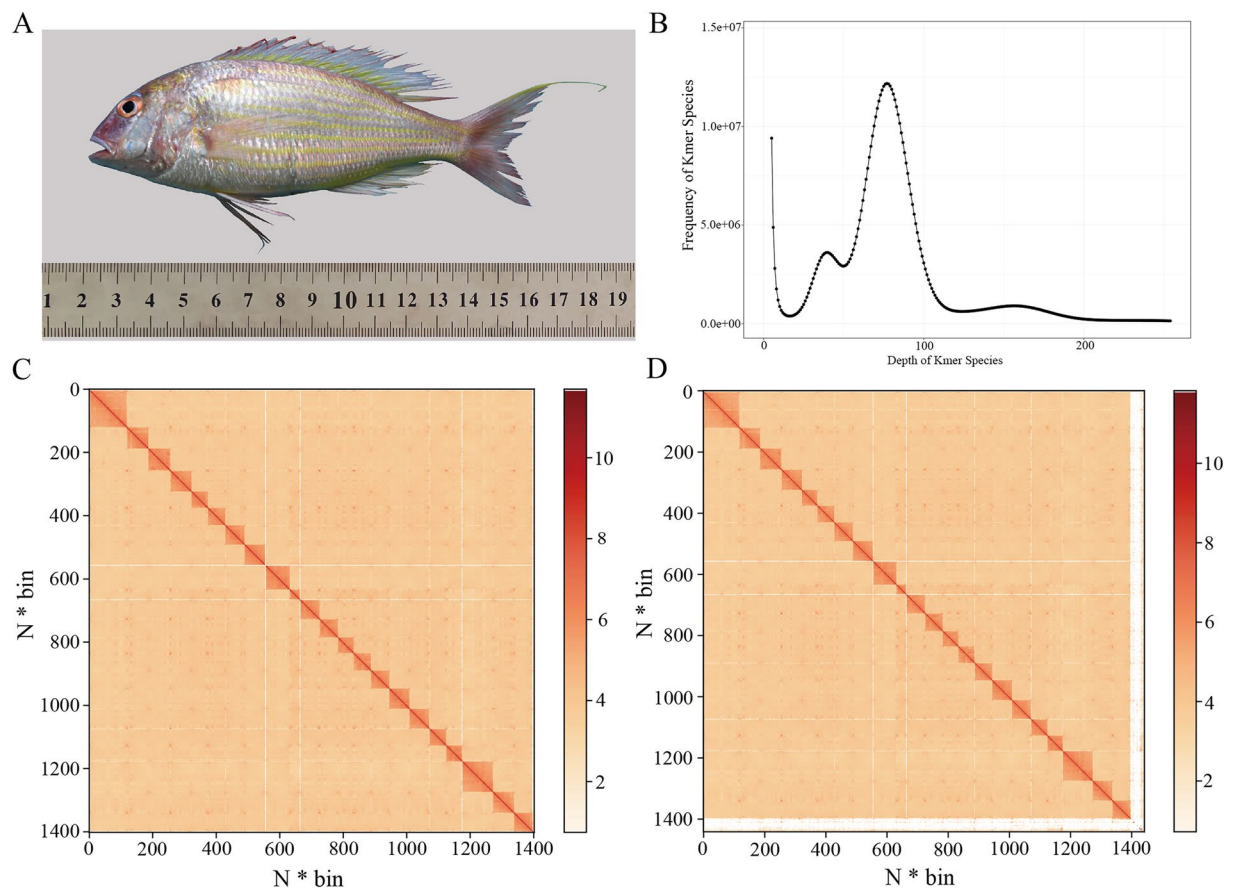


Fig. 1 Specimen, genome size estimation, and chromosome-level assembly for *N. japonicus* genome sequencing. **(A)** Specimen used for genome sequencing. **(B)** Genome-size estimation based on 17-mer distribution. **(C)~(D)** Chromosome **(C)** and genome-wide **(D)** heat maps at 500 kb resolution of Hi-C assembly of *N. japonicus*. The 500 kb corresponds to 1 bin and the color bar indicates the logarithm of the strength of the contact density.

of this species have been reported so far²³, with no published information on the whole genome. Thus, in-depth research on its adaptive evolutionary mechanisms, population genetic structure, evolutionary history, and germplasm resource conservation remains severely limited.

This study integrated Illumina Nova-Seq, PacBio SMRT sequencing, and Hi-C chromosome conformation capture technology to conduct deep sequencing and assembly of the *N. japonicus* genome, aiming to obtain high-quality chromosome-level genomic data. First, the RNA-seq data from Illumina and PacBio platforms were used to systematically annotate repetitive sequences, the structure and function of protein-coding genes, and non-coding RNAs (ncRNAs) in the genome by combining homology, *de novo*, and RNA-seq predictions. Second, comparative genomic analyses were performed between *N. japonicus* and 18 other fish species, including gene family clustering, expansion and contraction analysis, phylogenetic tree construction, and divergence time estimation, with synteny analysis using coding and whole-genome sequences of the closely related species, *Sparus aurata*. Finally, eight positive selection groups were established based on the morphological traits of *N. japonicus* for enrichment analysis and gene functional annotation to identify potential trait-associated genes. Thus, this study became the first whole-genome sequencing and assembly of a species in the family Nemipteridae, providing a genomic foundation for evolutionary biology, germplasm resource conservation, and sex-determination mechanism analyses of *N. japonicus*. Moreover, it serves as a reference genome for genetic investigation of other Nemipterid fish.

Methods

Sample collection and nucleic acid extraction. A live *N. japonicus* (specimen ID: 20210923002) (Fig. 1A) captured in the coastal waters of Maoming City, Guangdong Province, China, in September 2021, was used for genomic sequencing. The specimen measured 14.5 cm in body length and weighed 80.4 g. After anesthesia with MS-222 solution (200 mg/L) (ethyl 3-aminobenzoate methanesulfonate, Sigma-Aldrich, Shanghai, China), the muscle, liver, heart, and brain tissues were quickly excised and separately stored in 1.5 mL sterile tubes. The tissues were flash-frozen in liquid nitrogen and transferred to a -80°C refrigerator in the laboratory. Genomic DNA was extracted from the muscle tissue using the phenol/chloroform method²⁴, while RNA was extracted from the above four tissues using Trizol reagent (Invitrogen, Carlsbad, CA, USA).

Type	Platform	Library (bp)	Raw data (Gb)	Clean data (Gb)	Coverage (×)
Illumina Nova	Illumina NovaSeq-6000	350	73.39	63.60	106.36×
PacBio SMRT	PacBio Sequel II	15k	35.63	—	49.26×
Hi-C	Illumina HiSeq 2500	350	75.80	72.96	108.20×
Illumina RNA-seq	Illumina NovaSeq-6000	350	7.46	7.00	10.31×
ONT RNA-seq	NanopromethION	—	14.03	13.40	19.40×

Table 1. Sequencing data statistics of the *N. japonicus* genome assembly and annotation.

Level	Contig/Scaffold Length(bp)	Contig/Scaffold Number
N90	23,744,612	21
N80	26,166,573	18
N70	27,417,517	15
N60	29,042,877	13
N50	30,830,163	10
Total length	723,269,530	—
Number (≥ 100 bp)	—	257
Number (≥ 2 kb)	—	257
Max length	53,928,558	—

Table 2. Sequence statistics of the *N. japonicus* genome assembly.

