scientific data

Check for updates

OPEN

DATA DESCRIPTOR

First Chromosome-level Genome Assembly and Annotation of an Endangered Freshwater Stingray (*Fontitrygon garouaensis*) from Africa

Lotanna Micah Nneji[1] ✉, Segun Olayinka Oladipo[1,8], Oyetola Olusegun Oyebanji[2,8], Adeola Oluwakemi Ayoola[3], Olatunde Omotoso[4,5], Ifeanyi Christopher Nneji[6] & Kehinde Moruff Adelakun[7]

Fontitrygon garouaensis **(Smooth Freshwater Stingray) is an evolutionarily distinct and globally endangered species and is currently the only stingray known to be strictly adapted to African freshwater systems. Lack of a reference genome has limited studies of its adaptive evolution and genomic architecture. Here, we present the first chromosome-level genome assembly of *F. garouaensis* using a hybrid approach that integrates PacBio HiFi long reads, Illumina short reads, and Hi-C chromatin conformation capture. The assembly spans 4.19 Gb with 41 anchored chromosomes, a scaffold N50 of 86.49 Mb, and a contig N50 of 16.58 Mb. Overall, 84.07% of the genome was assigned to chromosomes, with 65.27% repetitive elements. BUSCO analysis showed 93.3% completeness, confirming a highly contiguous genome. We annotated 29,804 protein-coding genes, with 98.28% functionally annotated. Repetitive elements comprised 65.27% of the genome, including lineage-specific expansions of DNA transposons (7.13%) and LTRs (19.31%). This genomic resource establishes a foundation for systematic, evolutionary, and conservation research on this threatened species.**

Background & Summary

Elasmobranchs, including sharks, skates, sawfishes and rays, are among the most ancient and evolutionarily distinct vertebrates[1,2]. They diverged from the lineage leading to bony fishes approximately 420–450 million years ago, retaining many primitive traits such as cartilaginous skeletons while also exhibiting unique adaptive mechanisms[3]. Despite their evolutionary significance and ecological roles, there is a paucity of genomic resources for this group[4,5]. To date, whole-genome sequencing efforts have primarily focused on large marine elasmobranchs, greatly enhancing our understanding of their biology, evolution, and ecological adaptations[2,3,6]. Despite this recent surge in genomic resources, there are still limited genomic data for elasmobranchs, particularly those in African waterbodies. Thus, the availability of whole genome and transcriptome datasets for African elasmobranchs will fill critical gaps in understanding their evolutionary history, population structure, and adaptive mechanisms, as well as inform conservation and management strategies for these often-threatened taxa.

Among African elasmobranchs, *Fontitrygon garouaensis* (formerly *Dasyatis garouaensis*) stands out as the only stingray currently known to be fully adapted to freshwater, and it is endemic to West Africa[7–11]. A defining morphological feature that distinguishes *F. garouaensis* from its congeners (*F. margarita* and *F. margaritella*) is the

[1]Department of Biology, Howard University, Washington, DC, USA. [2]Botanical Research Institute of Texas, Fort Worth, Texas, USA. [3]Department of Human Genetics, University of California, Los Angeles, USA. [4]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [5]University of Chinese Academy of Sciences, Beijing, China. [6]Department of Marine Biology, Shantou University, Shantou, China. [7]Department of Wildlife and Ecotourism Management, Federal College of Wildlife Management, New Bussa, Nigeria. [8]These authors contributed equally: Segun Olayinka Oladipo, Oyetola Olusegun Oyebanji. ✉e-mail: Lotannamicah.nneji@howard.edu

	Illumina	PacBio HiFi	Hi-C
Insert size	350 bp	10-20 kb	200-350 bp
Total raw bases (bp)	93,751,433,700	—	—
GC Content (%)	42.02	—	—
Total data (Gb)	93.75	140.4	182
Read length (bp)	150 (PE)	—	—
Sequencing coverage (x)	21.37	32.00	112.2

Table 1. Statistics of hybrid sequencing data for *Fontitrygon garouaensis* genome assembly.

pronounced reduction or complete absence of dorsal denticles[7,9]. This species holds considerable significance due to its evolutionary distinctiveness and ecological specialization[9,10], particularly given that freshwater colonization is a rare phenomenon among elasmobranchs[12]. Nevertheless, there are fewer studies that investigate the mechanisms enabling freshwater colonization and adaptation in elasmobranchs[12–15]. Within this framework, *F. garouaensis* serves as a valuable model for exploring evolutionary processes in elasmobranchs, providing insights into habitat colonization, physiological adaptation, and factors shaping diversification within a predominantly marine lineage.

To date, no chromosome-level genome assembly or annotation are available for *F. garouaensis*, although partial mitochondrial genomes have been reported[9,16]. Previous studies have demonstrated the broad potential of reference genomes in advancing organismal biology[17,18]. Therefore, generating a high-quality, chromosome-level reference genome for *F. garouaensis* would not only enable comparative genomics with other elasmobranchs but also facilitate the identification of molecular mechanisms underlying freshwater adaptation and specialization, while providing essential genomic tools for effective conservation management. Considering its threatened status, evolutionary significance, and documented population declines[9], a complete genome assembly of *F. garouaensis* is urgently needed to advance evolutionary, ecological, and conservation genomic research.

