Supplementary Materials for

Stable genome structures in living fossil fishes

Cheng Wang[1#], Chase D. Brownstein[2#], Wejun Chen[1,3], Zufa Ding[1], Dan Yu[1], Yu Deng[1], Chenguang Feng[1], Thomas J. Near[2,4*], Shunping He[1*], Liandong Yang[1*]

[*] Corresponding author. Email: thomas.near@yale.edu; clad@ihb.ac.cn; yangliandong1987@163.com

Supplemental Results

Positive selection genes for DNA repair and genome stability in gar microchromosomes

To uncover potential drivers of genome stability in gars, we identified 91 genes on microchromosomes that are under positive selection in comparison to those in other vertebrates possessing microchromosomes, including *Gallus gallus*, *Chelonia mydas*, *Naja naja*, *Acipenser ruthenus*, and *Amblyraja radiata*. The three gar species exhibited lower ω (*K*a/*K*s) values than other taxa, indicating slower amino acid substitution rates in the former (Supplemental Fig. S12A). Among positively selected genes (PSGs), several stand out for their roles in genome stability maintenance. *CCP110*, a physiological centrosomal cyclin-dependent kinase, functions in centrosome duplication and regulates cytokinesis (Chen et al. 2002; Tsang et al. 2006) (Supplemental Fig. S12B). *MPG*, located in the cytosol and nucleoplasm, is involved in base excision repair, correcting DNA damage caused by oxidation, deamination, and alkylation (Zharkov 2008; Krokan and Bjørås 2013) (Supplemental Fig. S12C). *NLK*, a serine/threonine-protein kinase, is essential for *TP53* activation in response to DNA damage (Zhang et al. 2014). *NTHL1* produces a protein that excises DNA base damage caused by reactive oxygen species (Limpose et al. 2018; Das et al. 2020) (Supplemental Table S10). All these PSGs are all located on gar microchromosomes.

Comparative analysis of their expression levels revealed that these PSGs exhibit relatively higher expression compared to genes on macrochromosomes and non-PSG genes on microchromosomes. This finding suggests that these PSGs regulate genome stability and may partially explain the genomic conservation observed in gars (Supplemental Fig. S12D-E).

Supplemental Methods

Ethics statement, sampling and sequencing

All sampling and experimental procedures were approved by the Ethics Committee of the Institute of Hydrobiology, Chinese Academy of Sciences. We collected *Atractosteus spatula* and *Lepisosteus osseus* specimens from an aquarium market in Wuhan, Hubei Province, China. We extracted DNA from muscle tissues of *A. spatula* and *L. osseu*s using a modified cetyltrimethylammonium bromide (CTAB) method (Porebski et al. 1997). We extracted RNA from the eye tissue of *A. spatula* and from the liver, heart, brain, eyes, muscle, spleen, fin, gill, swim bladder, intestine, and ovary tissues of *L. osseus* using a TRIzol kit.

For long-read sequencing of *A. spatula*, we generated and sequenced nanopore libraries on 13 flow cells using a GridION X5 DNA sequencer (Oxford Nanopore). For the *L. osseus* assembly, we constructed and sequenced a SMRT cell sequencing library containing fragments approximately 15–20 kb in size on a PacBio Sequel II platform. We prepared high-throughput sequencing libraries using Illumina technology (Illumina, San Diego, USA) with 350 bp inserts. We generated RNA sequencing libraries with an NEBNext® Ultra™ RNA Library Prep Kit (Illumina®) following the manufacturer's recommendations and sequenced RNA on an Illumina platform (Wuhan Benagen Technology Co., Ltd.), yielding 150 bp paired-end reads. We prepared Hi-C libraries from liver tissue and conducted sequencing on an Illumina NovaSeq 6000 platform.

Genome assembly and evaluation

For all Illumina sequencing data, we removed low-quality reads, duplicated reads, and adapter sequences using fastp (v0.23.4) (Chen 2023) and Trimmomatic (v0.39) (Bolger et al. 2014) with default parameters. We then used filtered Illumina reads to estimate genome size based on *k*-mer analysis and conducted a 17-mer depth frequency distribution analysis. We calculated genome sizes using the formula: Genome size = total *k*-mer count / peak *k*-mer frequency depth.

For the *A. spatula* genome assembly, we used NextDenovo (v2.5.2) (Hu et al. 2024) to correct the Nanopore long reads and assemble the genome with default parameters. Next, we used NextPolish (v1.4.1) (Hu et al. 2020) to polish the draft assembly with both Nanopore and Illumina reads. To obtain the chromosome-level genome, we aligned clean Hi-C reads to the genome and ordered and oriented into chromosomes using Lachesis (Burton et al. 2013) with default parameters.

For the *L. osseus* genome assembly, we used HiFi long reads for draft genome assembly in hifiasm (v0.18.5) (Cheng et al. 2021) using default parameters. We corrected errors using gcpp in the SMRT Link (v8.0) toolkit and polished sequences with Racon (v1.4.3) (Vaser et al. 2017). We filtered redundant sequences using Purge Haplotigs (v1.2.5) (Roach et al. 2018) and filled gaps using TGS-GapCloser (v1.2.1) (Xu et al. 2020). We used the cleaned Hi-C data to scaffold the genome assembly to the chromosome level. We mapped clean paired-end reads to the genome using BWA (v0.7.15) (Li and Durbin 2009) and used the Juicer (Durand et al. 2016b) and 3D-DNA (v180922) (Olga Dudchenko 2017) workflow to produce the chromosome-level genome assembly, which we then manually reviewed and refined with Juicebox (v1.6) (Durand et al. 2016a) Assembly Tools.

Repeat and protein-coding gene annotation

For repeat sequence annotation, we used a combination of a custom *de novo* repeat sequence library and a public repeat sequence database to predict repeat elements in assembled genome (Liu 2018). First, we ran RepeatModeler (v2.0.3) (Flynn et al. 2020) on the genome assemblies with default parameters to generate a custom repeat

library. We combined this custom library with the repeat database and detected transposable elements (TEs) using RepeatMasker (v4.1.4) (Tarailo-Graovac and Chen 2009). Next, we determined the age profile of TEs by calculating the Kimura distances of individual TE copies from their corresponding consensus sequences.

For gene prediction, we soft-masked repetitive sequences in the genome using the maskfasta -soft command in BEDTools (v2.31) (Quinlan and Hall 2010).

We used a combination of *ab initio* gene prediction, homology-based gene prediction, and transcript-based prediction to identify protein-coding genes in gars. We predicted genes *ab initio* using AUGUSTUS (v3.4.0) (Hoff and Stanke 2019) using models trained on high-quality protein data generated from RNA-seq data processed by PASA (v2.4.1) (Haas et al. 2003). For homology-based prediction, we used Exonerate (v2.4.0) (Slater and Birney 2005) and GenomeThreader (v1.7.3) (Gremme et al. 2005) to align homologous protein sequences to the assembled genomes and predict coding sequences. Reference species for homology were *Danio rerio*, *Amia calva*, *Gallus gallus*, *Cetorhinus maximus*, *Homo sapiens*, and *Lepisosteus oculatus*.

For transcript-based prediction, we assembled clean RNA-seq reads into transcripts on our genomes using StringTie (v2.1.1) (Pertea et al. 2015) and refined gene structures using PASA (v2.4.1) (Haas et al. 2003). After completing all predictions, we integrated the three predicted gene structures using EvidenceModeler (v2.1.0) (Haas et al. 2008). Finally, we updated the combined results to add alternative splicing and untranslated regions (UTRs) with PASA (v2.4.1) (Haas et al. 2003).

For gene functional annotation, we aligned the predicted genes to the NR (non-redundant protein database), eggNOG, Pfam, Swiss-Prot, and InterPro databases using BLASTP (v2.6.0) (Camacho et al. 2009)

Population Structure and Admixture Among Gars

To examine population structure and infer episodes of historical introgression among living gar species, we collected single nucleotide polymorphism (SNP) data for

individuals of *A. spatula*, *L. osseus*, and *L. oculatus*. Muscle samples were collected from a total of 79 individual that collected from an aquarium market in Wuhan, Hubei Province, China, 23 *L. osseus*, 22 *L. oculatus*, 24 *A. spatula*. To ensure the accuracy of species, we conducted both morphological characteristic identification and molecular identification of COI sequences. We filtered raw sequences using Trimmomatic (v0.39) (Bolger et al. 2014), and mapped the clean paired-end reads to the *A. spatula* genome with BWA-MEM2 (v0.7.17) (Li 2013) using default parameters. We used SAMtools (v1.16.1) (Li et al. 2009) and Picard (v3.1) (Ebbert et al. 2016) to sort the BAM files and remove duplicated reads. We called variants of each sample and filtered SNPs using GATK (v 4.0.5) (McKenna et al. 2010). The SNP filter criteria were "QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0". After the initial filtering step, we obtained a total of 44,298,904 SNPs (Ts/Tv=1.31). Using VCFtools (v0.1.13) (Danecek et al. 2011), we calculated sequencing depth of the SNPs, the proportion of missing sites, site quality, individual heterozygosity and inbreeding coefficient, and allele frequency. We conducted another round of filtering parameters in VCFtools with the following parameters: "--maf 0.3 --max-missing 0.9 --minQ 30 --minDP 15 --maxDP 60". We obtained a total of 15,305,348 high quality SNPs (Ts/Tv=1.48) for downstream analysis.