Library construction and genome sequencing. A paired-end sequencing library with a 350 bp insert size was constructed following the standard Illumina experimental protocol²⁵ and sequenced on the Illumina NovaSeq-6000 platform (Illumina, San Diego, CA, USA), generating 489,256,470 raw reads, 73.39 Gb of data (Table 1). The raw data were deduplicated and filtered using FASTQ (v0.23.2)²⁶, resulting in 428,627,422 clean reads (63.60 Gb) with a 40.8% GC content (Table 1). Concurrently, a 15 kb-insert SMRTbell library was constructed using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on the PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA, USA) in CCS (Circular Consensus Sequencing) mode. The raw sequencing data were adapter-trimmed and filtered to remove adapter sequences. After multiple sequence alignment and consensus correction using SMRTlink²⁷, 2,449,579 high-precision HiFi reads generated 35.63 Gb of data (Table 1). The HiFi reads had an average length of 14,544 bp and an N50 length of 14,749 bp, with a 41.32% GC content.

Genome size estimation and assembly. A K-mer analysis (K = 17) was performed using Jellyfish (v1.11.1)²⁸ on Illumina sequencing-derived clean data, revealing a 17-mer depth of 77 (Fig. 1B). The estimated *N. japonicus* genome size was 700 Mb, and later adjusted to 690 Mb, and the genome heterozygosity rate was 0.56%. Subsequently, HiFi reads were rapidly assembled using Hifiasm (v0.16.1)²⁹. The initially assembled genome was de-redundant with Purge Haplotigs (v1.0.4)³⁰, yielding 257 contigs with maximum, N90, and N50 lengths of 53,928,558 bp, 23,744,612 bp, and 30,830,163 bp, respectively (Table 2). The *N. japonicus* genome size (based on contig-level assembly) was 723.27 Mb with GC content of 41.46%.

Hi-C library preparation, sequencing, and chromosome assembly. First, the sticky ends generated from muscle tissue DNA after cross-linking with paraformaldehyde solution (Sbjbio, Nanjing, China) and restriction enzyme Mbo I digestion were end-repaired using biotin labeling. Subsequently, adjacent DNA fragments were ligated using DNA ligase, followed by protease treatment to reverse cross-linking. The purified DNA was then randomly fragmented into 300–500 bp fragments. Finally, the biotin-labeled DNA fragments were captured using streptavidin magnetic beads, followed by end repair, A-tailing, adapter ligation, PCR amplification, and purification to complete the Hi-C library construction.

The qualified library was sequenced using paired-end sequencing on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA), yielding 505,351,562 raw reads (75.80 Gb). After quality control, 496,771,290 clean reads (72.96 Gb) were obtained (Table 1), with a 41.84% GC content. Further quality control and filtering using HiCUP (v0.8.0)³¹, 186,013,178 generated valid read pairs (accounting for 97.18% of all read pairs), representing 74.89% of total sequencing data. Next, deduplication resulted in 130,388,169 valid pairs (68.12%), accounting for 52.49% of the total sequencing data (Table 3).

The contigs were partitioned, anchored, ordered, oriented, and merged with 3D-DNA software³² to achieve a chromosome-level genome assembly. The interaction map of the assembled genome was constructed using Juicebox (v1.6)³³, then corrected and visualized in JuiceBox (v1.11.08)³⁴ to rectify chromosomal translocations and inversions. The resulting interaction map conformed to the principles of chromosomal interactions (Fig. 1C,D), confirming the effectiveness of Hi-C-assisted assembly. Ultimately, 97.04% of the sequences were successfully anchored onto 22 pseudochromosomes, with a total chromosome length of 700.57 Mb (Table 4). The contig N50 and scaffold N50 of the *N. japonicus* genome assembly were 30.83 Mb and 30.96 Mb, respectively (Table 4).

Di-Tag Type	Di-Tag Count	Percent in Paired(%)	Percent in Total Reads(%)
Same Circularised	317,783	0.17	0.13
Same Fragment Dangling Ends	787,768	0.41	0.32
Same Fragment Internal	2,357,537	1.23	0.95
Re-ligation	1,933,771	1.01	0.78
Contiguous Sequence	0	0.00	0.00
Wrong Size	0	0.00	0.00
Invalid Pairs	5,396,859	2.82	2.17
Valid Pairs	186,013,178	97.18	74.89
Valid Pairs (de-duplication)	130,388,169	68.12	52.49

Table 3. Statistics of the *N. japonicus* HiCUP filtering.

Type	Sequence length (bp)	Sequence number	Contig N50 (bp)	Scaffold N50 (bp)
Raw	723,269,530	257	30,830,163	30,830,163
Hi-C	721,909,522	152	30,830,163	30,963,390
Hi-C chr	700,571,217	22	30,830,163	30,963,390
Hi-C nochr	21,338,305	130	304,857	465,076

Table 4. Statistics of the *N. japonicus* Hi-C-assisted assembly.

RNA library construction and transcriptome sequencing. First, mRNA purified from the mixed RNA samples was reverse transcribed using the TURScript First Strand cDNA Synthesis Kit (AidLab, Beijing, China) to synthesize double-stranded cDNA libraries. An Illumina paired-end sequencing library with ~350 bp insert size was constructed. Subsequently, the library was sequenced on the Illumina NovaSeq-6000 platform (Illumina, San Diego, CA, USA), yielding 7.46 Gb of raw data and 7.00 Gb of clean data after filtering (Table 1). Furthermore, a Nanopore full-length transcriptome library was prepared using the PCR Barcoding (SQK-PBK004) and PCR-cDNA Sequencing Kits (SQK-PCS109, Oxford Nanopore Technologies, Oxford, UK). Sequencing was performed on the PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK), generating 14.03 Gb of raw data and 13.40 Gb of clean data after filtering (Table 1).

Genome annotation. Gene annotation in *N. japonicus* was performed by integrating homology-based prediction, *de novo* prediction, and RNA-seq/EST-based prediction methods. First, the whole-genome sequences of 11 reference species, including *Homo sapiens*, *Danio rerio*, *Oryzias latipes*, *Lates calcarifer*, *Larimichthys crocea*, *Nibea albiflora*, *Pagrus major*, *Acanthopagrus latus*, *Oplegnathus fasciatus*, *Gasterosteus aculeatus*, and *Takifugu rubripes*, were downloaded from the NCBI database. The genome sequence of *N. japonicus* was aligned with the protein-coding sequences of these 11 species using tblastn (E-value $\leq 1 \times 10^{-5}$)³⁵ and Genewise (v2.4.1)³⁶ to predict the gene structure of *N. japonicus*. Second, Augustus (v3.3.2)³⁷ was utilized to calculate the codon usage frequencies, exon-intron distributions, and other factors. Third, RNA-seq data were assembled to obtain transcripts using Tophat (v2.1.1) and Cufflinks (v2.2.1)³⁸. Finally, the three gene sets were integrated into a non-redundant, more complete gene set using Maker2 (v2.31.10)³⁹. After incorporating the CEGMA results and employing the HiCESAP workflow, 23,514 protein-coding genes were predicted. The average lengths of the predicted genes, coding sequences (CDS), exons, and introns were 16,089 bp, 1,732 bp, 274.90 bp, and 1,451 bp, respectively. Their distribution in the *N. japonicus* genome is highly consistent with that of the 11 reference species (Table 5, Fig. 2A), reflecting their evolutionary conservation. *N. japonicus* has 23,514 protein-coding genes, all of which were aligned to the NR (Non-Redundant Protein), SwissProt (Swiss Protein Institute), KEGG (Kyoto Encyclopedia of Genes and Genomes), TrEMBL (Translation of European Molecular Biology Laboratory), and GO (Gene Ontology) databases for gene functional annotation. Moreover, the Pfam (Protein Family) and InterPro (Integrated Resource of Protein) databases were employed to predict conserved functional domains and protein family information. Ultimately, 22,858 genes (97.21%) were successfully annotated, while 656 genes (2.79%) remained unannotated (Table 6).