In the present study, we employed a hybrid sequencing approach that integrates PacBio HiFi long reads, Illumina short reads, and Hi-C chromatin conformation capture to generate a high-quality, chromosome-level genome assembly of *F. garouaensis*. This genomic resource not only provides critical insights into the genomics of *F. garouaensis* but also establishes a foundation for future investigations of the molecular mechanisms driving adaptive evolution in elasmobranchs across salinity gradients. Furthermore, this study lays an important foundation for genomic research on other endangered African elasmobranch species.

Methods

Sample acquisition and DNA extraction. On November 15, 2023, an adult female *F. garouaensis* specimen with a disc length of approximately 308 mm, which is within the average length (748.9 mm) reported by Oladipo *et al.*[9], was collected from local fishermen on the River Niger at Jebba, Kwara State, Nigeria (N 9.149356, E 4.865930). The specimen was initially identified based on morphological characteristics following the identification guide by Olaosebikan and Raji[19]. Muscle tissue was subsequently sampled for whole-genome sequencing (Illumina short reads, PacBio HiFi long reads, and Hi-C). For the transcriptome sequencing, assembly, and annotation, five tissues (liver, eye, gill, body tissue, and spleen) were sampled. All tissue samples were preserved in RNAlater™ Stabilization Solution (Thermo Fisher Scientific, USA) and stored at −80 °C in Dr Nneji's Laboratory, Department of Biology, Howard University, USA. Samples were later shipped to Novogene Inc. for hybrid genome sequencing and transcriptomic analyses.

Genomic DNA extraction, library preparation and sequencing. High-quality, high-molecular-weight genomic DNA (gDNA) was extracted from muscle tissue using a combination of cetyltrimethylammonium bromide (CTAB) and sodium dodecyl sulfate (SDS) protocols, following Healey *et al.*[20]. The quality of the extracted gDNA was quantified using a Qubit 4.0 fluorometer (Invitrogen, USA), and DNA purity was assessed with 1% agarose gel electrophoresis. The gDNA was randomly sheared into shorter fragments using a Covaris M220 system (Covaris, Woburn, MA), after which fragments were end-repaired and A-tailed, followed by ligation with Illumina adapters. Adapter-ligated products were size-selected, PCR-amplified, and purified. Each library was quantified using a Qubit 4.0 fluorometer and qPCR, and the fragment size distribution was analyzed with a Fragment Analyzer (Agilent, CA, USA) to assess DNA integrity. Thereafter, pooling was done based on quantified library concentrations and the required sequencing amount.

A hybrid sequencing strategy was employed to construct the genome survey and assembly using three sequencing libraries. First, for Illumina short-read sequencing, quantified libraries were pooled and sequenced on the Illumina HiSeq 2000 platform (San Diego, CA, USA) with an average insert size of 350 bp and a read length of 2 × 150 bp paired-end (PE150) reads (Table 1). Second, the PacBio HiFi long-read library was prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, USA) according to the manufacturer's protocol and sequenced on the PacBio Sequel II platform (Pacific Biosciences, USA) across two independent runs. Contigs were *de novo* assembled using Hifiasm with PacBio HiFi reads. Third, the high-throughput chromosome conformation capture (Hi-C) library was constructed using the Frasergen Hi-C Kit (Frasergen, China) according to the manufacturer's protocol. For Hi-C, muscle tissue was processed by cell cross-linking, DNA digestion and fragmentation with the restriction enzyme DpnII, biotin labeling of repaired sticky ends, ligation, enrichment of biotin-labeled fragments, and purification. The genome was fragmented at 6-bp recognition sites (approximately every 4,000 bp). Ligation was performed at low DNA concentration in permeabilized nuclei using T4 DNA ligase, and interacting loci were subsequently quantified by PCR amplification of ligated junctions.

RNA extraction and transcriptome sequencing. Total RNA was extracted from the tissues (liver, eye, gill, body tissue, and spleen) using the TRIzol reagent kit (Thermo Fisher Scientific Inc., USA). Messenger RNA (mRNA) was purified from total RNA using poly-T oligo-attached magnetic beads (Illumina, San Diego, CA, USA). Following fragmentation, first-strand cDNA was synthesized using random hexamer primers with fragmented mRNA as the template, and second-strand cDNA was synthesized incorporating deoxyuridine triphosphate (dUTP). The resulting double-stranded cDNA was subjected to end repair, A-tailing, adapter ligation, size selection, USER enzyme digestion, amplification, and purification with AMPure XP beads (Beckman Coulter Inc., USA). Libraries were initially quantified with a Qubit 4.0 fluorometer, and size distributions as well as the absence of adapter dimers were confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Final library concentrations were determined by quantitative real-time PCR (qPCR) following the manufacturer's protocols. After quality control, libraries were pooled and sequenced on the Illumina HiSeq platform (Illumina Inc., USA). Among the five tissues sampled, only the brain and gill tissues yielded high-integrity RNA with sufficient concentration, passed quality profiling, and were used in our study.