We examined the ancestry of gar individuals using Admixture (v1.3.0) (Alexander et al. 2009). We selected a K value = 3 as optimal according to the CV errors. Next, we conducted a principal component analysis of the SNP dataset using PLINK (v1.90b7) (Purcell et al. 2007), and constructed a phylogenetic tree using FastTree (v2.1) (Price et al. 2009) and PHYLIP (v3.697) (Alzohairy 2009). We calculated linkage disequilibrium values across all chromosomes using PopLDdeacy (v3.42) (Zhang et al. 2019) with default parameters. We calculated pairwise fixation indices (*F*st) and average nucleotide diversity (π) with a 100kb window size and 10kb step size using VCFtools (v0.1.13) (Danecek et al. 2011). For identity-by-descent (IBD) analysis, we used Beagle (v4.0) (Browning and Browning 2013) with the parameter "window=100000 overlap=10000 ibdtrim=100 ibdold=10" to phase and

impute the SNP data, and uncover shared IBD blocks between species. In order to explore whether one gar species had admixed with the other two populations, we performed the F3 test using the “qp3Pop” function in Admixtools (v8.0.2) and “threepop” in Treemix (v1.13) (Molloy et al. 2021). In order to infer population splits and ancient migration events, we used the phylogeny built using the SNP data and inferred migration edges to account for gene flow among populations in Treemix (v1.13). We also used Momi2 (v2.1.19) (Kamm et al. 2020) to explore demographic models for various sets of three populations. For this analysis, we filtered out SNPs corresponding to coding and repetitive sequence regions, fitted 5 independent runs with random starting parameters, and selected the demographic model with the biggest log-likelihood value of all runs. Finally, we calculated bootstraps supports for the best model by resampling blocks of the SFS (site frequency spectrum) to generate confidence intervals.

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**: 1655-1664.

Alzohairy A. 2009. Phylip and Phylogenetics. *Genes, genomes and genomics* **3**: 46-49.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Browning BL, Browning SR. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**: 459-471.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol* **31**: 1119-1125.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**: e107.

Chen Z, Indjeian VB, McManus M, Wang L, Dynlacht BD. 2002. CCP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells. *Dev Cell* **3**: 339-350.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**: 170-175.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156-2158.

Das L, Quintana VG, Sweasy JB. 2020. NTHL1 in genomic integrity, aging and cancer. *DNA Repair* **93**: 102920.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016a. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**: 99-101.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016b. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**: 95-98.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* **117**: 9451-9457.

Gremme G, Brendel V, Sparks ME, Kurtz S. 2005. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology* **47**: 965-978.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Jr., Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD et al. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**: 5654-5666.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**: R7.

Hoff KJ, Stanke M. 2019. Predicting Genes in Single Genomes with AUGUSTUS. *Curr Protoc Bioinformatics* **65**: e57.

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**: 2253-2255.

Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K et al. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol* **25**: 107.

Kamm J, Terhorst J, Durbin R, Song YS. 2020. Efficiently inferring the demographic history of many populations with allele count data. *J Am Stat Assoc* **115**: 1472-1487.

Krokan H, Bjørås M. 2013. Base Excision Repair. *Cold Spring Harbor perspectives in biology* **5**.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Quantitative Biology*.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Limpose KL, Trego KS, Li Z, Leung SW, Sarker AH, Shah JA, Ramalingam SS, Werner EM, Dynan WS, Cooper PK et al. 2018. Overexpression of the base excision repair NTHL1 glycosylase causes genomic instability and early cellular hallmarks of cancer. *Nucleic Acids Res* **46**: 4515-4532.

Liu X. 2018. *An analytical pipeline of assembly and annotation of the Betta splendens genome. v1*.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297-1303.

Molloy EK, Durvasula A, Sankararaman S. 2021. Advancing admixture graph estimation via maximum likelihood network orientation. *Bioinformatics* **37**: i142-i150.

Olga Dudchenko, 2,3,4 Sanjit S. Batra,1,2,3* Arina D. Omer,1,2,3* Sarah K. Nyquist,1,3 Marie Hoeger,1,3 Neva C. Durand,1,2,3 Muhammad S. Shamim,1,2,3 Ido Machol,1,2,3 Eric S. Lander,5,6,7 Aviva Presser Aiden,1,2,8,9 Erez Lieberman Aiden1,2,3,4,5†. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *science*.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290-295.

Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* **15**: 8-15.

Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641-1650.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**: 559-575.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**: 460.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics* **6**: 31 doi:10.1186/1471-2105-6-31.

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**: Unit 4 10.

Tsang WY, Spektor A, Luciano DJ, Indjeian VB, Chen Z, Salisbury JL, Sánchez I, Dynlacht BD. 2006. CCP110 cooperates with two calcium-binding proteins to regulate cytokinesis and genome stability. *Mol Biol Cell* **17**: 3423-3434.

Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**: 737-746.

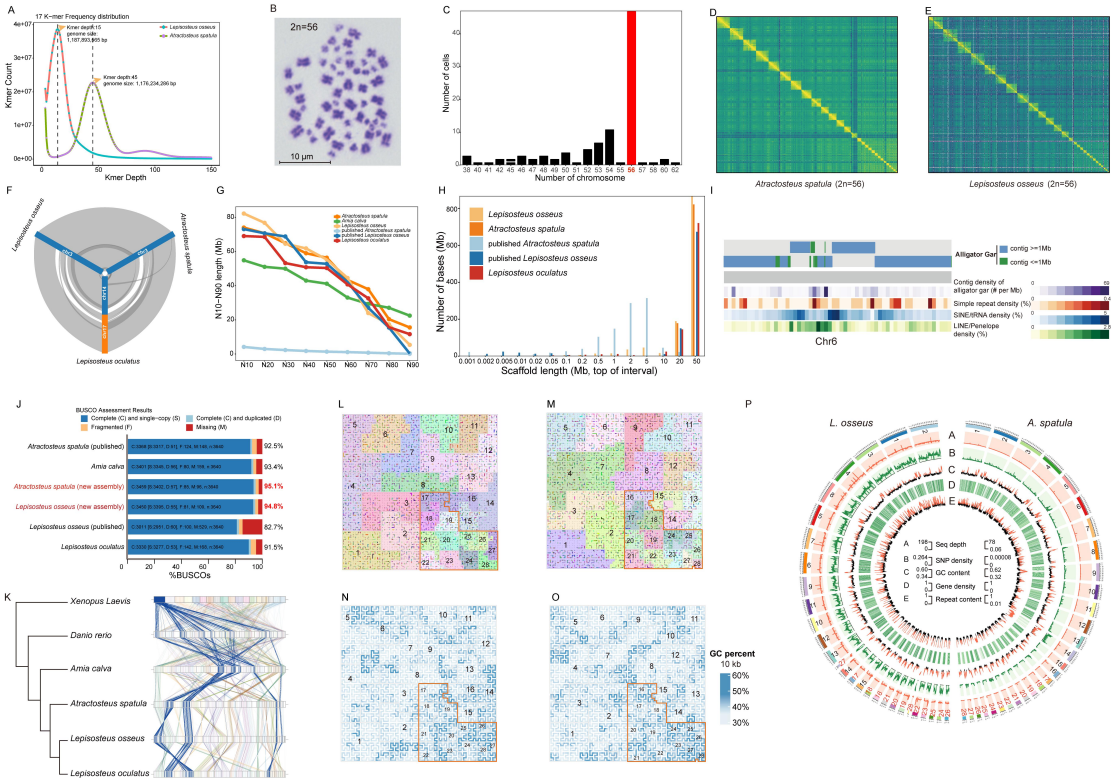
Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, Fan G, Liu X, Xu X, Deng L et al. 2020. TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **9**.

Zhang C, Dong SS, Xu JY, He WM, Yang TL. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**: 1786-1788.

Zhang HH, Li SZ, Zhang ZY, Hu XM, Hou PN, Gao L, Du RL, Zhang XD. 2014. Nemo-like kinase is critical for p53 stabilization and function in response to DNA damage. *Cell Death & Differentiation* **21**: 1656-1663.

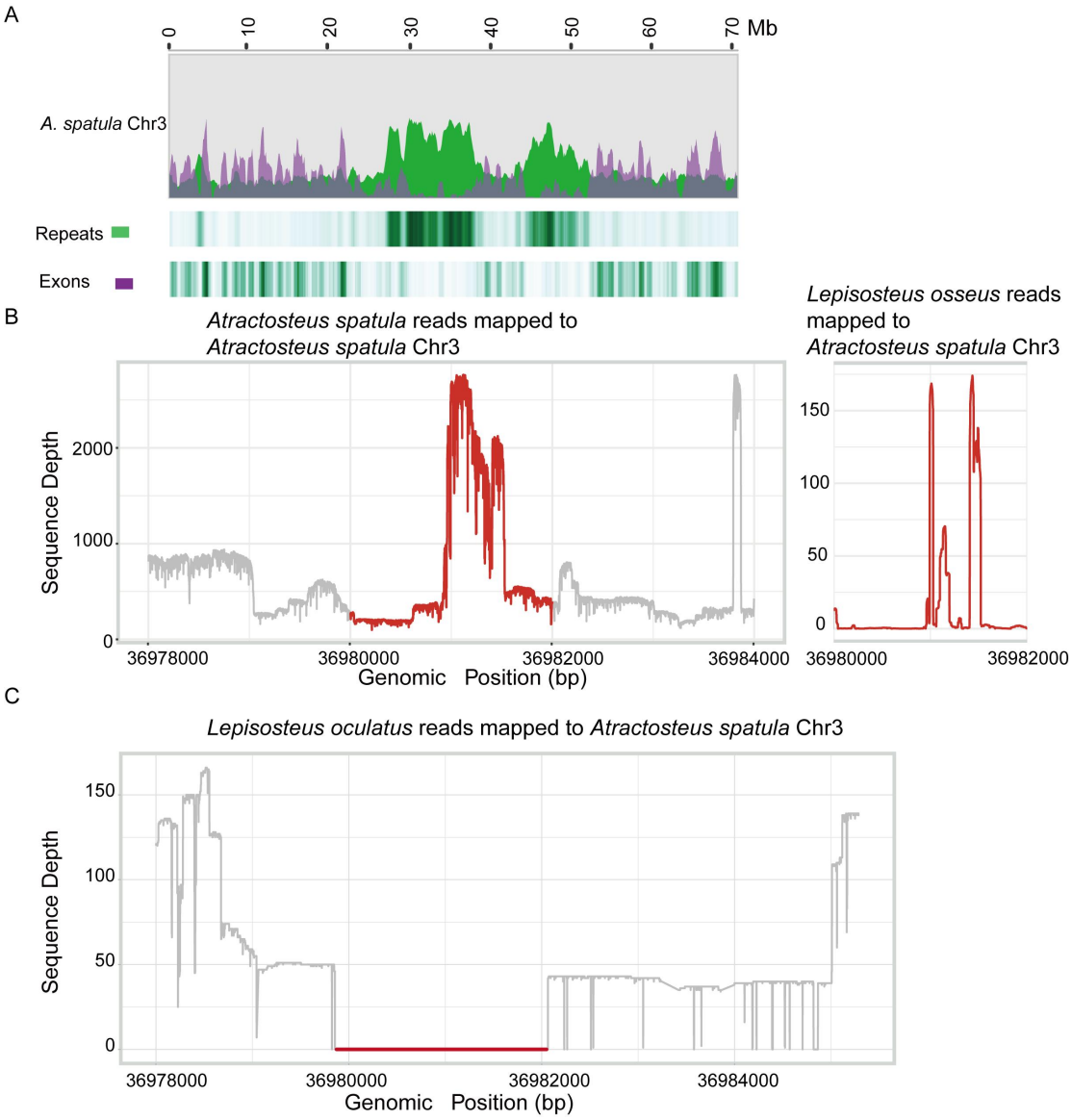
Zharkov DO. 2008. Base excision DNA repair. *Cellular and Molecular Life Sciences* **65**: 1544-1565.