Repetitive sequence annotation was performed by combining homologous sequence alignment and *de novo* prediction. First, RepeatMasker (v4.1.2)⁴⁰ and RepeatProteinMask (v4.1.2)⁴¹ were run against the RepBase database⁴² to identify sequences homologous to known nucleic acid and amino acid repetitive sequences, respectively. Second, a *de novo* repetitive sequence library was constructed using LTR-Finder (v1.0.7)⁴³, RepeatModeler (v2)⁴⁴, and RepeatScout (v1.0.5)⁴⁴, followed by repetitive sequence prediction with RepeatMasker (v4.1.2)⁴⁰. The integration of both methods predicted 157,659,348 bp of repetitive sequences in *N. japonicus*, accounting for 21.84% of its genome. Further, tandem repetitive sequences predicted by TRF (v4.10.0)⁴⁵ generated 42,505,501 bp (5.89% of the genome). Overall, the proportions of repetitive DNAs, LINEs, SINEs, and LTRs in the *N. japonicus* genome were 10.55%, 3.54%, 0.33%, and 2.33%, respectively (Table 7). Additionally, the distribution of protein-coding genes, repetitive sequences, LTR, LINE, and DNA-TE across the 22 pseudochromosomes of *N. japonicus* is illustrated in Fig. 2B.

Methods	Gene set	Protein coding gene number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	AUGUSTUS	33,606	9885.41	1255.69	6.95	180.64	1450.06
Homolog	<i>Nibeal albiflora</i>	29,648	10,506	1,151	6.20	185.53	1,798
	<i>Danio rerio</i>	34,108	13,533	1,391	7.65	181.67	1,825
	<i>Homo sapiens</i>	26,288	13,518	1,288	7.97	161.49	1,754
	<i>Pagrus major</i>	41,578	11,955	1,327	7.06	188.04	1,755
	<i>Oplegnathus fasciatus</i>	39,067	11,230	1,266	7.22	175.22	1,601
	<i>Gasterosteus aculeatus</i>	33,579	12,773	1,453	8.07	179.95	1,601
	<i>Oryzias latipes</i>	33,257	12,953	1,475	7.97	185.00	1,646
	<i>Lates calcarifer</i>	35,248	13,702	1,453	7.99	181.90	1,753
	<i>Takifugu rubripes</i>	30,883	12,913	1,481	8.33	177.82	1,560
	<i>Larimichthys crocea</i>	36,198	13,053	1,445	7.84	184.37	1,697
	<i>Acanthopagrus latus</i>	37,639	14,102	1,456	7.81	186.40	1,856
BUSCO		3,673	12,098	1,852	12.03	153.92	928.81
MAKER		24,806	15,033	1,697	9.81	263.82	1,412
HiCESAP		23,514	16,089	1,732	10.16	274.90	1,451

Table 5. Structure and parameters of the predicted genes in the *N. japonicus* genome.

The non-coding RNAs (ncRNAs) were also annotated. tRNAs were predicted using tRNAscan-SE (v2.0)⁴⁶. The rRNA sequences of closely related species were used as references to identify rRNAs of *N. japonicus* through BLAST alignment. miRNA and snRNA sequences were predicted using Infernal (v1.1.4)⁴⁷ integrated with Rfam (v1.1.4)⁴⁸. Thus, 6,532 non-coding RNAs were detected in the *N. japonicus* genome, comprising 894 miRNAs (77,062 bp; 0.010675%), 2,240 tRNAs (168,686 bp; 0.023367%), 2,215 rRNAs (269,715 bp; 0.037361%), and 1,183 snRNAs (179,014 bp; 0.024797%) (Table 8).

Gene family clustering, expansion, and contraction analyses. The genomes of 18 fish species were downloaded from NCBI for gene family clustering and phylogenetic analyses, including *Rhincodon typus*, *D. rerio*, *Megalobrama amblycephala*, *Gadus morhua*, *Clupea harengus*, *Protosalanx hyalocranius*, *Tachysurus fulvidraco*, *Perca flavescens*, *Epinephelus akaara*, *A. latus*, *Acanthopagrus schlegelii*, *S. aurata*, *L. crocea*, *Scatophagus argus*, *Amphiprion ocellaris*, *Chelmon rostratus*, *Channa argus*, and *Pundamilia nyererei*. First, protein-encoding genes with < 30 amino acids or containing internal stop codons were removed from the genomes of these 18 species and *N. japonicus*, retaining only the longest transcript in the coding region. Then, all-vs-all blastp (v2.11.0+)⁴⁹ was used to establish similarity relationships between protein sequences of all species (E-value = 1e-5). OrthoMCL (v2.0.9)⁵⁰ (inflation coefficient = 1.5) was subsequently employed to cluster the 23,514 protein-coding genes of *N. japonicus* into 15,525 gene families. Among these, all 19 species shared 5,430 gene families, while 83 were unique to *N. japonicus*. The 19 species had 1,666 single-copy genes (Fig. 3A,B). Additionally, comparative clustering analysis between *N. japonicus* and three Sparidae species (*S. aurata*, *A. latus*, and *A. schlegelii*) revealed 11,699 shared gene families and 475 unique gene families for *N. japonicus* (Fig. 3C).

The *N. japonicus* gene family expansion and contraction analyses using CAFÉ5 (v5.0.0)⁵¹ revealed 218 expanded and 1,076 contracted gene families (Fig. 3E). KEGG and GO enrichment analyses were performed on the 302 genes from 43 significantly expanded gene families and 190 from 42 significantly contracted gene families. The genes from the expanded gene families primarily enriched pathways such as neurodegeneration-multiple diseases, Huntington disease, calcium signaling pathway, and amyotrophic lateral sclerosis. The genes from the contracted gene families mainly enriched the parathyroid hormone synthesis, secretion, and action, NOD-like receptor signaling pathway, cGMP-PKG signaling pathway, and adrenergic signaling in cardiomyocytes. Synteny analyses between *N. japonicus* and *S. aurata* were performed using Mummer (v4.0.0rc1)⁵² (based on whole-genome sequence) and JCVI (v1.1.22)⁵³ (based on coding sequence); the results demonstrated strong chromosomal collinearity between the two species (Fig. 4A,B).

Phylogenetic analysis. Protein sequences from each single-copy gene family across the 19 species were aligned using MAFFT (v7.487)⁵⁴, and the maximum-likelihood phylogeny was inferred using RAXML (v8.2.12)⁵⁵. Next, six calibration points were selected from the Timetree database⁵⁶, including the divergence time between *R. typus* and *C. harengus* (442.7–515.5 Mya), *D. rerio* and *P. hyalocranius* (180.0–264.0 Mya), *G. morhua* and *C. argus* (131.2–177.0 Mya), *T. fulvidraco* and *D. rerio* (132.0–170.0 Mya), *S. argus* and *C. argus* (103.7–176.0 Mya), and *P. flavescens* and *L. crocea* (59.3–142.9 Mya). The divergence times among the 19 species were then estimated using McMcTree (v4.9) in PAML⁵⁷. The ML tree clustered *N. japonicus* with three Sparidae species (*S. aurata*, *A. latus*, and *A. schlegelii*), with their divergence time estimated at 56.9–85.9 million years ago (Mya) (Fig. 3D), earlier than the divergence time among the three Sparidae species (17.3–41.2 Mya).