***De novo* genome assembly and Hi-C scaffolding.** Read preprocessing was performed using Trimmomatic v0.39[21] to remove adapter sequences and low-quality reads ($N > 10\%$; Q-score ≤ 5). Read quality was further assessed with FastQC[22] prior to alignment against the Atlantic stingray (*Hypanus sabinus*, GCA_030144855.1)[23] reference genome using BWA-MEM (http://bio-bwa.sourceforge.net/) with default parameters. An initial 17-mer frequency analysis was conducted with Jellyfish v2.3.0[24], and the resulting frequency table was analyzed with the SOAPdenovo package[25]. SAMTools (Danecek *et al.*[26]; available from http://samtools.sourceforge.net/) was used to convert SAM to BAM format, sort and deduplicate reads, and subsequently call single nucleotide polymorphisms (SNPs). Assembly accuracy was assessed using three complementary approaches: (1) mapping rates and sequencing coverage; (2) a 10-kb sliding window analysis to calculate GC content and average sequencing depth, thereby detecting potential GC bias or contamination (a uniform GC distribution indicating clean sequencing); and (3) genome completeness assessment with BUSCO (Benchmarking Universal Single-Copy Orthologs; Simão *et al.*[27]; http://busco.ezlab.org/) in combination with *tblastn*, *Augustus*, and *HMMER*.

The genome size of *F. garouaensis* was estimated through k-mer analysis of Illumina clean reads, revealing a total genome size of 4,342.44 Mbp, a heterozygosity rate of 0.23%, and a repeat sequence content of 72.69% (Supplementary Table 1a). The quality and coverage of the Illumina reads were further evaluated using 17-mer analysis (Fig. 1A). A major peak was observed at approximately 32 × sequencing depth, indicating sufficient and relatively uniform coverage. A minor peak at low depths (1–5 ×) likely represents low-coverage regions, while the extended tail at higher depths suggests the presence of genomic regions with unusually high coverage. The estimated genome size of *F. garouaensis* is consistent with patterns observed in other elasmobranchs. Specifically, the Smooth Freshwater Stingray genome (4.19 Gb, this study) is larger than that of the whale shark (*Rhincodon typus*, 3.44 Gb[28]) and the white-blotched river stingray (*P. leopoldi*, 4.11 Gb[6]), but smaller than that of the white shark (*Carcharodon carcharias*, 4.63 Gb[29]).

The genome assembly integrated multiple sequencing strategies, including Illumina whole-genome shotgun (WGS; 93.75 Gb) for genome survey, PacBio Revio sequencing (2 SMRT cells; 140.4 Gb) for de novo assembly, Hi-C sequencing (182 Gb) for chromosome scaffolding (Table 1), and mRNA-seq (6 Gb) for gene annotation (Fig. 1). The initial contig-level assembly was generated from PacBio HiFi reads using Hifiasm v0.19.5[30,31], resulting in a ~4.19 Gb draft genome comprising 2,688 contigs. This assembly exhibited high contiguity, with a contig N50 of 16.58 Mb and a maximum contig length of 128.46 Mb (Table 2). Illumina short reads were subsequently used for base-level polishing and downstream assembly evaluation. To achieve chromosome-level resolution, draft contigs were scaffolded using Hi-C data with the ALLHiC pipeline[32].

The final assembly retained the estimated genome size (~4.19 Gb) and exhibited high contiguity, with a scaffold N50 of 86.49 Mb (Table 2). A total of 1,379 scaffolds were generated, of which 41 large scaffolds were anchored to putative chromosomes, collectively representing 84.07% of the genome (Supplementary Table 1b). This anchoring proportion likely reflects the absence of species-specific chromosomal data and structural divergence from the closest chromosome-level reference, *Hypanus sabinus* (GCA_030144855.1)[23], which limits synteny-based placement. The remaining 1,338 unplaced scaffolds were retained in the final assembly to ensure completeness. Hi-C paired-end reads were subsequently mapped to construct a 3D chromatin interaction model for genome validation and organization (Fig. 1B). Chromosome-scale sequences ranged from 42.3 Mb to 157.8 Mb, consistent with the 41-chromosome karyotype reported for the first time in this species, though closely related taxa, *Hypanus* sabinus (GCA_030144855.1)[23], containing 35 chromosomes with the chromosome-level genome, have been reported (Fig. 2; Supplementary Table 2). This high-quality, chromosome-scale genome assembly of *F. garouaensis* provides a foundational reference for advancing evolutionary, ecological, and functional genomic research in African cartilaginous fishes and across the broader elasmobranch lineage.

Phylogenetic analyses. To confirm the identity of our species within the genus *Fontitrygon*, we extracted the standard DNA barcode marker (COI gene sequence) from the newly assembled *F. garouaensis* genome using Geneious Prime v2025.1.1[33]. We queried it against the National Center for Biotechnology Information (NCBI) database using BLASTn to verify species identity. For the whole mitochondrial genome phylogenetic analysis, 12 publicly available whole mitochondrial genome sequences from closely related species within the subfamily Urogymninae (family Dasyatidae) were downloaded, with *Neotrygon kuhlii* designated as the outgroup. Sequences were aligned with MAFFT v7.450[34] in Geneious Prime. A maximum-likelihood (ML) phylogenetic tree was constructed using IQ-TREE2[35,36] with the "-m MFP" option to automatically select the best-fit substitution model[37]. Node support was assessed with 1,000 ultrafast bootstrap replicates using the "-bb 1000" option[38]. The resulting consensus tree was visualized in FigTree v1.4.3[39]. A BLAST search of the COI sequence from our