Supplementary Figures



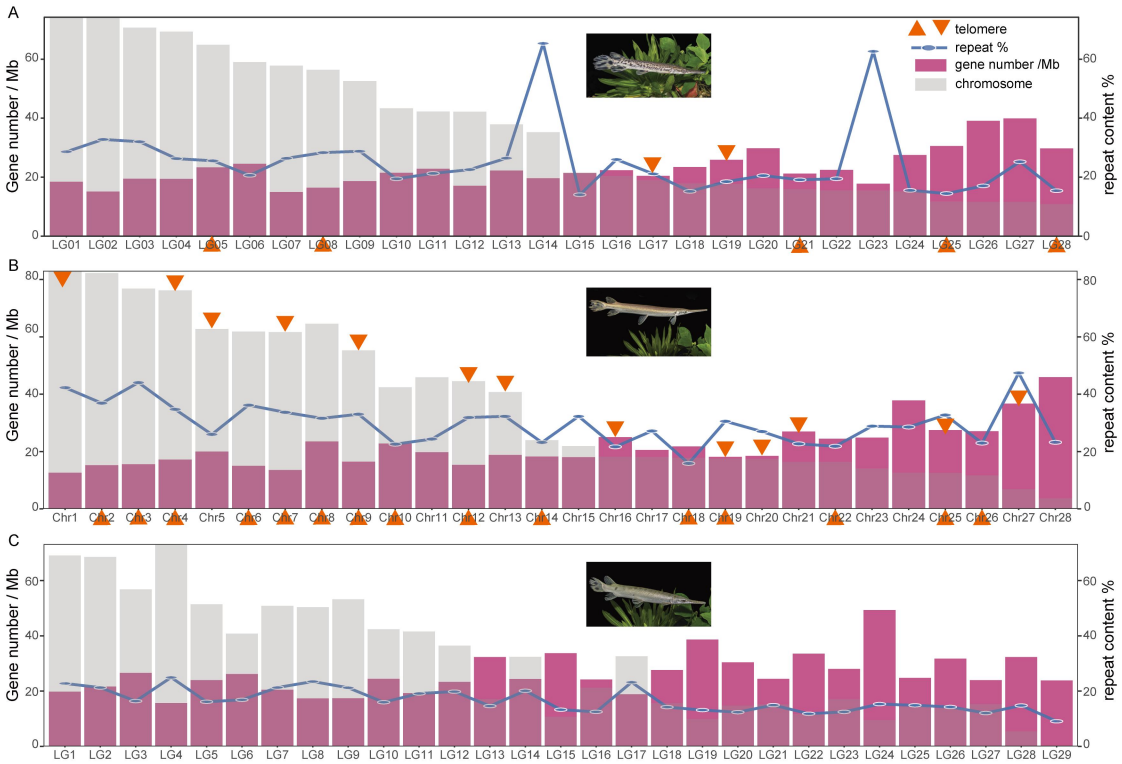
Supplemental Fig. S1:
Genomic landscape analysis of Alligator gar *Atractosteus spatula* and Longnose gar *Lepisosteus osseus*. (*A*) *K*-mer based genome survey of *L. osseus* and *A. spatula*. The frequency distribution plots of *k*-mer count and depth (*k*-mer = 17). (*B*) Homologous pairing karyotype analysis in *A. spatula*. (*C*) A bar graph displaying the absolute numbers of Alligator Gar chromosomal counts as determined by cell counting. (*D-E*) The assembled chromosome Hi-C heatmaps (n=28) of *A. spatula* and *L. osseus*, respectively. High-resolution heatmaps show distinct intra-chromosomal interaction and chromosomal territories. (*F*) the gene synteny map of the three gars shows that Chromosomes 14 and 17 of *Lepisosteus oculatus* have fused with Chromosome 3 of *L. osseus* and *A. spatula*. (*G*) N10-N90 size comparison of *A. spatula*, *Amia calva*, *L. osseus* and *L. oculatus*. (*H*) comparison of sequence length distribution among *A. spatula*, *L. osseus*, and *L. oculatus* genome assemblies, including both chromosome-level and published non-chromosome-level assemblies. (*I*) an example demonstrating that our newly assembled *A. spatula* genome is more continuous than previously published versions. In all cases, a single newly assembled *A. spatula* contig corresponds to multiple contigs in the published assembly, with regions of high repeat content located at the breakpoints between these contigs. (*J*) Evaluation of the completeness of the assembled genomes using BUSCO (actinopterygii_odb10 database). (*K*) comparative analysis of BUSCO gene synteny across the phylogenetic diversity of bony fishes. (*L-M*) The gene distribution across the chromosomes of the *A. spatula* and *L. osseus* visualizes that microchromosomes tend to exhibit a more concentrated distribution of genes compared to the

macrochromosomes. (*N*-*O*) The GC content distribution across the chomosomes of the *A. spatula* and *L. osseus*, respectively. The microchromosomes exhibit higher GC content than the macrochromosomes. (*P*) the genomic landscapes of *A. spatula* and *L. osseus*. A: sequence depth; B: SNP density; C: GC content; D: Gene density; E: Repeat content.



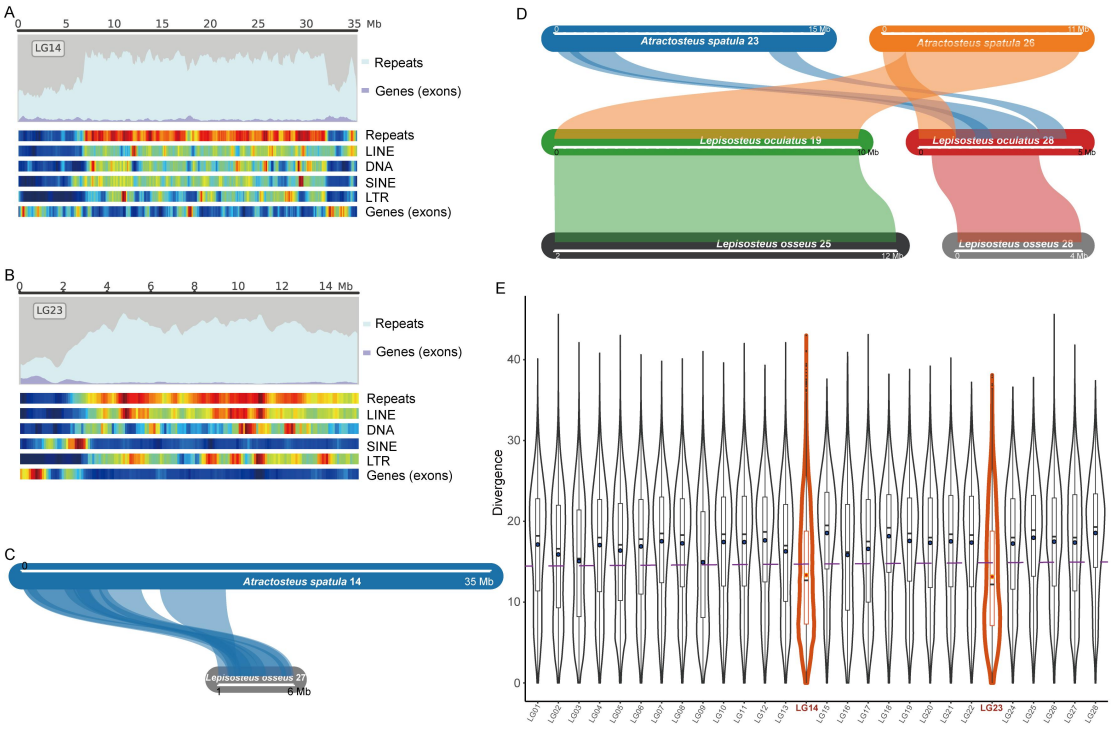
Supplemental Fig. S2:

Sequence depth analysis and chromosomal structural features confirm the accuracy of chromosomal fusion. (*A***)** Analyzing repetitive sequences and gene density on Chromosome 3 of *A. spatula*. (*B*) Sequencing reads of *A. spatula* and *L. osseus* mapped to Chromosome 3 show higher sequencing depth in the fusion region. (*C*) sequencing reads of *L. oculatus* mapped to Chromosome 3 of *A. spatula* showed that the sequencing depth of a 2 kb fragment in the fusion region was zero.