Positive selection analysis. This study established eight positive selection groups for genome-wide comparative analysis to identify candidate genes potentially linked to phenotypic traits in *N. japonicus*. Group 1 (*N. japonicus*, *A. ocellaris*, and *P. nyererei*) vs. (*M. amblycephala*, *A. schlegelii*, *P. hyalocranius*, and *C. harengus*) was

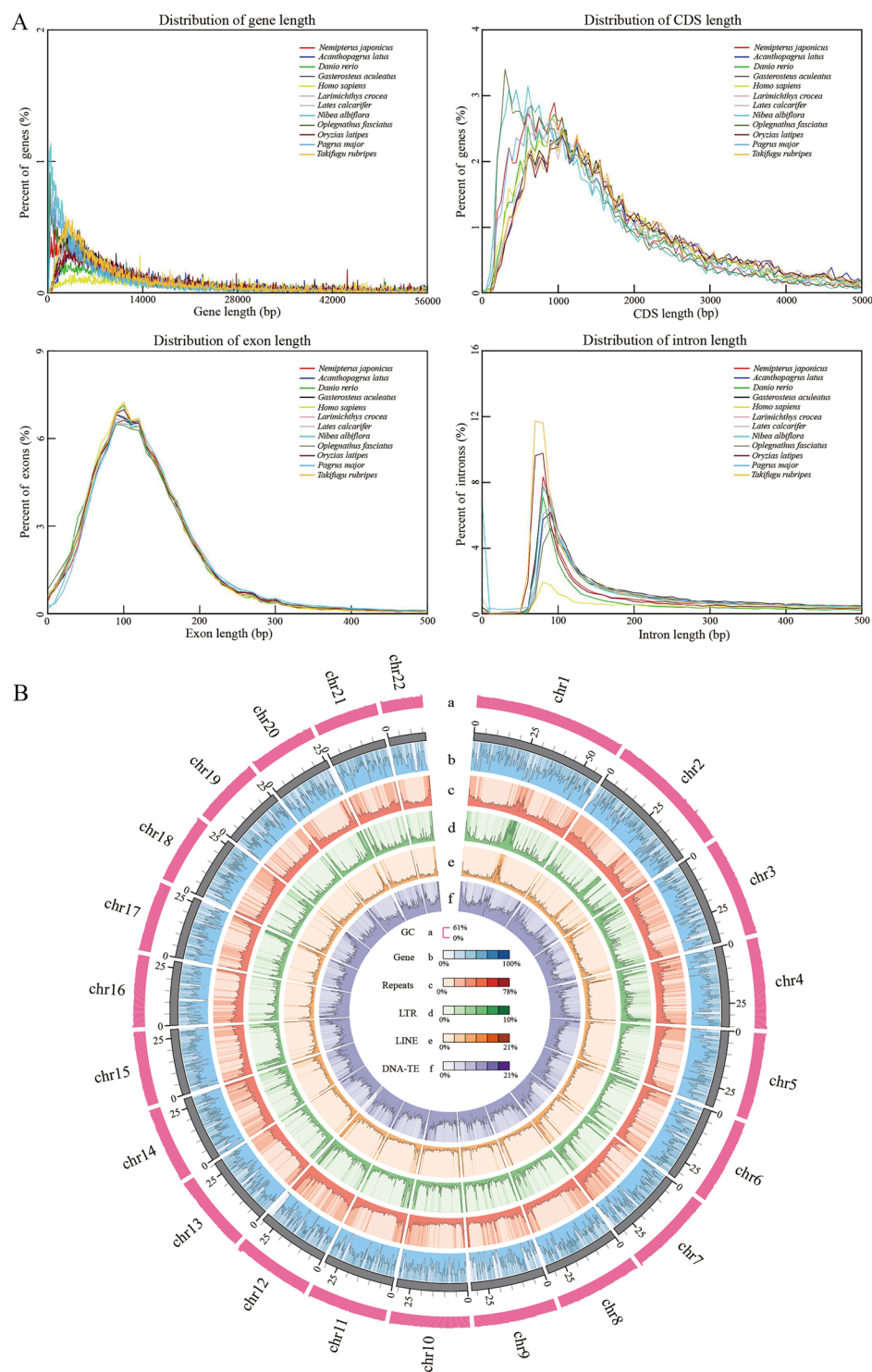


Fig. 2 The *N. japonicus* gene elements length distribution and genomic characteristics circle figure. **(A)** The comparisons of different gene elements in the *N. japonicus* genome with other fish species. **(B)** Circle figure of the genomic characteristics of *N. japonicus*, including (a) the GC content of the genome, (b) the distribution of genes, (c) the distribution of repeats, (d) the distribution of long tandem repeats, (e) the distribution of long interspersed nuclear elements, and (f) the distribution of DNA transposable elements.

designed to identify genes associated with the bright-red body coloration of the species. Group 2 (*N. japonicus*, *A. latus*, *P. flavescens*, *L. crocea*, and *T. fulvidraco*) vs. (*M. amblycephala*, *A. schlegelii*, *P. hyalocranius*, and *C. harengus*) and Group 3 (*N. japonicus*) vs. (*A. latus*, *P. flavescens*, *L. crocea*, *T. fulvidraco*, and *Sinocyclocheilus grahami*) were further screened for genes related to the yellow coloration and its position on the body surface of *N. japonicus*.

Type	Number	Percent (%)
Total	23514	—
Annotated	22858	97.21
InterPro	20722	88.13
GO	15729	66.89
KEGG_ALL	22465	95.54
KEGG_KO	14409	61.28
Swissprot	19931	84.76
TrEMBL	22630	96.24
TF	3574	15.2
Pfam	19944	84.82
NR	22800	96.96
KOG	18472	78.56
Unannotated	656	2.79

Table 6. Functional annotation of the genes in the *N. japonicus* genome.

Type	RepBase TEs		TE Protein		De novo		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	42,114,538	5.83	4,203,692	0.58	45,617,87	6.32	76,192,342	10.55
LINE	15,411,205	2.13	8,890,320	1.23	14,480,114	2.01	25,527,614	3.54
SINE	1,563,033	0.22	0	0.00	1,233,439	0.17	2,373,014	0.33
LTR	11,685,076	1.62	3,357,687	0.47	6,674,779	0.92	16,826,206	2.33
Satellite	4,728,176	0.65	0	0.00	8,395,252	1.16	10,138,320	1.40
Simple repeat	0	0.00	0	0.00	90,356	0.01	90,356	0.01
Other	7,119	0.00	831	0.00	0	0.00	7,950	0.00
Unknown	796,836	0.11	14,589	0.00	49,243,903	6.82	49,687,176	6.88
Total	68,173,879	9.44	16,449,475	2.28	112,695,424	15.61	157,659,348	21.84

Table 7. Statistics of repetitive sequences in the *N. japonicus* genome.

Type		Copy number	Average length(bp)	Total length(bp)	% in genome
miRNA		894	86	77,062	0.010675
tRNA		2,240	75	168,686	0.023367
rRNA	rRNA	2,215	122	269,715	0.037361
	18S	0	0	0	0.000000
	28S	0	0	0	0.000000
	5.8S	190	154	29,258	0.004053
	5S	2,025	119	240,457	0.033308
snRNA	snRNA	1,183	151	179,014	0.024797
	CD-box	167	119	19,869	0.002752
	HACA-box	76	149	11,300	0.001565
	splicing	932	157	145,939	0.020216
	scaRNA	8	238	1,906	0.000264

Table 8. Annotation statistics of the non-coding RNAs in the *N. japonicus* genome.