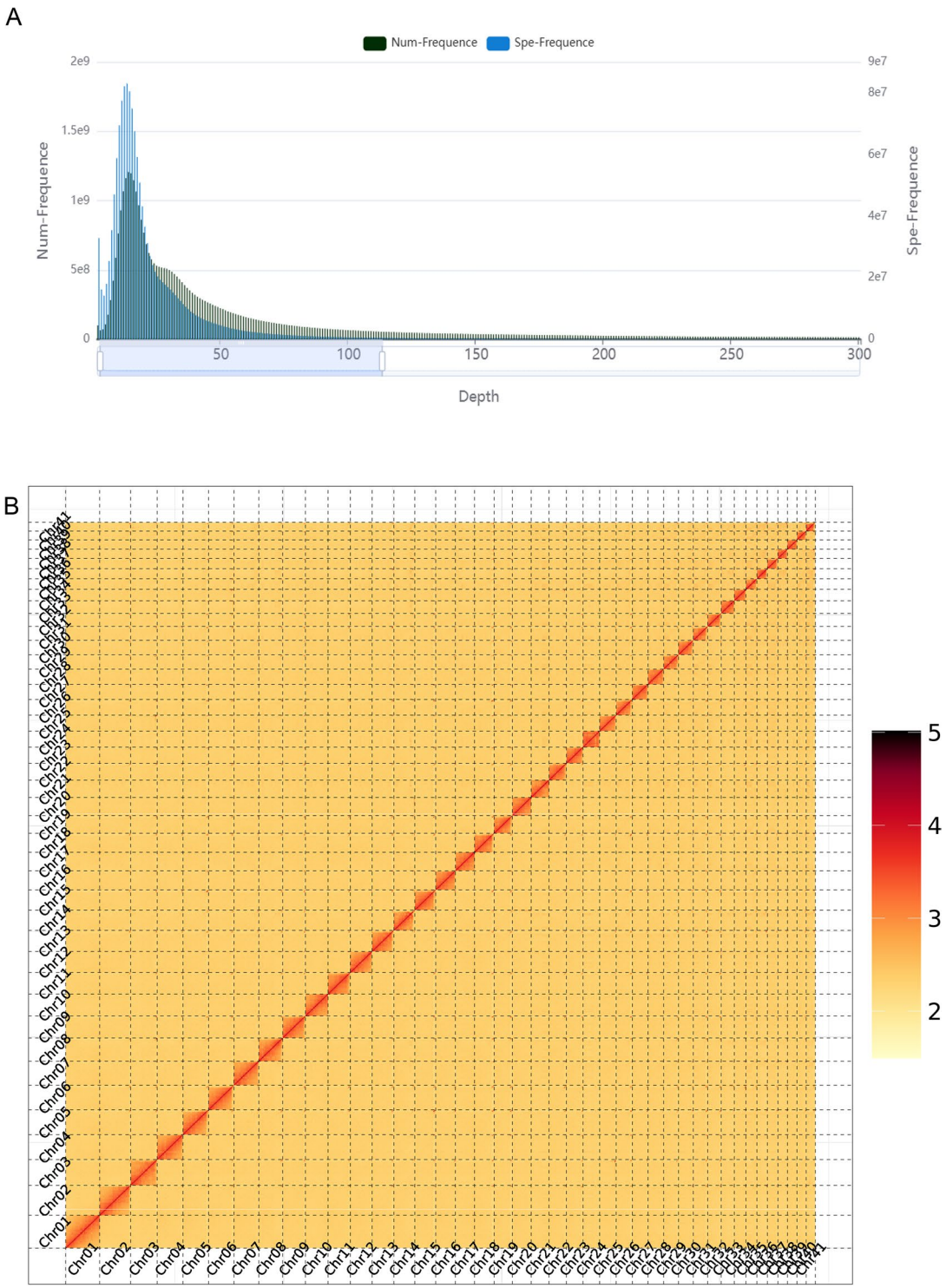

Fig. 1 Genome survey and Hi-C interaction map of *Fontitrygon garouaensis*. (**A**) K-mer frequency distribution of cleaned PacBio HiFi reads using 17-mers. *Spe-Frequency* indicates specific frequency, and *Ne-Frequency* indicates number of frequency. (**B**) Hi-C chromosomal contact heatmap generated using the 3D-DNA pipeline.

specimen revealed 100% identity with *F. garouaensis* (GenBank accession: PP865962), a specimen previously collected from Jebba, Kwara State, Nigeria[16]. In the ML phylogenetic tree, our specimen clustered within a monophyletic group containing other *F. garouaensis* sequences and formed a sister clade to all other *Fontitrygon* species within the subfamily Urogymninae, with strong nodal support (100% bootstrap confidence; Supplementary Fig. 1). These results provide robust molecular phylogenetic confirmation of our sample as *F. garouaensis* within Urogymninae (family Dasyatidae), consistent with previous studies[9,16].

Assembly Statistics	Contig length	Scaffold length	Contig number	GC content (%)	Number of chromosomes
Total	4,185,027,496	4,185,158,396	2,688	41.98	41
Max	128,460,713	157,764,182	—		
Number ≥ 2000	—	—	2,687		
N50	16,576,105	86,485,802	58		
N60	12,679,680	77,506,048	87		
N70	8,644,714	69,283,552	127		
N80	4,237,787	52,219,673	195		
N90	1,243,743	11,232,548	387		

Table 2. Summary statistics of *Fontitrygon garouaensis* genome assembly.