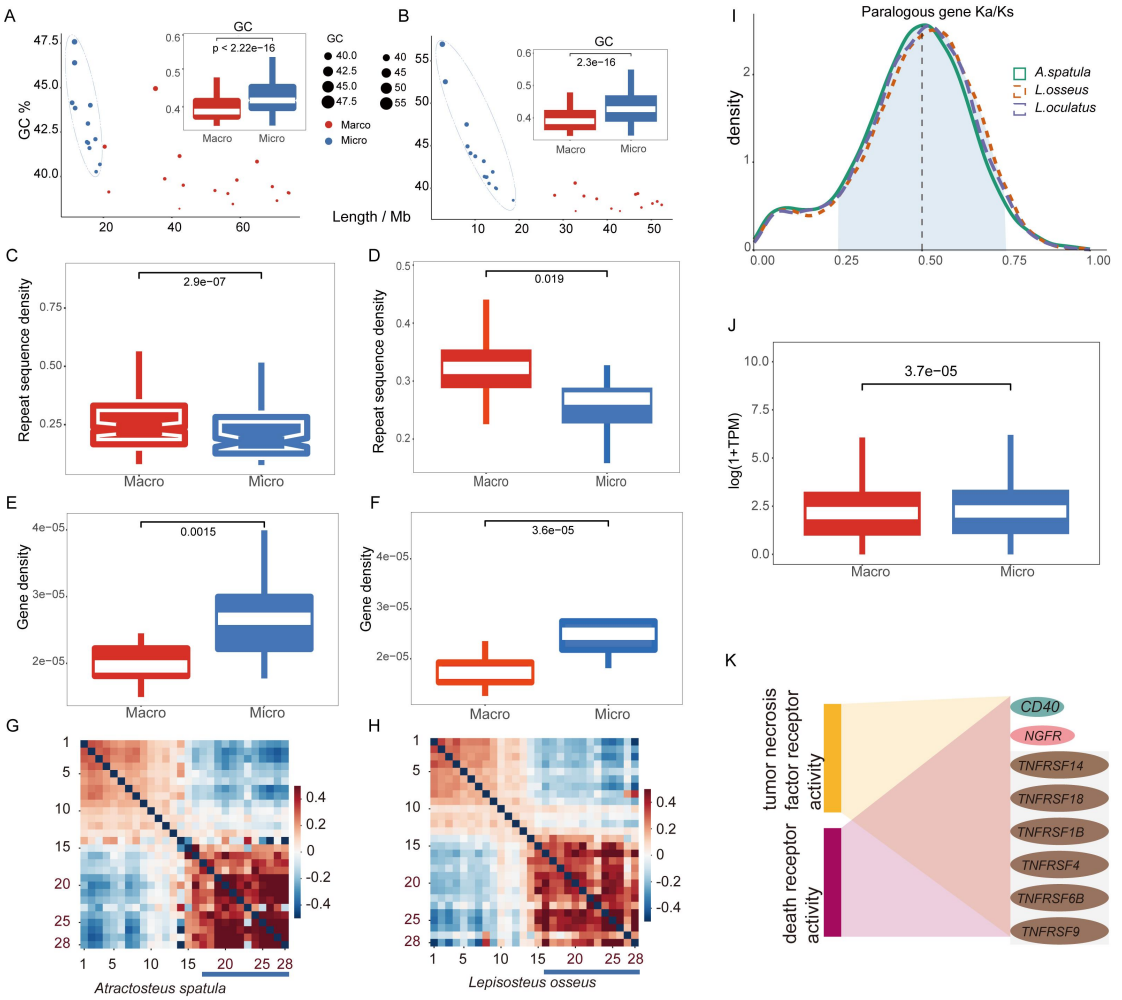
Supplemental Fig. S3:

Relationship between gene density, repeat content and chromosome length (macrochromsomes (>20Mb) and microchromosomes (<20 Mb)) in *Atractosteus spatula* (*A*), *Lepisosteus osseus* (*B*), and *L. oculatus* (*C*).



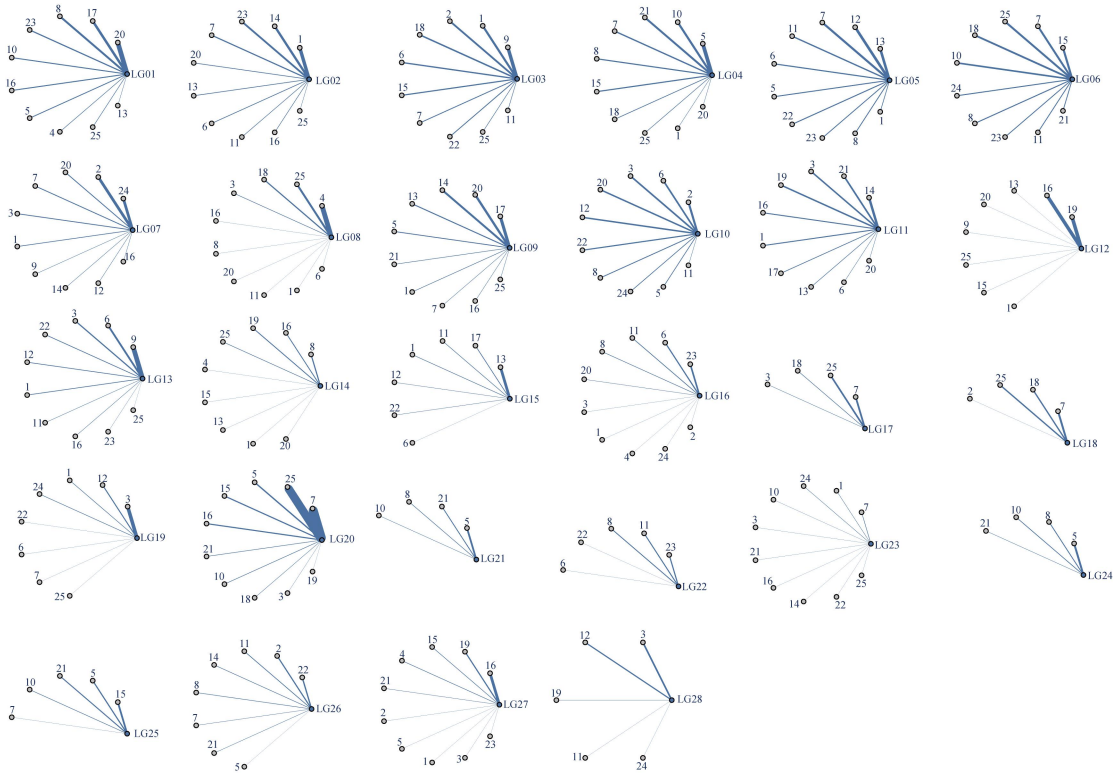
Supplemental Fig. S4:

Sequence Characterization of Chromosomes 14 and 23 in *A. spatula.* (*A, B*) Heatmap showing the repeat sequence content and gene density distribution; (*C*) Gene synteny map of Chromosome 14 between *A. spatula* and *L. osseus*; **(***D*) Gene synteny map of Chromosome 23 among *A. spatula*, *L. oculatus*, and *L. osseus*; (*E*) Comparison of repeat sequence divergence across the chromosomes of *A. spatula*.



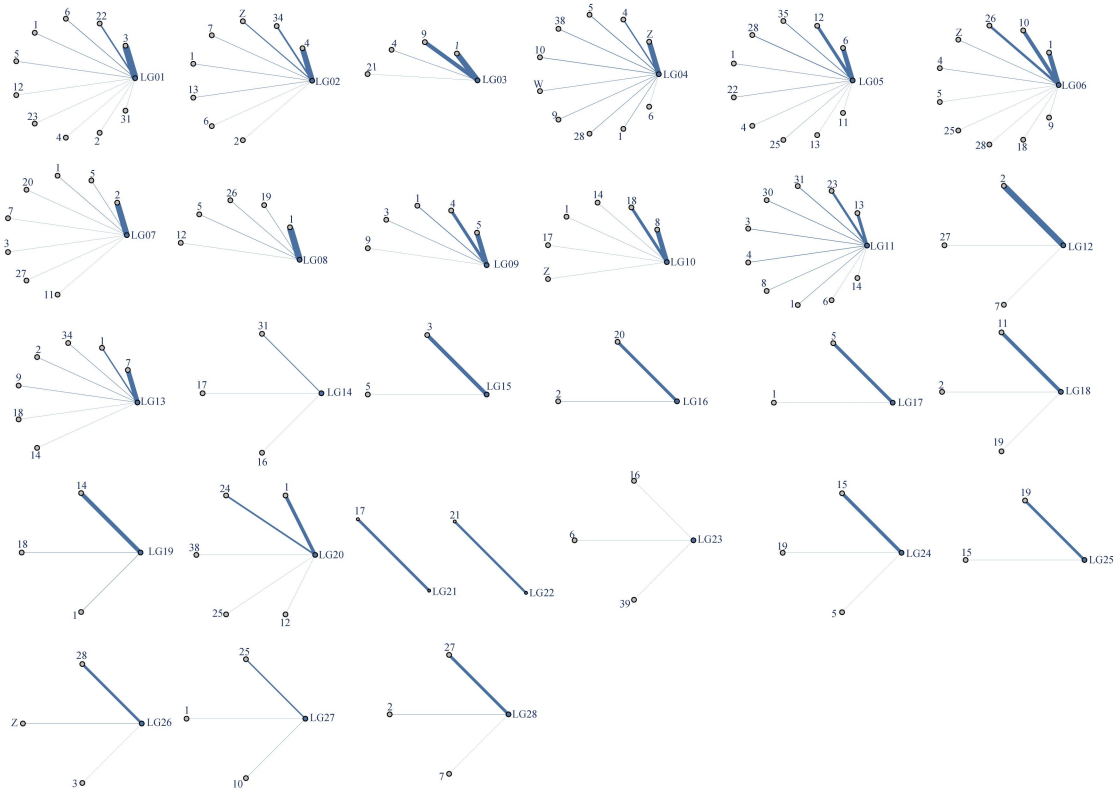
Supplemental Fig. S5:

The macro- and microchromosomes characteristics of gars. (*A-B***)** Comparison of global GC content between macrochromosomes and microchromosomes in *A. spatula* and *L. osseus*. **(***C-D***)** Comparison of repeat sequence density in *A. spatula* and *L. osseus*. **(***E-F***)** Comparison of gene density between micro- and macrochromosomes in *A. spatula* and *L. osseus*. (*G-H*) Comparison of pairwise Hi-C contact density between macrochromosomes and microchromosomes in *A. spatula* and *L. osseus*. (*I*) Comparison of the *K*a/*K*s ratios of paralogous genes between macrochromosomes and microchromosomes in three gar species. **(***J***)** The RNA-seq expression levels between macrochromosomes and microchromosomes. (*K*) GO functional enrichment annotation of genes located on microchromosomes.