Group 4 (*N. japonicus*, *D. rerio*, *P. nyererei*, *A. ocellaris*, and *P. flavescens*) vs. (*E. akaara*, *S. grahami*, *S. argus*, and *C. argus*) focused on genes linked to striped or spotted body patterns. Group 5 (*N. japonicus*, *D. rerio*, and *T. fulvidraco*) vs. (*A. ocellaris*, *P. flavescens*, *C. rostratus*, *A. schlegelii*, and *P. nyererei*) targeted genes responsible for longitudinal or transverse body bands. Group 6 (*N. japonicus*, *A. latus*, *S. aurata*, and *A. schlegelii*) vs. (*C. argus*, *G. morhua*, *L. crocea*, *E. akaara*, and *S. argus*) examined caudal fin shape-related genes. Group 7 (*N. japonicus*, *A. latus*, and *A. schlegelii*) vs. (*C. harengus*, *G. morhua*, and *D. rerio*) investigated genes related to the presence or absence of fin spines. Group 8 (*N. japonicus*) vs. (*A. latus*, *A. schlegelii*, and *S. aurata*) detected positively selected genes between *N. japonicus* and its closely related species. Positive selection acting on protein-coding sequences was detected using CodeML (v4.9) in PAML⁵⁸ based on the branch-site model⁵⁹, revealing 163 positively selected genes, which were subjected to KEGG and GO enrichment analyses and visualization. Pathways

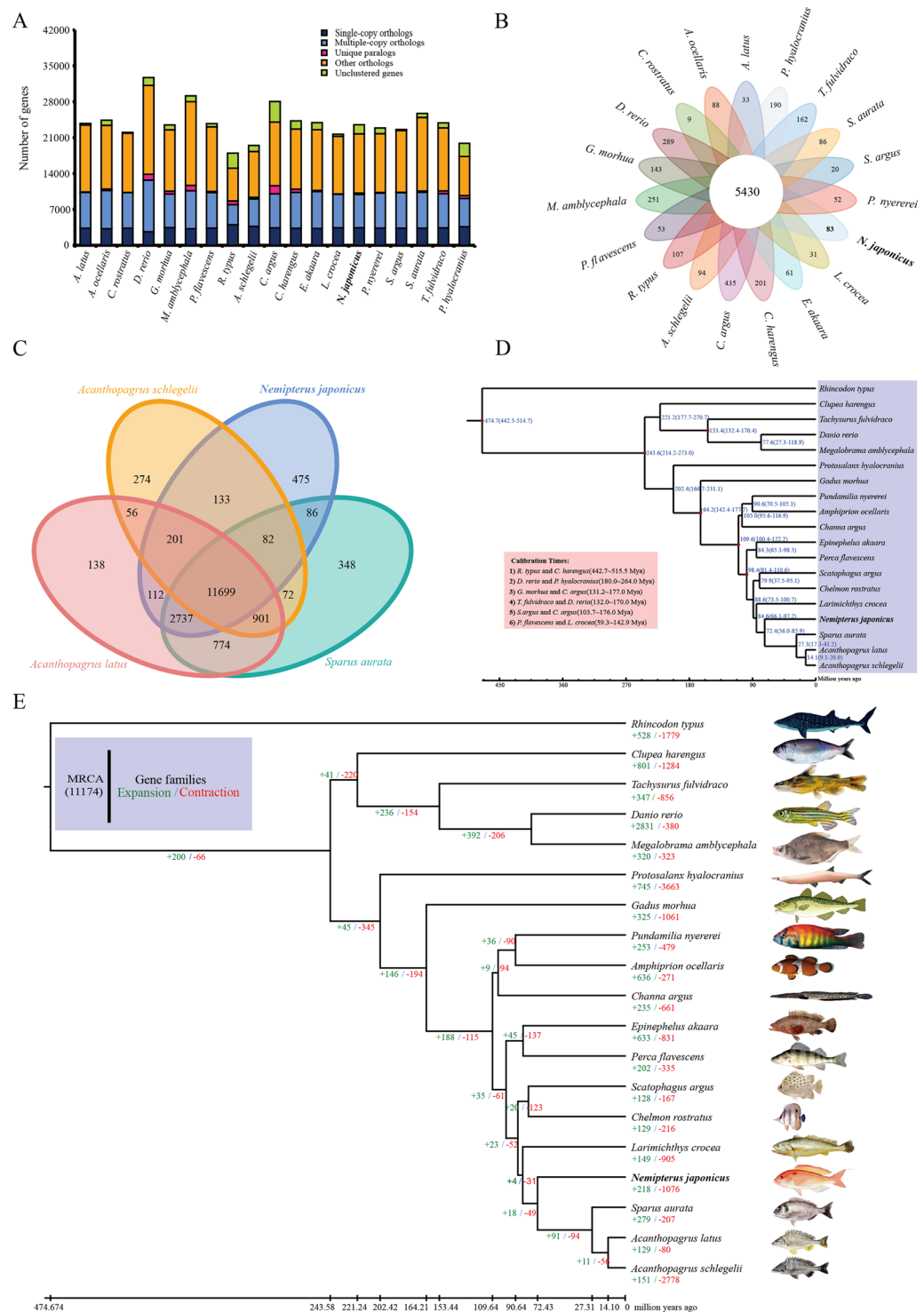


Fig. 3 The *N. japonicus* phylogenetic position and genome evolution. (A) Statistics of the orthologous gene families in 19 fish species. (B) Cluster petal map of gene families of 19 fish species. (C) Venn diagram of the number of homologous gene families in the four species. (D) Phylogenetic analysis and divergence time tree of *N. japonicus* and 18 other fish species. (E) Phylogenetic tree and gene family contractions and expansions for the 19 fish species. Green (+) and red (–) numbers represent the number of expanded and contracted gene families, respectively.

and terms meeting the criteria of $P < 0.05$ and $FDR < 0.05$ were filtered to explore the functions of these genes. Consequently, the main KEGG-enriched pathways included cytokine-cytokine receptor interaction, base excision repair, Fanconi anemia pathway, homologous recombination, mismatch repair, and Th1 and Th2 cell differentiation.