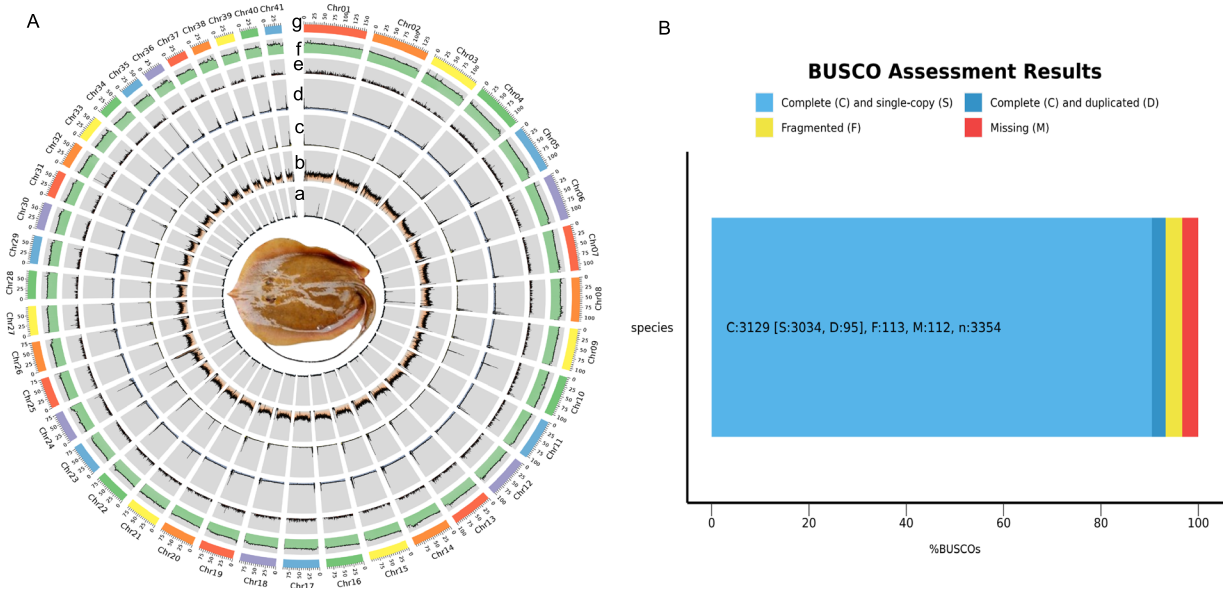

Fig. 2 Features of *Fontitrygon garouaensis* genome. (**A**) Circos plot showing genome features from inside to outside: (a) DNA-TE content, (b) LINE content, (c) LTR content, (d) repeat content, (e) gene density, (f) GC content, and (g) the 41 pseudo-chromosomes, and (**B**) BUSCO analysis of genome assembly completeness.

Genome annotation and gene prediction. Structural annotation of the *F. garouaensis* genome was conducted using an integrative approach that combined ab initio prediction, homology-based prediction, and RNA-Seq-assisted prediction. Ab initio gene models were predicted using Augustus v3.5 (http://bioinf.uni-greifswald.de/augustus/) and SNAP (http://homepage.mac.com/iankorf/) with default parameters. For homology-based prediction, protein sequences from closely related species [*Potamotrygon leopoldi* (PRJCA003510)[40], *Hypanus* sabinus (GCF_030144855.1)[23], *Oreochromis niloticus* (GCF_001858045.2)[41], *Gasterosteus aculeatus* (GCF_016920845.1)[42] and *Danio rerio* (GCF_049306965.1)[43]] were obtained from the NGDC and NCBI databases. These sequences were aligned to the *F. garouaensis* genome using TBLASTN v2.2.26 (E-value ≤ 1e−5), and the resulting hits were refined with GeneWise v2.4.1 to produce accurate spliced alignments and gene models within protein-matching regions.

For transcript-based gene prediction, paired-end (PE) RNA-Seq reads from brain and gill tissues were processed with fastp to remove adapter sequences, poly-N regions, and low-quality reads. Clean reads were de novo assembled using Trinity v2.8.5[44], and quality metrics (Q20, Q30, and GC content) were calculated to ensure high sequencing quality. All downstream analyses were conducted using these high-quality reads. The reference genome (*H. sabinus*) was indexed with HISAT2 v2.2.1[45], and cleaned PE reads were aligned to the genome. Resulting alignments were used as input for StringTie v2.2.1[46], which performed genome-guided transcript assembly under default parameters.