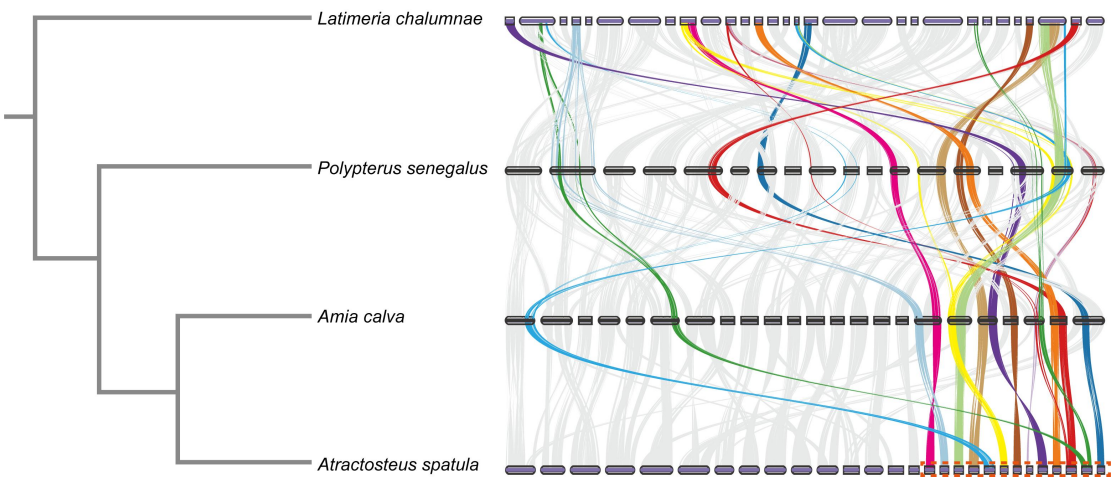
Supplemental Fig. S6:

Homologous chromosomes between *Atractosteus spatula* and *Danio rerio*. Network diagrams display the relative proportions of single-copy orthologous genes between *A. spatula* and *D. rerio*. The thickness of the connecting lines indicates the number of orthologous genes shared between chromosomes. Each chromosome of *A. spatula* shows homology with multiple chromosomes of *D. rerio*, likely due to teleost-specific genome duplication (TGD), which has resulted in extensive chromosomal rearrangements.



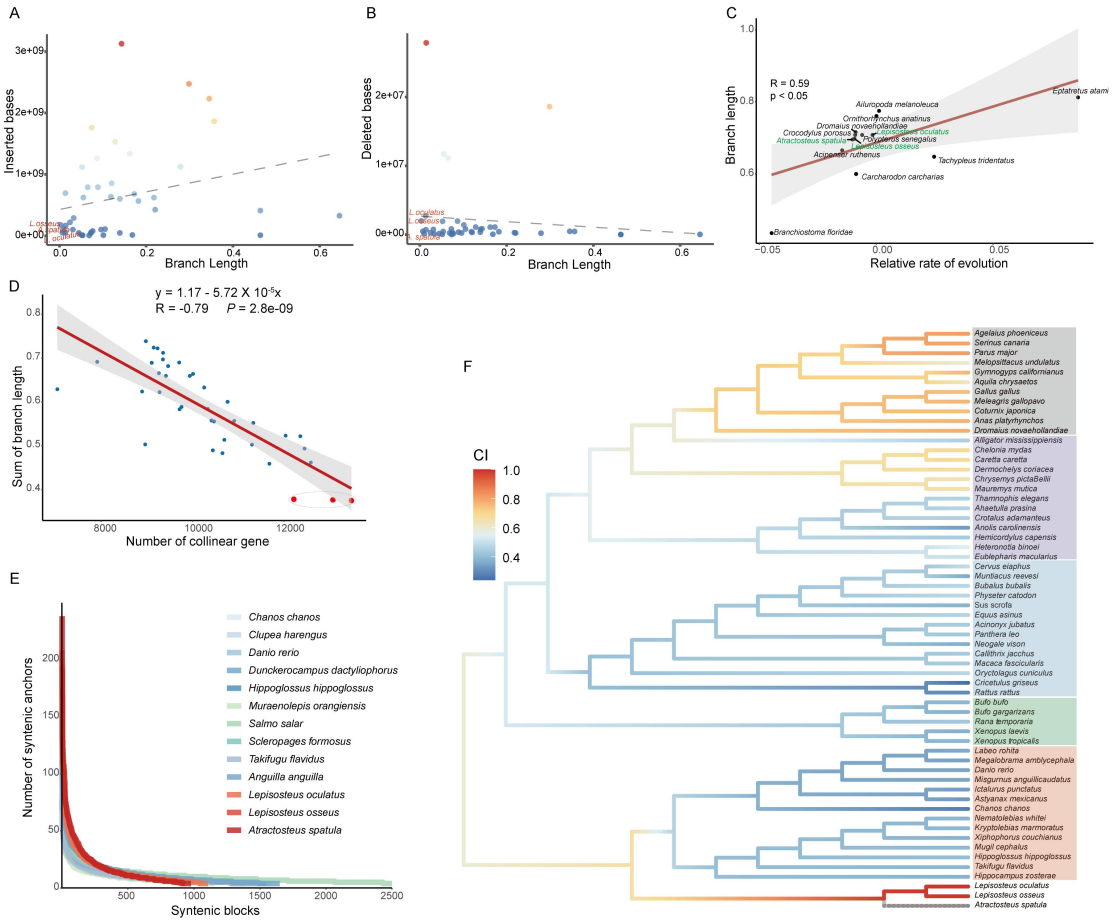
Supplemental Fig. S7:

Homologous chromosomes between *Atractosteus spatula* and *Gallus gallus*. Network diagrams display the relative proportions of single-copy orthologous genes between *A. spatula* and *G. gallus*. The thickness of the connecting lines indicates the number of orthologous genes shared between chromosomes. Compared to macrochromosomes, microchromosomes between *A. spatula* and *G. gallus* exhibit a 1:1 homologous relationship.



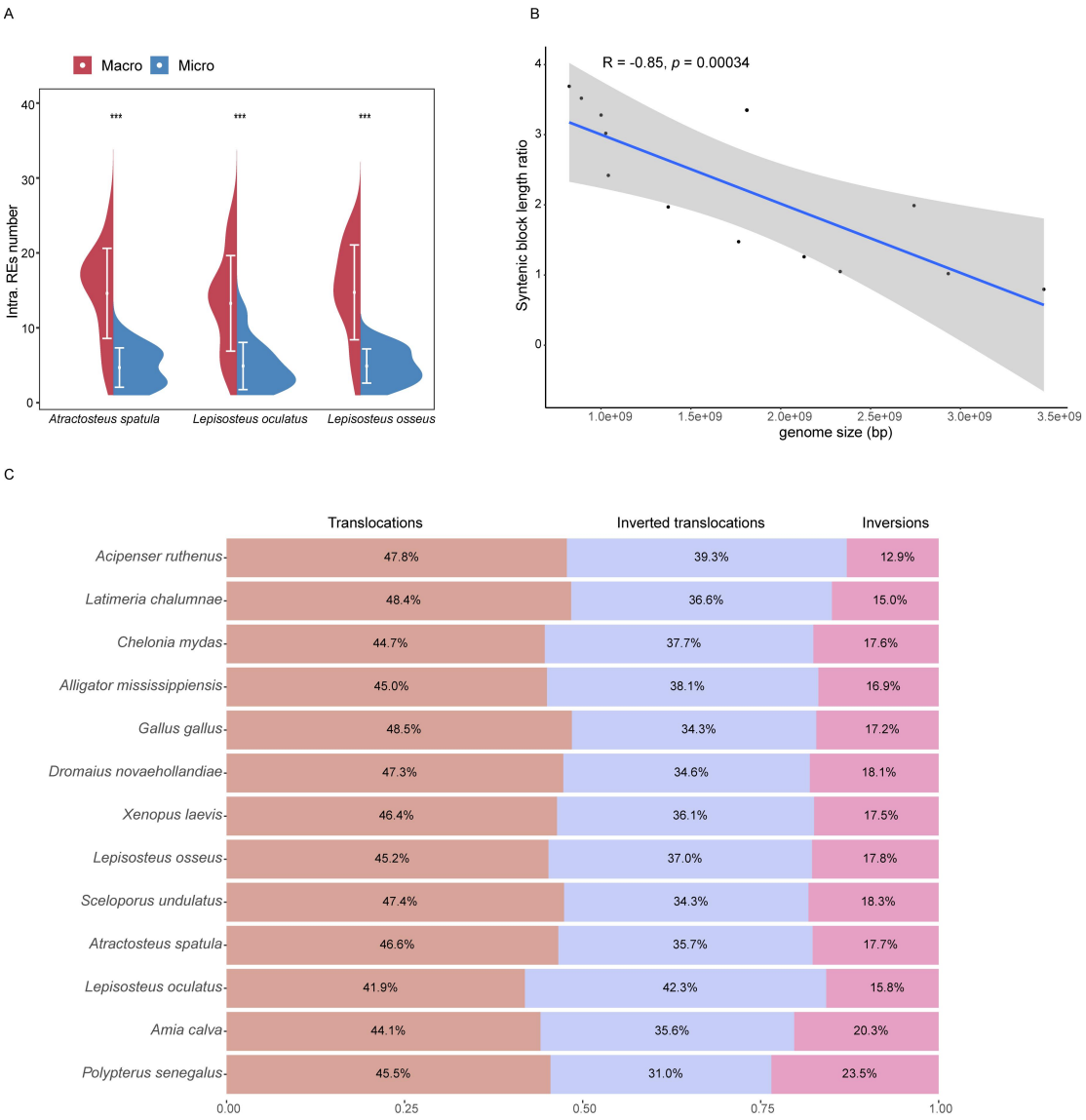
Supplemental Fig. S8:

Comparison of macrosynteny conservation across diverse phylogenetic lineages. The gene synteny plots indicate that *A. spatula* microchromosomes exhibit higher conservation than macrochromosomes throughout evolution. Colored highlights indicate microchromosomes.