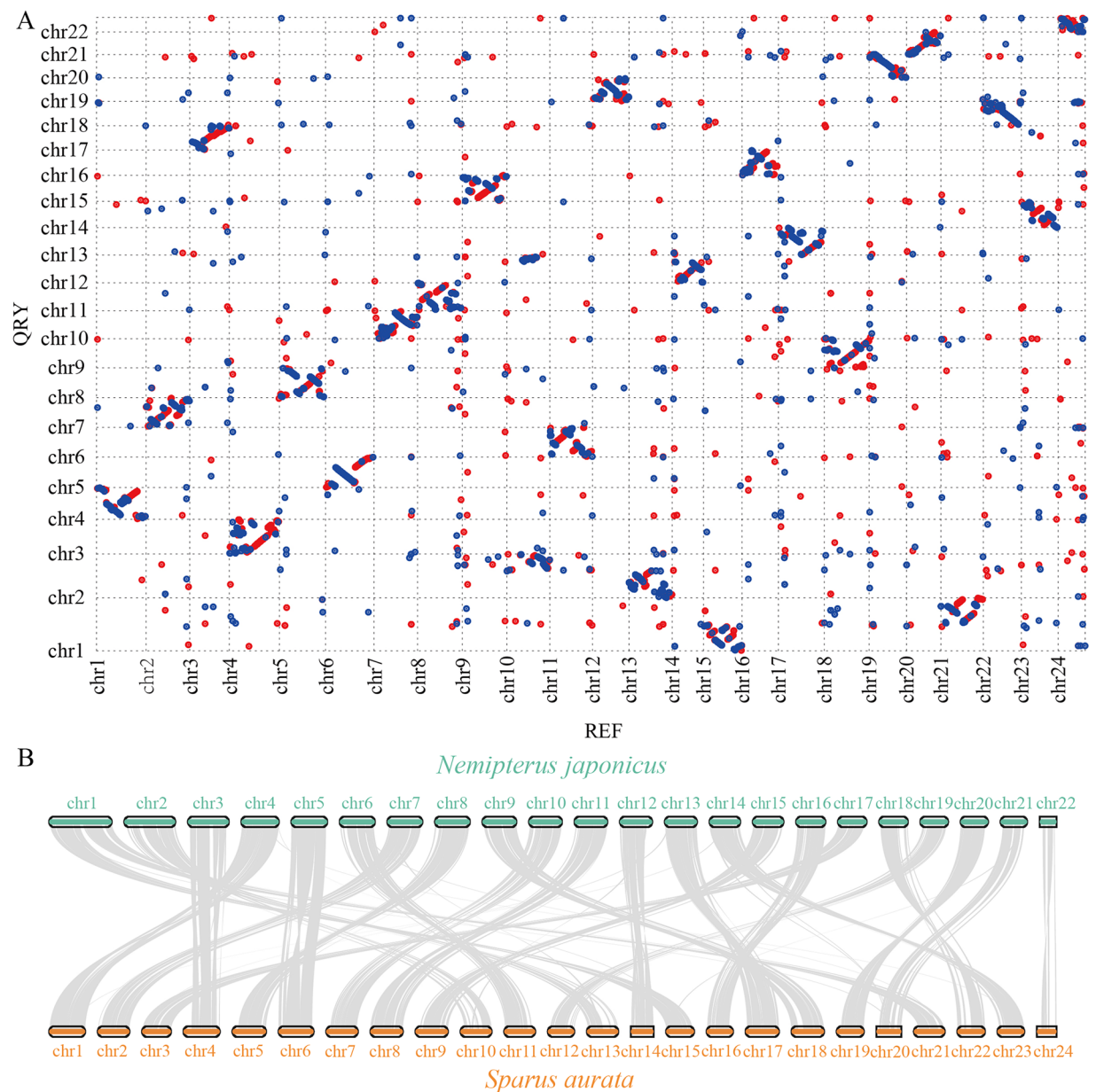


Fig. 4 Collinearity analyses of *N. japonicus* and *S. aurata* based on genome sequence (A) and coding sequence (B).

A combined functional analysis of positive selection genes with the biological traits of the species, identified several genes related to growth and development, interleukins, and DNA repair that have undergone significant evolution in the *N. japonicus* genome (Table S1). Specifically, the *LEPR*, *GHR*, *PRLR*, *PDGFD*, *BMP9*, *BMP14*, and *FOS* genes appear to be significantly involved in growth and skeletal development, lipid metabolism, and reproductive processes. The interleukin-related genes are dominated by the IL gene family, including *IL1B*, *IL1R1*, *IL1R2*, *IL1RAP*, *IL6ST*, *IL10RA*, *IL12B*, *IL12RB2*, *IL13RA2*, *IL17RB*, *IL17RC*, *IL17C*, and *IL17F*. DNA damage repair pathways include genes involved in interstrand cross-link repair (*FANCB*, *FANCD2*, *FANCL*, *FANCI*, and *UBE2T*), mismatch repair (*EXO1*, *RFC2*, *MLH3*, *PMS2*, and *SSBP1*), double-strand break and homologous recombination repair (*BRCA1*, *MRE11*, *NBN*, *RAD51D*, *RAD52*, *RAD54B*, *BRIP1*, *EME1*, *MUS81*, *SLX1A*, and *SLX4*), and translesion synthesis (*POLD3*, *POLE*, *POLH*, *POLK*, *REV1*, *PARP1*, *PARP2*, *PARP4*, and *XRCC1*).

Ethics declarations. The experimental animal protocols in the present study were reviewed and approved by the Animal Experimental Ethics Committee of Guangdong Ocean University, China (approval number: 202108001). Experiment procedures were performed in accordance with the Provisions and Regulations for the National Experimental Animal Management Regulations (China, July 2013) and the Experimental Animal Policies and Regulations of Guangdong Province (China, October 2010).

Type	Genome		Annotation	
	Proteins	Percentage (%)	Proteins	Percentage (%)
Complete BUSCOs (C)	3,558	97.75	3,507	96.3
Complete and single-copy BUSCOs (S)	3,505	96.29	3,459	95.0
Complete and duplicated BUSCOs (D)	53	1.46	48	1.3
Fragmented BUSCOs (F)	6	0.16	36	1.0
Missing BUSCOs (M)	76	2.09	97	2.7
Total BUSCO groups searched	3,640	100	3,640	100

Table 9. The completeness and accuracy of the *N. japonicus* genome assembly and annotation.

Data Records

The genome assembly data of *N. japonicus* has been deposited in public repositories^{60,61}. The DNA and RNA sequencing data are archived in the Sequence Read Archive (SRA) under BioProject number PRJNA1256535 and BioSample number SAMN48153383, with SRA accession number SRR33394854⁶², SRR33398644⁶³, SRR33393141⁶⁴, SRR33394373⁶⁵, SRR33406548⁶⁶ and SRR33419198⁶⁷. The genome annotation and comparative genomic analyses files have been deposited in figshare^{68,69}.

Technical Validation

Quality assessment of genomic DNA and RNA. The concentration, purity, and integrity of the extracted DNA were assessed using Qubit 3.0 fluorometer (Invitrogen, Carlsbad, CA, USA), Nanodrop 2000 microvolume spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and Agilent 4200 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), respectively. The integrity and concentration of extracted RNA were detected using Agilent RNA ScreenTape Assay (Agilent Technologies, Palo Alto, CA, USA) and Agilent 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). Equal amounts of RNA from four qualified tissues were pooled and rechecked for integrity, concentration, and purity. Subsequently, the high-quality DNA and pooled RNA samples that passed quality control were used for library preparation and sequencing.

Evaluation of genome assembly integrity. The genome sequences were fragmented and aligned to the NCBI nucleotide database using blastn (v2.11.0+) ⁷⁰ to confirm that the assembly result belonged to the target species. Illumina and PacBio sequencing data were then aligned to the assembled genome using the BWA (v0.7.12)⁷¹ and Minimap2 (v2.22)⁷² software, respectively. Alignment rates reached 99.35% with 99.97% coverage for Illumina clean reads, and 99.70% with 99.80% coverage for PacBio HiFi reads. A joint plot of GC content versus coverage depth distribution analysis showed a stable GC content of around 41.46% without significant GC bias, indicating no exogenous contamination in the genome assembly. Variants in the genome were identified using Samtools (v1.9)⁷³, Picard (v1.124) (<https://broadinstitute.github.io/picard/>), and GATK (v4.2.0.0)⁷⁴. The observed homozygous SNP (0.000%) and InDel (0.001%) rates were extremely low, indicating high assembly accuracy. Finally, the integrity of genome assembly based on the BUSCO (Benchmarking Universal Single-copy Orthologs method⁷⁵ was evaluated with Augustus (v3.3.2)³⁷ using 3,640 orthologous single-copy genes as references. The results revealed 3,558 (97.75%) complete BUSCO genes in the *N. japonicus* genome, including 3,505 (96.29%) complete single-copy and 53 (1.46%) complete duplicated genes (Table 9), demonstrating high gene coverage.