Prior to gene prediction, repetitive elements were identified using RepeatMasker (http://www.repeatmasker.org/) with the RepBase repeat sequence database (http://www.girinst.org/repbase/). Predictions from all three approaches, along with transcriptome alignments, were integrated using EVidenceModeler (EVM v1.1.1 (Haas *et al.*[47]; Available from: http://evidencemodeler.sourceforge.net/)) to produce a non-redundant reference gene set. This process incorporated terminal exon support from the Program to Assemble Spliced Alignments (PASA; Haas *et al.*[47]) and masked transposable elements during gene prediction. To ensure robust transcript support during evidence integration and improve the reliability of the final gene models, we removed low-abundance or potentially spurious transcripts and retained only transcripts with reads per kilobase of exon model per million

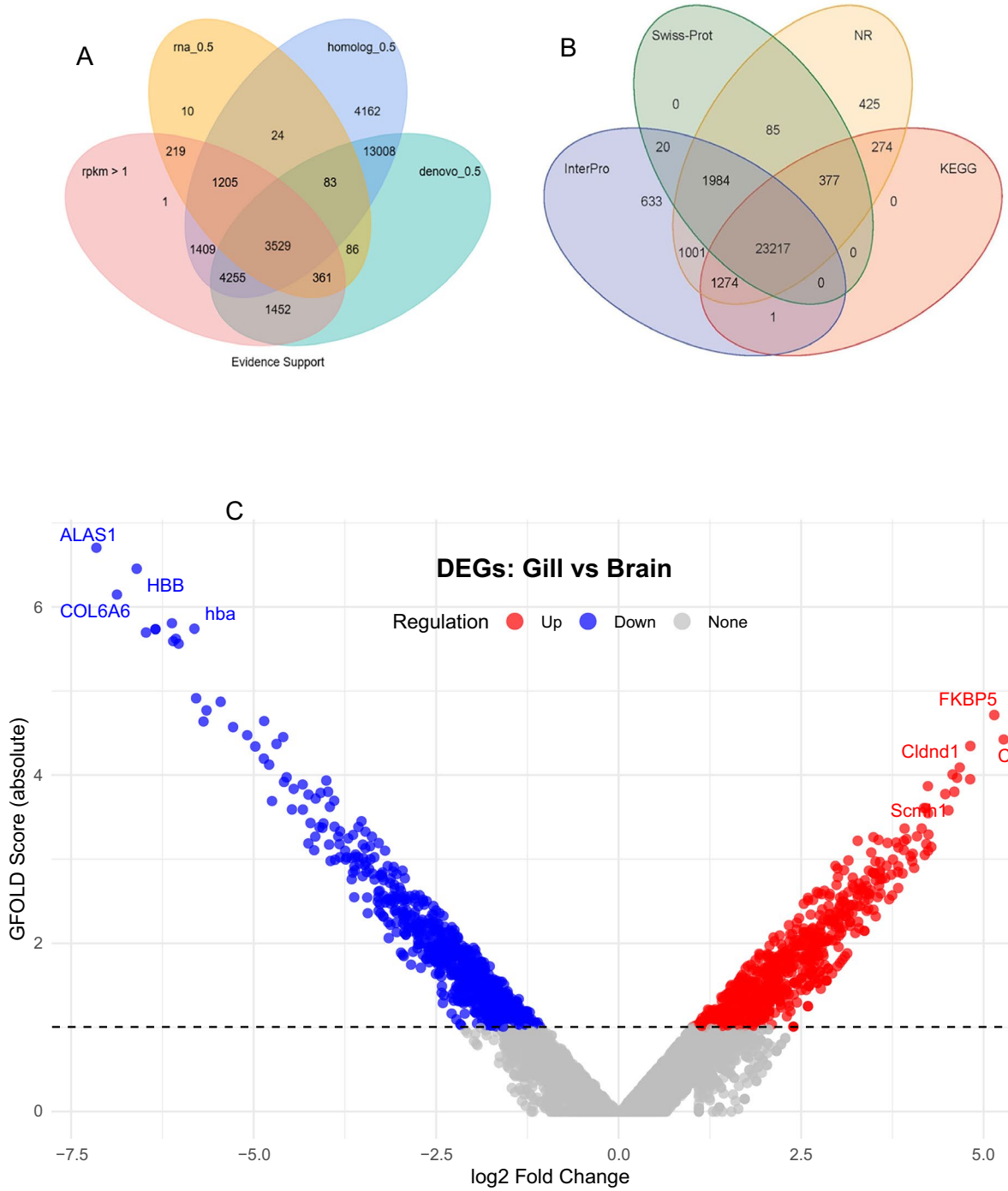

Fig. 3 Genome annotations and differential expression analysis. (**A**) Venn diagram showing gene set evidence support. *De novo* indicates genes supported by ab initio prediction at EVM integration; *Homolog* indicates genes supported by homology-based prediction at EVM integration; *RNA* indicates genes supported by RNA-Seq at EVM integration. Each evidence category is based on gene overlap >50%, and numbers indicate the total number of genes supported. (**B**) Shared and unique gene annotations across functional databases. (**C**) Volcano plot of differentially expressed genes in gill tissue relative to brain tissue of *F. garouaensis*.

mapped reads (RPKM) > 1[48]. Finally, transcriptome data were used to refine annotations by adding untranslated regions (UTRs) and alternative splicing information, resulting in a high-confidence final gene set.