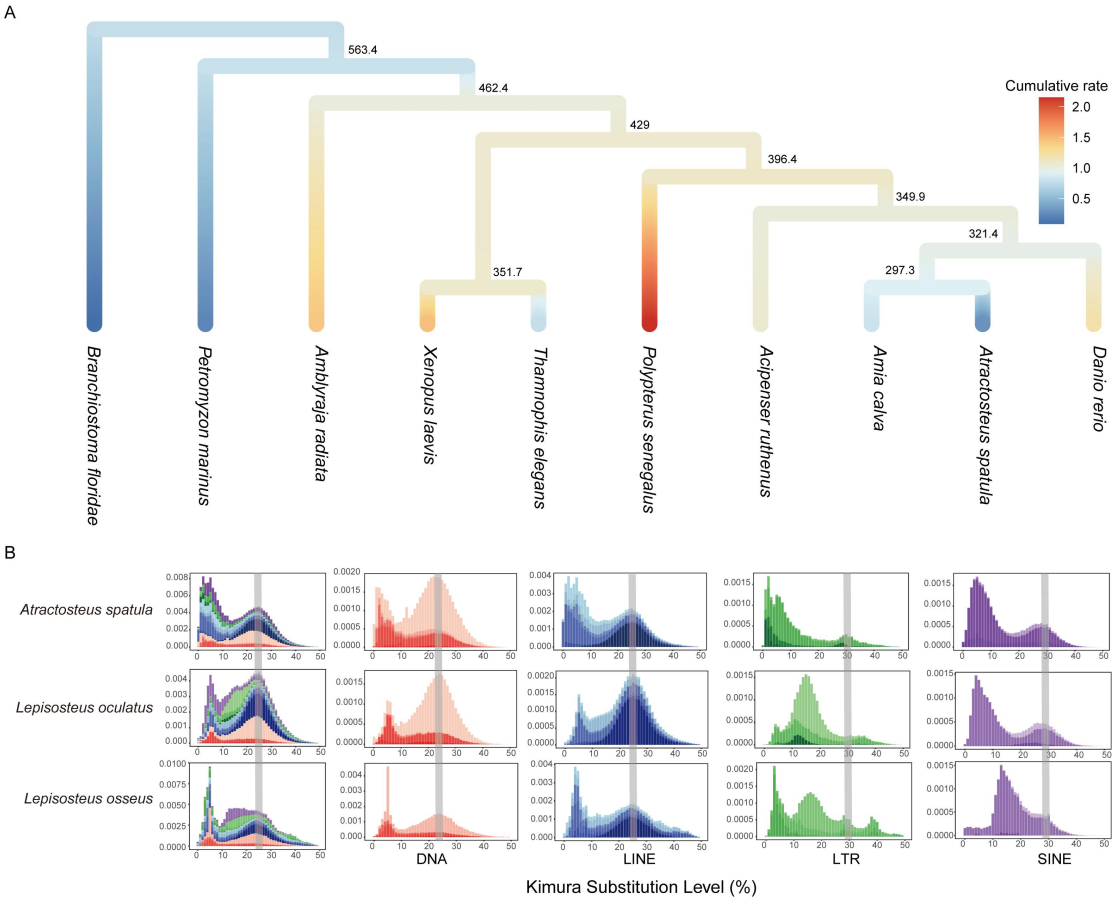
Supplemental Fig. S9:

Evolutionary rates of gar genomes compared with diverse lineages. (*A***)** Scatter plot showing the relative increase in base pairs and branch length across the whole genome alignment among gars and other animals (including a lancelet, cyclostomes, chondrichthyans, non-teleost actinopterygians, teleosts, birds, squamates, turtles and mammals) (Main figure 2A). **(***B***)** Scatter plot showing the relative decrease in base pairs and branch length across the whole genome alignment among selected lineages. **(***C***)** Correlation between relative evolutionary rate and branch length among some putative "living fossils.". (*D*) Correlation between the number of collinear genes and branch length, with three red dots representing the three gar species. **(***E***)** Comparison of syntenic block size decay among different ray-finned fishes in comparison with the reference species *Chiloscyllium plagiosum*. **(***F***)** The phylogenetic tree illustrates relationships among various vertebrate groups, and the macrosynteny conservation index (CI) highlights dynamic changes in genomic conservation across species, with *Atractosteus spatula* as the reference species.



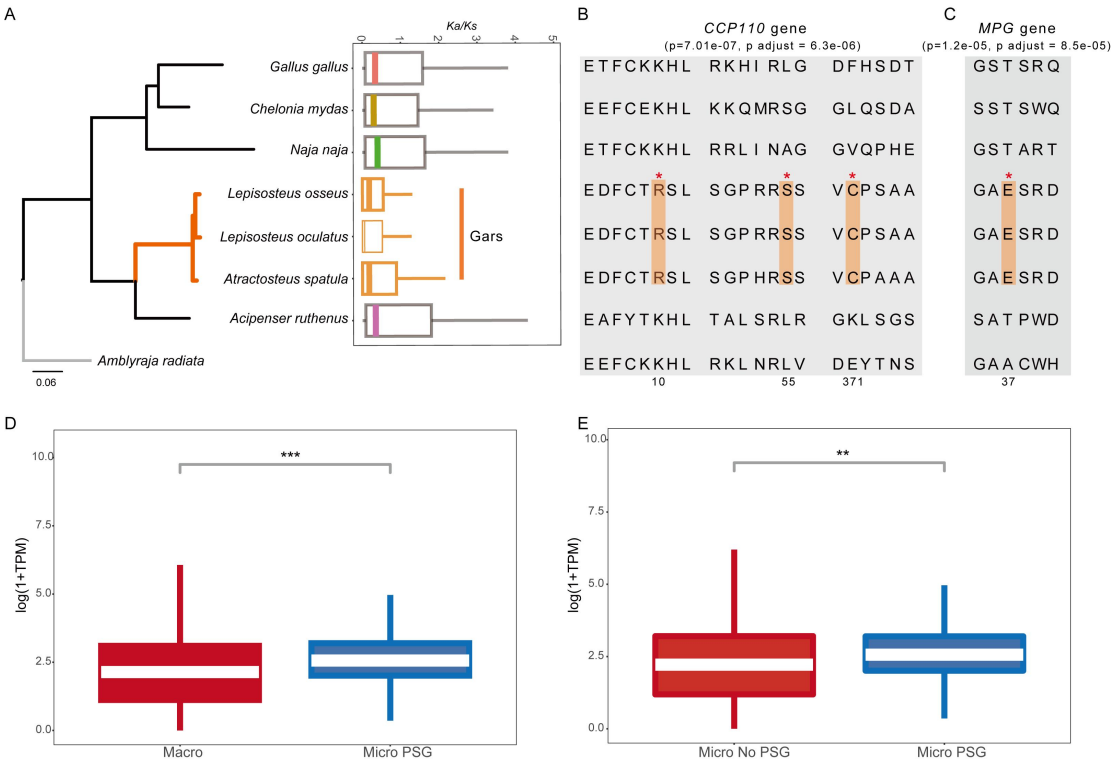
Supplemental Fig. S10:

Comparison of intrachromosomal rearrangements between macrochromosomes and microchromosomes in three gar species. (*A***)** The results indicate that the number of rearrangements in microchromosomes is lower than in macrochromosomes, suggesting that microchromosomes are more stable than macrochromosomes in gars. **(***B***)** the correlation between the overall syntenic block length ratio (*N. bancroftii* versus query species) and genome sizes. (*C*) Results statistics of intra-chromosome rearrangement types, including translocations, inverted translocations and inversions.



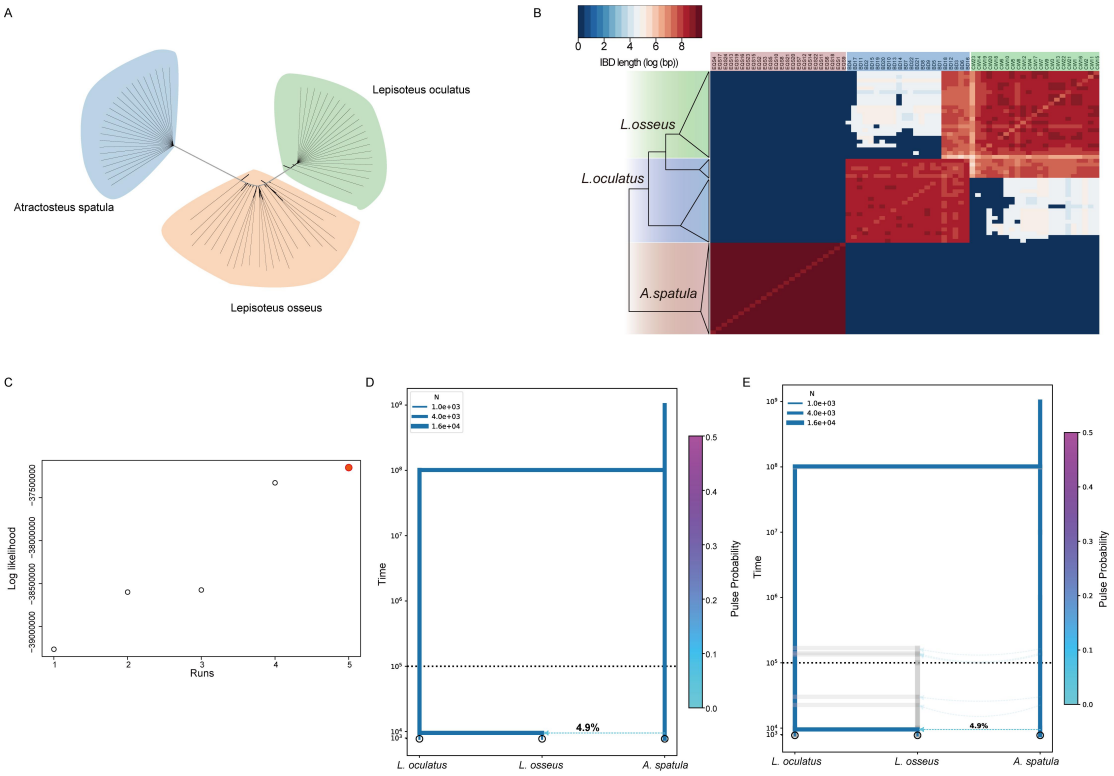
Supplemental Fig. S11:

Sequence divergence distribution of transposable elements (TEs) and insertion rate of young TEs. (*A*) Phylogenetic branches are color-coded to reflect the evolutionary rate of recently accumulated TEs with less than 10% K2P divergence from consensus TEs across genomes of diverse species. (*B*) The landscape plot is based on Kimura 2-parameter (K2P) sequence divergence, comparing individual TE copies with consensus sequences. The landscape distribution of TEs in three gar species, showing similar distribution patterns with two distinct peaks. The gray dashed line indicates the primary ancient peak in *Atractosteus spatula* for reference.



Supplemental Fig. S12:

Positive selection gene (PSG) analysis in gars. (*A*) Species phylogeny tree and distribution of *K*a/*K*s values. **(***B***,** *C***)** Partial alignment of positively selected genes, with asterisks indicating amino acids in gars with BEB posterior probabilities higher than 95%. **(***D***)** Comparison of expression levels between microchromosomes of PSG genes and macrochromosomes genes. **(***E***)** Comparison of expression levels between PSGs on microchromosomes and non-PSG genes on microchromosomes.



Supplemental Fig. S13:

Population structure and ancient introgression of living gar genera. (*A***)** Unrooted neighbor joining tree of population data from three gar species. (*B*) Heatmap indicating the total length of identity-by-descent (IBD) blocks for each pair of comparisons. **(***C***)** five rounds of simulations were performed, with the model achieving the highest log-likelihood value selected as the optimal model. **(***D***)** The genetic introgression event between *Lepisosteus osseus* and *Atractosteus spatula* using momi2. **(***E***)** Bootstrap confidence intervals were estimated by refitting the model on the bootstrap datasets, where blocks of the Site Frequency Spectrum (SFS) were resampled, and quantiles of the re-inferred parameters were examined.

Supplementary Information Tables

Table S1. Summary of the raw sequencing data for *Atractosteus spatula* and *Lepisosteus osseus* assembly.

A. spatula (1.02 Gb)

Data statistics	Illumina	Nanopore	Hi-C
Total (G)	65	239	137
Q20	97.97	14.39	99.18
Q30	94.54	0	96.69
GC	40.73	40.27	42.02
Coverage (X)	63.7	234.3	134.3

L.osseus (1.2 Gb)

Data statistics	Illumina	HiFi	Hi-C
Total (G)	44	24	77
Q20	96.31	98.64	99.4
Q30	88.67	96.98	96.58
GC	39.83	40.47	42.09
Coverage (X)	36.7	20	64.17

Table S2. *Atractosteus spatula* genome gene function annotation statistics.

Database	Annotated gene number	% Ratio
NR	20,411	96.94
eggNOG	20,124	95.58
Pfam	19,508	92.65
InterPro	18,416	87.47
SwissProt	18,130	86.11

Table S3. *Lepisosteus osseus* genome gene function annotation statistics.

Database	Annotated gene number	%Ratio
eggNOG	19,590	97.09
InterPro	17,009	84.29
Nr	18,561	91.99
Pfam	18,014	89.28

Table S4. Comparison of assembly statistics between our newly generated gar genomes (bold size) and previously published gar genomes.

	A.spatula	*A.spatula*	*L.oculatus*	***L.osseus***	*L.osseus*
Assembly sizes	1019.9 Mb	1055.2 Mb	945 Mb	1195.1 Mb	1014.9 Mb
GC content (%)	40.3	40.1	39.6	40.6	40.1
Number of sequences	454	81,747	1,926	231	22,745
Chromosome number	28	/	29	28	/
Contig N50	5.16 Mb	21.2 Kb	68.3 Kb	28.18 Mb	116.61 Kb
Scaffold N50	56 Mb	1.4 Mb	50 Mb	55 Mb	53 Mb
Scaffold L50	8	209	8	8	9
Scaffold N90	15.54 Mb	132 Kb	11.6 Mb	4.868 Mb	5.560 Mb
Scaffold L90	22	1,019	24	32	26
Longest sequence	74 Mb	11.8 Mb	73.2 Mb	83 Mb	74.2 Mb
Number of gaps	1,434	787,078	955,774	185	27,358
Number of telomeres (pairs)	7 (0)	/	/	28 (6)	/
TE size (%)	31.24	25.075	18.63	37.39	28.11

Table S5. Based on whole-genome alignment, nucleotide similarity levels were assessed using *Atractosteus spatula* genome as the reference.

Genome	Nucleotide identity (%)	Total number of identical sites	Total number of aligned sites
Atractosteus spatula	100	1,007,672,558	1,007,672,558
Ancestor of *Lepisosteus*	83.35	839,854,398	861,105,715
Lepisosteus osseus	81.96	825,925,787	850,110,412
Lepisosteus oculatus	77.15	777,436,861	799,824,958

Table S6. Types of gars repeat sequence frequency statistics. Bold text denotes the newly assembled genomes.

Species	**% DNA**	**% LINE**	**% SINE**	**%LTR**	**% Other**	**% Total**
Alligator Gar	4.50	7.79	4.80	2.06	12.09	31.24
Spotted Gar	3.49	4.66	2.90	2.31	6.87	20.23
Alligator Gar	3.95	6.50	3.55	6.64	4.44	25.08
Longnose Gar	4.62	6.81	4.28	3.84	17.84	37.39
Longnose Gar	3.72	6.53	2.06	5.42	10.38	28.11

Table S7. Conservation index (CI) comparison between *A. spatula* all chromosomes and microchromosomes and those of diverse species.

Species	All chromosome CI	Sign	Micro CI
Sus scrofa	0.38	<	0.59
Canis lupus	0.46	<	0.59
Homo sapiens	0.47	<	0.57
Mus musculus	0.31	<	0.59
Monodelphis domestica	0.30	<	0.52
Naja naja	0.43	<	0.67
Gallus gallus	0.75	<	0.91
Chelonia mydas	0.68	<	0.92
Xenopus laevis	0.37	<	0.54
Oryzias latipes	0.40	<	0.63
Hippocampus zosterae	0.38	<	0.61
Danio rerio	0.35	<	0.53
Scleropages formosus	0.38	<	0.63
Amia calva	0.64	<	0.69
Acipenser ruthenus	0.63	<	0.85
Polypterus senegalus	0.53	<	0.67
Amblyraja radiata	0.83	<	0.92
Petromyzon marinus	0.39	<	0.72
Branchiostoma floridae	0.29	<	0.61
Latimeria chalumnae	0.70	<	0.82

Table S8. Relative evolutionary rates of species analyzed by LINTRE.

Note: The *Latimeria chalumnae* was used as the reference species, and *Nautilus pompilius* was used as the outgroup species. Z-statistic was used to test whether the distances between ingroups (bA, bB) and outgroup were significantly different. Delta is the absolute difference between bA and bB (delta = | bA - bB |). Z-statistics (Z score) is calculated by the formula of Z = delta/s.e., where delta represents the absolute difference between bA and bB, and s.e represents the standard error; CP (confident probability) is equal to 1 - P-value (CP = 1 - P-value).

Outgroup	Ingroup1	Ingroup2	bA	bB	delta	Z score	CP
Nautilus pompilius	***Tachypleus tridentatus***	***Latimeria chalumnae***	0.385198	0.341957	0.043241	30.739841	99.96%
Nautilus pompilius	***Atractosteus spatula***	***Latimeria chalumnae***	0.15128	0.142931	0.008349	10.254581	99.96%
Nautilus pompilius	***Lepisosteus Oculatus***	***Latimeria chalumnae***	0.159376	0.142492	0.016884	20.113438	99.96%
Nautilus pompilius	***Lepisosteus Osseus***	***Latimeria chalumnae***	0.151214	0.142048	0.009166	11.197308	99.96%
Nautilus pompilius	***Carcharodon carcharias***	***Latimeria chalumnae***	0.155404	0.145538	0.009866	12.004181	99.96%
Nautilus pompilius	***Branchiostoma floridae***	***Latimeria chalumnae***	0.291968	0.318239	0.026271	22.009923	99.96%
Nautilus pompilius	***Polypterus senegalus***	***Latimeria chalumnae***	0.158807	0.146367	0.01244	14.948249	99.96%
Nautilus pompilius	***Acipenser ruthenus***	***Latimeria chalumnae***	0.165809	0.161905	0.003903	2.811337	99.50%
Nautilus pompilius	***Crocodylus porosus***	***Latimeria chalumnae***	0.14285	0.133382	0.009468	11.875559	99.96%
Nautilus pompilius	***Eptatretus atami***	***Latimeria chalumnae***	0.341447	0.23666	0.104787	81.966224	99.96%
Nautilus pompilius	***Ornithorhynchus anatinus***	***Latimeria chalumnae***	0.161011	0.14246	0.018551	22.307175	99.96%
Nautilus pompilius	***Dromaius_novaehollandiae***	***Latimeria chalumnae***	0.145742	0.136056	0.009686	12.151619	99.96%
Nautilus pompilius	***Ailuropoda melanoleuca***	***Latimeria chalumnae***	0.1649	0.145208	0.019692	23.266565	99.96%

Table S9 Relative evolutionary rates of species analyzed using Tajima's D test.