Data availability

The genome assembly data of *N. japonicus* has been deposited to the National Center for Biotechnology Information (NCBI) GenBank at JBNJRZ000000000 (https://identifiers.org/ncbi/insdc.gca:GCA_053574625.1)⁶⁰ and figshare (<https://doi.org/10.6084/m9.figshare.30630179>)⁶¹. The genomic Illumina sequencing data were available in the Sequence Read Archive at SRR33394854 (<https://identifiers.org/ncbi/insdc.sra:SRR33394854>)⁶². The genomic PacBio sequencing data were available in the SRA at NCBI SRR33398644 (<https://identifiers.org/ncbi/insdc.sra:SRR33398644>)⁶³. The RNA-seq data were available in the SRA at NCBI SRR33393141 (<https://identifiers.org/ncbi/insdc.sra:SRR33393141>)⁶⁴ and SRR33394373 (<https://identifiers.org/ncbi/insdc.sra:SRR33394373>)⁶⁵. The Hi-C sequencing data were available in the SRA at NCBI SRR33406548 (<https://identifiers.org/ncbi/insdc.sra:SRR33406548>)⁶⁶ and SRR33419198 (<https://identifiers.org/ncbi/insdc.sra:SRR33419198>)⁶⁷. The genome annotation files were available in figshare (<https://doi.org/10.6084/m9.figshare.29061083>)⁶⁸. The protein sequences used in the comparative genomic analyses and the output files including gene family clustering, and expansion and contraction analyses, and positive selection analysis were also available in figshare (<https://doi.org/10.6084/m9.figshare.29063051>)⁶⁹.

Code availability

No specific code was used in this study. All data analyses were performed using standard bioinformatics tools. The commands and parameter settings for each tool were executed according to their official manuals and established pipelines. Detailed descriptions have been provided in the Methods section.

Received: 5 September 2025; Accepted: 24 December 2025;
Published online: 06 January 2026

References

- Russell, B. C. FAO species catalogue. Vol. 12. Nemipterid fishes of the world. (Threadfin breams, Whiptail breams, Monocle breams, Dwarf monocle breams, and Coral breams). Family Nemipteridae. An annotated and illustrated catalogue of Nemipterid species known to date. *FAO Species Catalogue FAO Fisheries Synopsis No. 125*, **12** (Rome, FAO, <https://openknowledge.fao.org/handle/20.500.14283/t0416e> 1990).
- Chen, G. B., Li, Y. Z., Zhao, X. Y., Chen, Y. Z. & Jin, X. S. Acoustic assessment of five groups commercial fish in South China Sea. *Acta Oceanol. Sin.* **28**, 128–134, <https://doi.org/10.3321/j.issn.0253-4193.2006.02.016> (2006).
- Wu, X. M., Zhou, L. B. & Ou, L. X. Analysis on muscle nutrition component of *Nemipterus japonicus*. *Guangdong Agric. Sci.* **39**, 109–112, <https://doi.org/10.16768/j.issn.1004-874x.2012.06.071> (2012).
- Vagh, S. *et al.* Food and feeding habits of *Nemipterus japonicus* (Bloch, 1791) off Gujarat, north-west coast of India. *Indian J. Geo-Mar Sci.* **50**, 473–479, <https://doi.org/10.56042/ijms.v50i06.66274> (2021).
- Sudheesan, D., Jaiswar, A. K., Chakraborty, S. K. & Pazhayamadom, D. G. Predatory diversity of finfish species inhabiting the same ecological niche. *Indian J. Fish.* **56**, 169–175, <https://www.researchgate.net/publication/225066522> (2009).
- Wang, J. X. *et al.* Feeding Ecology of *Nemipterus japonicus* in Beibu Gulf. *Prog. Fish. Sci.* **44**, 47–57, <https://doi.org/10.19663/j.issn2095-9869.20210606001> (2023).
- Sreekanth, G. B. *et al.* Analysis of meristic characters of the Japanese threadfin bream, *Nemipterus japonicus* (Bloch, 1791) along Indian coast. *Indian J. Fish.* **60**, 119–121, <https://www.researchgate.net/publication/263587444> (2013).
- Sreekanth, G. B. *et al.* Comparison of length-weight progression for Japanese threadfin bream, *Nemipterus japonicus* (Bloch, 1791) from different regions along Indian coast using selected regression models. *Indian J. Geo-Mar Sci.* **43**, 1533–1542, <https://www.researchgate.net/publication/271198369> (2014).
- Sasikumar, R. G. *et al.* Biological Traits and Population Characteristics of Japanese Threadfin Bream, *Nemipterus japonicus* from the Coromandel Coast of Tamil Nadu, India. *Thalassas* **41**, 54, <https://doi.org/10.1007/s41208-025-00808-4> (2025).
- Ghosh, S. *et al.* Trophic interactions and the diet composition of sympatric finfish predators in the tropical demersal ecosystem in the Bay of Bengal. *J. Fish Biol.* **104**, 69–91, <https://doi.org/10.1111/jfb.15560> (2024).
- Tonnie, N. *et al.* Food and feeding habits of *Nemipterus japonicus* and *Nemipterus peronii* from coastal water of Bintulu, Sarawak, South China Sea. *J. Environ. Biol.* **39**, 857–864, [https://doi.org/10.22438/jeb/39/5\(SI\)/10](https://doi.org/10.22438/jeb/39/5(SI)/10) (2018).
- Nettely, T. *et al.* Reproductive biology of *Nemipterus japonicus* (Bloch, 1791) from the coastal waters of Bintulu (South China Sea), Sarawak, Malaysia. *J. Environ. Biol.* **37**, 715–724, <https://www.researchgate.net/publication/308360523> (2016).
- Rao, M. V. H. *et al.* Reproductive biology of *Nemipterus japonicus* (Bloch, 1791) in the trawl grounds along the north-east coast of India. *Indian J. Fish.* **64**, 21–27, <https://doi.org/10.21077/ijf.2017.64.4.62693-03> (2017).
- Joshi, K. K. Population dynamics of *Nemipterus japonicus* (Bloch) in the trawling grounds off Cochin. *Indian J. Fish.* **57**, 7–12, <https://www.researchgate.net/publication/267848214> (2010).
- Kalhor, M. A., Liu, Q., Memon, K. H., Chang, M. S. & Zhang, K. Population dynamics of Japanese threadfin bream *Nemipterus japonicus* from Pakistani waters. *Acta Oceanol. Sin.* **33**, 49–57, <https://doi.org/10.1007/s13131-014-0401-1> (2014).
- Dong, L. N. *et al.* Sequence analysis of mitochondrial COI gene in three *Nemipterus* species from Beibu Gulf. *J. Fish. Sci. China* **18**, 508–514, <https://doi.org/10.3724/SP.J.1118.2011.00508> (2011).
- Dong, L. N., Huang, Z. R., Li, Y. Z. & Chen, G. B. Sequence analysis of mitochondrial CO I and Cyt b gene of *Nemipterus* species in South China Sea. *J. Fish. Sci. China* **19**, 355–363, <https://doi.org/10.3724/SP.J.1118.2012.00355> (2012).
- Liang, R. S., Wang, C., Zou, Q., Zhou, A. G. & Zou, J. X. Molecular phylogenetic relationship among seven *Nemipterus* species based on partial S7 ribosomal protein gene sequences. *J. Fish. Sci. China* **20**, 506–513, <https://doi.org/10.3724/SP.J.1118.2013.00506> (2013).
- Liang, R. S. *et al.* Molecular Phylogenetic Relationship of 15 Nemipteridae Fishes Based on Mitochondrial COI and Nuclear RAG2 Gene Sequences. *J. Ocean Univ. China* **47**, 74–81, <https://doi.org/10.16441/j.cnki.hdx.20160407> (2018).
- Ning, P., Sha, Z., Hebert, P. D. N. & Russell, B. The Taxonomic Status of Japanese Threadfin Bream *Nemipterus japonicus* (Bloch, 1791) (Perciformes: Nemipteridae) with a Redescription of this Species from the South China Sea Based on Morphology and DNA Barcodes. *J. Ocean Univ. China* **14**, 178–184, <https://doi.org/10.1007/s11802-015-2609-x> (2015).
- Farivar, S., Jalil-Piran, Z., Zarei, F. & Hosseinzadeh Sahafi, H. Intraspecific phylogeography of the Japanese threadfin bream, *Nemipterus japonicus* (perciformes: nemipteridae), from the Persian Gulf and Indo-West Pacific: a preliminary study based on mitochondrial DNA sequence. *Iran. J. Fish. Sci.* **16**, 587–604, <https://www.researchgate.net/publication/303245132> (2017).
- Rao, M. V. H. *et al.* Dynamics of *Nemipterus japonicus* (Bloch) stocks along the north-east coast of India. *Indian J. Geo-Mar Sci.* **47**, 1761–1768, <http://nopr.niscair.res.in/handle/123456789/44924> (2018).
- Li, Y. L., Chen, J. H., Yan, B. L. & Meng, X. P. The complete mitochondrial genome of the Japanese threadfin bream, *Nemipterus japonicus* (Teleostei, Nemipteridae). *Mitochondrial DNA A* **27**, 429–430, <https://doi.org/10.3109/19401736.2014.898290> (2014).
- Sambrook, J. & Russell, D. W. The Inoue Method for Preparation and Transformation of Competent *E. Coli*: “ultra-competent” cells. *CSH Protocols* **2006**, pdb.prot3944–pdb.prot3944, <https://doi.org/10.1101/pdb.prot3944> (2006).
- Quail, M. A., Swerdlow, H. & Turner, D. J. Improved Protocols for the Illumina Genome Analyzer Sequencing System. *Curr. Protoc. Hum. Genet.* **18**, Unit 18.2, <https://doi.org/10.1002/0471142905.hg1802s62> (2009).
- Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890, <https://doi.org/10.1093/bioinformatics/bty560> (2018).
- Gordon, S. P. *et al.* Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. *PLoS ONE* **10**, e0132628, <https://doi.org/10.1371/journal.pone.0132628> (2015).
- 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).
- Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W. & 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).
- Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460, <https://doi.org/10.1186/s12859-018-2485-7> (2018).
- Wingett, S. W. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310, <https://doi.org/10.12688/f1000research.7334.1> (2015).
- Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95, <https://doi.org/10.1126/science.aal3327> (2017).
- 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).
- Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101, <https://doi.org/10.1016/j.cels.2015.07.012> (2016).
- Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067, <https://doi.org/10.1093/bioinformatics/btm071> (2007).
- Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995, <https://doi.org/10.1101/GR.1865504> (2004).
- Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439, <https://doi.org/10.1093/NAR/GKL200> (2006).

38. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562–578, <https://doi.org/10.1038/nprot.2012.016> (2012).
39. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491, <https://doi.org/10.1186/1471-2105-12-491> (2011).
40. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinform.* **25**, Unit 4.10, <https://doi.org/10.1002/0471250953.bi0410s25> (2009).
41. Bergman, C. M. & Quesneville, H. Discovering and detecting transposable elements in genome sequences. *Brief. Bioinform.* **8**, 382–392, <https://doi.org/10.1093/bib/bbm048> (2007).
42. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467, <https://doi.org/10.1159/000084979> (2005).
43. Zhao, X. & Hao, W. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–8, <https://doi.org/10.1093/nar/gkm286> (2007).
44. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358, <https://doi.org/10.1093/bioinformatics/bti1018> (2005).
45. 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).
46. 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).
47. 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).
48. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124, <https://doi.org/10.1093/nar/gki081> (2004).
49. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9, <https://doi.org/10.1093/nar/gkn201> (2008).
50. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* **13**, 2178–2189, <https://doi.org/10.1101/gr.1224503> (2003).
51. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518, <https://doi.org/10.1093/BIOINFORMATICS/BTAA1022> (2021).
52. Delcher, A. L., Salzberg, S. L. & Phillippy, A. M. Using MUMmer to Identify Similar Regions in Large Sequence Sets. *Curr. Protoc. Bioinform.* **10**, 10.3.1–10.3.18, <https://doi.org/10.1002/0471250953.bi1003s00> (2003).
53. Tang, H. *et al.* Synteny and Collinearity in Plant Genomes. *Science* **320**, 486–488, <https://doi.org/10.1126/science.1153917> (2008).
54. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492, <https://doi.org/10.1093/bioinformatics/bty121> (2018).
55. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, <https://doi.org/10.1093/bioinformatics/btu033> (2014).
56. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819, <https://doi.org/10.1093/molbev/msx116> (2017).
57. Yang, Z. H. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13**, 555–556, <https://doi.org/10.1093/bioinformatics/13.5.555> (1997).
58. Yang, Z. H. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591, <https://doi.org/10.1093/molbev/msm088> (2007).
59. Yang, Z. H. & Nielsen, R. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Mol. Biol. Evol.* **19**, 908–917, <https://doi.org/10.1093/oxfordjournals.molbev.a004148> (2002).
60. Lei, F. L. Genbank https://identifiers.org/ncbi/insdc.gca:GCA_053574625.1 (2025).
61. Lei, F. L. The genome assembly of *Nemipterus japonicus*. [figshare https://doi.org/10.6084/m9.figshare.30630179](https://doi.org/10.6084/m9.figshare.30630179) (2025).
62. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33394854> (2025).
63. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33398644> (2025).
64. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33393141> (2025).
65. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33394373> (2025).
66. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33406548> (2025).
67. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR33419198> (2025).
68. Lei, F. L. Genome annotation of *Nemipterus japonicus*. [figshare https://doi.org/10.6084/m9.figshare.29061083](https://doi.org/10.6084/m9.figshare.29061083) (2025).
69. Lei, F. L. Comparative genomic analysis of *Nemipterus japonicus*. [figshare https://doi.org/10.6084/m9.figshare.29063051](https://doi.org/10.6084/m9.figshare.29063051) (2025).
70. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410, [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
71. 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).
72. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100, <https://doi.org/10.1093/bioinformatics/bty191> (2018).
73. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, <https://doi.org/10.1093/bioinformatics/btp352> (2009).
74. McCormick, R. F., Truong, S. K. & Mullet, J. E. RIG: Recalibration and Interrelation of Genomic Sequence Data with the GATK. *G3-Genes Genom. Genet.* **5**, 655–665, <https://doi.org/10.1534/g3.115.017012> (2015).
75. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212, <https://doi.org/10.1093/bioinformatics/btv351> (2015).

Acknowledgements

This research was funded by the National Natural Science Foundation of China (41006084) and Project for Outstanding Young Teachers in Higher Education of Guangdong, China (Yq2013093). We thank Wuhan Wanmo-Tech Co., Ltd. for their technical support in genome sequencing and comparative genomic analysis.

Author contributions

Conceptualization, R.-X.Wu and F.-L.Lei.; methodology, F.-L.Lei, R.-X.Wu and F.-Y.Han; software, F.-L.Lei, F.-L.Long and F.-Y.Han; validation, F.-L.Lei, F.-L.Long and R.-X.Wu; formal analysis, F.-L.Lei, R.-X.Wu, F.-L.Long and F.-Y.Han; investigation, F.-Y.Han, F.-L.Long and Z.-B.Liang; resources, R.-X.Wu, F.-Y.Han and Z.-B.Liang; data curation, F.-L.Lei, R.-X.Wu and F.-L.Long; writing-original draft preparation, F.-L.Lei, R.-X.Wu and F.-Y.Han; writing-review and editing, F.-L.Lei and R.-X.Wu; visualization, F.-L.Lei and R.-X.Wu; supervision, R.-X.Wu; project administration, R.-X.Wu; funding acquisition, R.-X.Wu. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-06536-x>.

Correspondence and requests for materials should be addressed to R.-X.W.

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

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



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

© The Author(s) 2026