A total of 29,804 protein-coding genes were predicted in the *F. garouaensis* genome (Supplementary Table 3). Of these, 29,291 genes (~98.28%) were successfully annotated across multiple databases, with 23,217 genes (77.90%) receiving comprehensive support from all four databases (Fig. 3A,B). In addition, we identified 872

Type	Database	Gene Number	Gene Percent (%)
Total		29,804	100
Annotated		29,291	98.28
	NR	28,637	96.08
	Swissprot	25,683	86.17
	KEGG	25,143	84.36
	InterPro	28,130	94.38
	Pfam	22,143	74.3
	GO	21,965	73.7
Unannotated		513	1.72

Table 3. Functional annotation statistics of *Fontitrygon garounensis*.

miRNAs, 7,380 rRNAs, 16,380 tRNAs, and 870 snRNAs (Supplementary Table 4). Detailed statistical information is presented in Fig. 3, Table 3, and Supplementary Tables 3–4.

Functional annotation of predicted genes was conducted by aligning protein sequences to the Swiss-Prot database using BLASTp (E-value $\leq$ 1e−5). Protein motifs and domains were identified with InterProScan v5.39[49] across multiple databases, including ProDom, PRINTS, Pfam, SMART, PANTHER, and PROSITE. Gene Ontology (GO) terms were assigned based on corresponding InterPro entries. Protein function was further inferred by transferring annotations from homologous sequences in Swiss-Prot[50] and the non-redundant protein (NR) database using DIAMOND v0.8.22[51] and BLASTp with an E-value cutoff of 1e−5. Additionally, KEGG pathway mapping was performed to assign genes to relevant biological pathways.

Transcriptomics. Differentially expressed genes (DEGs) between gill and brain tissues of *F. garouaensis* were identified using the Generalized Fold Change software GFOLD v1.1.4[52]. Genes were considered significant when GFOLD(0.01) > 1.045. To investigate the functional significance of the identified DEGs, Gene Set Enrichment Analysis (GSEA) was performed using the easyGSEA module within the easy Visualization and Inference Toolbox for Transcriptome Analysis (eVITTA) platform[30,31]. The GSEA pre-ranked method was applied to genes with GFOLD(0.01) > 1 to identify enriched pathways. Upregulated and downregulated genes were analyzed separately to capture direction-specific functional enrichment. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were also conducted to identify significantly overrepresented biological processes and molecular functions. A significance threshold of adjusted $p < 0.05$ was applied, with multiple testing corrections performed using the Benjamini–Hochberg method to control the false discovery rate (FDR). Enrichment results were visualized within eVITTA, and Manhattan plots were generated to illustrate the distribution of enriched pathways across functional categories. Differential expression analysis revealed distinct tissue-specific gene expression patterns between the brain and gills of *F. garouaensis*, reflecting their specialized physiological roles (Fig. 3). Gene Ontology (GO) enrichment analysis revealed that genes upregulated in the gill were predominantly associated with metabolic processes, particularly fatty acid metabolism and gluconeogenesis (Fig. 4**;** Supplementary Fig. 2).

Data Records

All sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1302590[53]. The DNA sequencing data from the PacBio HiFi library are available under the SRA accession number SRR35772892[54], the Illumina short-read genomic sequencing data used for assembly polishing under SRR35557154[55], the RNA-seq data used for genome annotation under SRR35772888 & SRR35772889[56,57]. The chromosome-level genome assembly of *F. garouaensis* has been deposited in NCBI GenBank under the Whole Genome Shotgun (WGS) accession number JBSTRW000000000[58].

Technical Validation

Gene annotation and repetitive elements. Our final de novo assembly produced a total contig length of 4.18 Gb with a contig N50 of 16.58 Mb, yielding a highly contiguous chromosome-scale assembly with 99.90% genome coverage across 41 chromosomes and strong concordance with the assembled genome. The GC content was 41.98%, and the proportion of ambiguous bases (N) was 0.00%, well below the commonly accepted threshold of < 10%, further supporting the high accuracy of the assembly. Genome completeness, assessed using BUSCO with the vertebrata_odb10 database (3,354 conserved orthologs), revealed 93.3% completeness, comprising 90.5% single-copy and 2.8% duplicated BUSCOs, with 3.4% fragmented and 3.3% missing genes, thereby confirming the integrity of the genome assembly (Fig. 2).

Further validation by Illumina short-read alignment demonstrated uniform sequencing depth, with most genomic regions exceeding 10 × and 20 × coverage thresholds, indicating even sequencing depth and robust assembly reliability (Supplementary Table 5a). Variant calling identified 1,128,320 SNPs, of which 99.5% were heterozygous. The homozygous SNP rate was extremely low (0.000137%; Supplementary Table 5b), further supporting the high base-level accuracy of the assembly. GC content analysis showed a balanced composition (~42%), with no evidence of abnormal base distribution or sequencing bias, suggesting the absence of contamination or technical artifacts.