Note: The significant analysis was calculated by Chi-Square test with the default parameters in this study. *Latimeria chalumnae* was used as the reference species, and *Nautilus pompilius* was used as the outgroup species. The "identical site" represents the number of identical sites observed among groups (outgroup, A species, and B species). The "ingroup A specific" represents the number of specific sites observed in group A, and "ingroup B specific" represents the number of specific sites observed in group B. Chi-score represents the statistic values calculated from Chi-Square tests.

Outgroup	Ingroup A	Ingroup B	Identical site	Ingroup A specific	Ingroup B specific	Chi-score	P-value
Nautilus pompilius	*Tachypleus tridentatus*	*Latimeria chalumnae*	205,061	70,023	58,979	945.49	<0.00001
Nautilus pompilius	*Atractosteus spatula*	*Latimeria chalumnae*	285,885	30,724	28,234	105.16	<0.00001
Nautilus pompilius	*Lepisosteus Oculatus*	*Latimeria chalumnae*	279,491	32,664	27,721	404.62	<0.00001
Nautilus pompilius	*Lepisosteus Osseus*	*Latimeria chalumnae*	281,876	30,310	27,615	125.39	<0.00001
Nautilus pompilius	*Carcharodon carcharias*	*Latimeria chalumnae*	286,521	31,795	28,839	144.11	<0.00001
Nautilus pompilius	*Branchiostoma floridae*	*Latimeria chalumnae*	247,832	57,376	65,079	484.56	<0.00001
Nautilus pompilius	*Polypterus senegalus*	*Latimeria chalumnae*	282,460	32,106	28,428	223.47	<0.00001
Nautilus pompilius	*Acipenser ruthenus*	*Latimeria chalumnae*	111,594	13,499	13,041	7.9	<0.005
Nautilus pompilius	*Crocodylus porosus*	*Latimeria chalumnae*	283,667	28,930	26,143	141.04	<0.00001
Nautilus pompilius	*Eptatretus atami*	*Latimeria chalumnae*	212,333	62,688	36,783	6746.38	<0.00001
Nautilus pompilius	*Ornithorhynchus anatinus*	*Latimeria chalumnae*	283,645	33,231	27,723	497.72	<0.00001
Nautilus pompilius	*Dromaius_novaehollandiae*	*Latimeria chalumnae*	288,506	30,006	27,102	147.67	<0.00001
Nautilus pompilius	*Ailuropoda melanoleuca*	*Latimeria chalumnae*	280,176	33,863	28,072	541.47	<0.00001

Table S10. List of 91 positively selected genes (PSGs) located on gar microchromosomes.

The gene functions were primarily annotated using information from the GeneCards database (https://www.genecards.org). The gene names were confirmed from gene names database (https://www.genenames.org).

Gene	P value	P adjust	Gene function
HSPA5	0	0	protein folding and quality control in the endoplasmic reticulum lumen
SLC2A11	2.6e-05	0.0001	facilitative glucose transporter
PBX3	0.0067	0.0190	transcriptional activator
PIP5K1C	0.0009	0.0034	endocytosis and cell migration
SEZ6L	2.6e-05	0.0001	specialized endoplasmic reticulum functions in neurons
ATCAY	0.0004	0.0016	development of neural tissues
KCNN1	2.0e-05	0.0001	forms a voltage-independent potassium channel activated by intracellular calcium
DOK4	6.3e-05	0.0003	nervous system development
APLP2	0.000312426	0.0013	regulation of hemostasis
ARID3A	2.5e-08	3.2e-07	normal embryogenesis
CREB3L3	0.0002	0.0008	transcription factor
COMP	0	0	structural integrity of cartilage
FBN3	0	0	
ANKRD13A	8.3e-05	0.0004	enables ubiquitin-dependent protein binding activity
ADAM11	0.0118	0.0309	spatial learning and motor coordination
LRRC75B	0.00018	0.0005	suppresses myogenic differentiation
SLC25A25	3e-09	5.4e-08	transport proteins
SPNS2	7.75e-07	6.7e-06	lymphocyte trafficking
PTBP1	0.0187	0.047	pre-mRNA splicing and regulation of alternative splicing events
HNF1B	7.9e-05	0.0004	regulates development of the embryonic pancreas
WSCD2	2.9e-08	3. 6e-07	integral component of membrane
SLC25A33	1.8e-05	0.0001	mitochondrial carrier proteins
FZD9	6e-09	8.8e-08	neuromuscular junction
SIAH1	0	0	
TMEM120B	0.0033	0.0104	efficient adipogenesis
ADAP1	6.4e-05	0.0003	enables GTPase activator activity
PSME3	4.9e-05	0.0003	
TPRG1L	5.5e-05	0.0003	regulation of glutamatergic synaptic transmission
IFFO2	3.2e-08	3.8e-07	
AIF1L	5.4e-05	0.0003	promotes actin bundling
FAM57A	1.9e-05	0.0001	
SPPL2B	3.2e-05	0.0002	regulation of innate and adaptive immunity
FSD1	7.7e-06	5.5e-05	microtubule organization and stabilization

	0.0003	0.0011	embryonic development of the urogenital tract and the lung
WNT4			
TRAF2	3.47e-07	3.19e-06	regulation of cell survival and apoptosis
DSCAML1	7.8e-08	7.8e-07	neuronal self-avoidance
NMT1	1.9e-06	1.56e-05	
MYO1D	4.6e-08	4.9e-07	endosomal protein trafficking
ZFHX3	0	0	regulates myogenic and neuronal differentiation
GPT2	0	0	gluconeogenesis and amino acid metabolism
FBXL20	4e-08	4.6e-07	
ACAD10	6.8e-08	7.06e-07	
NLK	5e-09	8.05e-08	protein stabilization and transforming growth factor beta receptor signaling pathway
RPL3L	0.0025	0.008	regulates muscle function
TMC7	6e-09	8.78e-08	enable mechanosensitive ion channel activity
GP1BB	0	0	participates in the formation of platelet plugs
RPS15A	0.0023	0.0075	
GNPTG	0.0034	0.0104	
COQ7	0.0002	0.0008	
	0.0001	0.0005	interacts with its substrate, Kin17, which is involved in DNA repair and replication and mRNA processing
METTL22			
NOMO1	7.2e-06	5.4e-05	
	0.0013	0.0042	base excision repair (BER), the primary repair pathway for the repair of oxidative DNA damage
NTHL1			
JMJD8	0	0	regulates angiogenesis and cellular metabolism
	1.04e-06	8.89e-06	proteasome-mediated ubiquitin-dependent protein catabolic process
SPSB3			
KIAA0513	1.6e-05	0.0001	
KIAA0355	0	0	
MVD	2.34e-05	0.00014	
RNF166	6.4e-05	0.0003	
	0	0	SNARE-pin assembly and Golgi-to-ER retrograde transport
COG4			
SRL	5e-09	8.05e-08	
TEKT4	2e-09	3.8e-08	cilium assembly and cilium movement
	7.01e-07	6.3e-06	regulating cytokinesis and genome stability via cooperation with CALM1 and CETN2
CCP110			
ERCC4	0	0	nucleotide excision repair
POLR3K	0.0002	0.0007	
MPG	1.2e-05	8.5e-05	base-excision repair, DNA repair enzyme
TMEM130	8.3e-06	5.8e-05	
LMF1	0.0007	0.0027	
PEMT	0.0004	0.0016	
TIFA	4.3e-08	4.77e-07	adaptive and innate immunity

CWC25	4.48e-05	0.0002	
LRRC46	7.3e-06	5.5e-05	
MRPL45	1.09e-07	1.06e-06	
CDC6	0.0093	0.0248	involved in the initiation of DNA replication
PLXDC1	0.0026	0.0083	endothelial cell capillary morphogenesis
STARD3	0.0169	0.0435	
DHX8	0	0	ATP-dependent RNA helicase
DCAKD	1.89e-07	1.79e-06	
CCDC47	1.6e-08	2.1e-07	
PNPO	0.0006	0.0023	
RUNDC1	0.0002	0.00084	
MAP3K14	0	0	
CCDC97	0	0	
WDR59	1e-09	2.0e-08	involved in cellular response to amino acid starvation and positive regulation of TOR signaling
CCDC135	2.012e-06	1.62e-05	
LONP2	0.0047	0.0136	maintaining overall peroxisome homeostasis
FUK	0.0015	0.0049	
CYB5D2	0.0002	0.0008	nervous system development
TRIM50	0.0009	0.0034	
MORN3	0.0009	0.0034	located in nucleus
TMEM240	0.0048	0.0139	
DNLZ	0.0096	0.02537	protein folding; protein import into mitochondrial matrix; and protein stabilization.

Table S11. The F3 statistics for *Atractosteus spatula, Lepisosteus osseus* and *L. oculatus*. If F3(C; A, B) > 0, this suggests that population C is unadmixed; if F3(C; A, B) < 0 with a Z-scores < -3, this indicates that population C has undergone admixture between populations A and B.

The qp3Pop results:

A	B	C	F3	Standard error	Z-scores	SNPs
L. oculatus	*L. osseus*	*A. spatula*	1218.64	64.75	18.82	14,144,084
A. spatula	*L. oculatus*	*L. osseus*	-0.236	0.015	-14.889	14,144,084
A. spatula	*L. osseus*	*L. oculatus*	6.792	0.259	26.174	14,144,084

The threepop calculated F3 results in blocks of 500 SNPs (a total of 30,554 blocks):

A	B	C	F3	Standard error	Z-scores	SNPs
L. oculatus	*L. osseus*	*A. spatula*	0.851	0.0003	2654.79	15,277,036
A. spatula	*L. Oculatus*	*L. osseus*	-0.017	0.0002	-103.691	15,277,036
A. spatula	*L. osseus*	*L. oculatus*	0.114	0.0003	434.652	15,277,036