Collectively, these results confirm the integrity, completeness, and high fidelity of the *F. garouaensis* genome assembly (Table 2**;** Figs. 1–2 and 5 Supplementary Tables 1, 2 and 5). The assembly revealed that 65.27% of

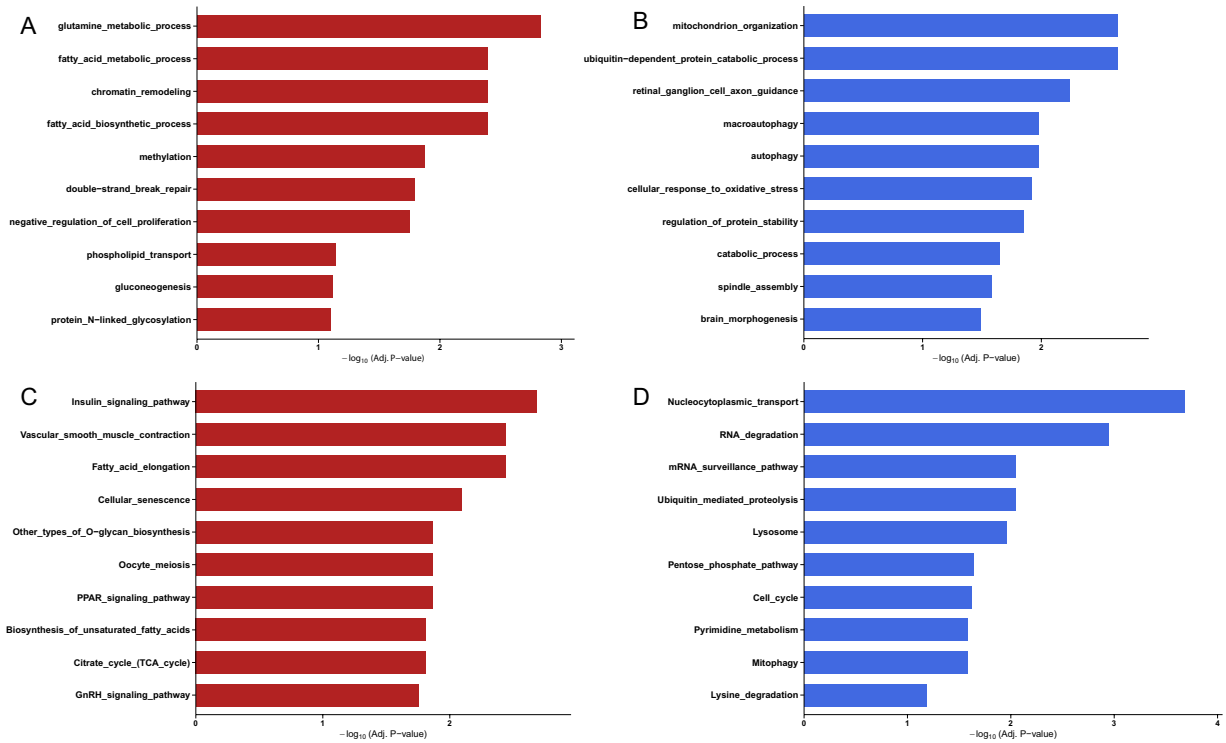

Fig. 4 Gene Ontology (GO) and KEGG pathway enrichment of differentially expressed genes in the gill compared to the brain of *Fontitrygon garouaensis*. (**A**) GO biological process (BP) terms significantly enriched among genes upregulated in the gill relative to the brain. (**B**) GO BP terms significantly enriched among genes downregulated in the gill relative to the brain. (**C**) KEGG pathways significantly enriched among genes upregulated in the gill relative to the brain. (**D**) KEGG pathways significantly enriched among genes downregulated in the gill relative to the brain.

Type	Number	Length(bp)	Percentage (%)
SINE	55,888	9,349,036	0.22
LINE	4,060,165	1,531,154,672	36.59
LTR	1,958,803	808,020,844	19.31
DNA	1,778,819	298,567,069	7.13
Unknown	215,539	52,966,089	1.27
Total		2,731,608,694	65.27

Table 4. Repetitive sequences in the genome of *Fontitrygon garouaensis*. SINE (Short Interspersed Nuclear Elements): Repetitive elements typically shorter than 1,000 bp. LINE (Long Interspersed Nuclear Elements): Repetitive elements usually longer than 1,000 bp. LTR (Long Terminal Repeats): Retrotransposons characterized by the presence of long terminal repeats at both ends. DNA transposons: Mobile elements that transpose directly via a DNA intermediate without involving RNA reverse transcription. Unknown: Repetitive sequences that could not be classified by RepeatMasker. Total: Non-redundant results obtained after integrating all classifications and removing overlapping regions.

the genome is composed of repetitive sequences (Table 4). This high repeat content is consistent with patterns observed in other large elasmobranch genomes, such as the white shark, and highlights the challenge of assembling repeat-rich, large genomes typical of cartilaginous fishes[14,15]. Comparative studies have also emphasized the slow evolutionary rates, gene family expansions, and elevated LINE content characteristic of shark genomes[15].

To investigate the genomic synteny between *F. garouaensis* and related elasmobranch species (*H. sabinus, H. akajei, L. erinacea, C. plagiosum*), we conducted a comprehensive comparative genomic analysis. Initially, an all-versus-all protein sequence alignment was performed using BLASTP v2.11.0 + [59] with a stringent e-value cutoff of 1e-5. The chromosomal location data and chromosome length information were extracted for syntenic block identification. These data were analyzed with JCVI v1.1.22[60], facilitating the interpretation of conserved genomic architecture across the studied species.

We identified extensive chromosomal rearrangements in *F. garouaensis* at both the intragenomic level and in comparisons with other sampled elasmobranch species (Fig. 5). Microchromosome synteny revealed dense inter-chromosomal links (e.g., between chromosome 2 and chromosome 14), indicative of a history of large-scale

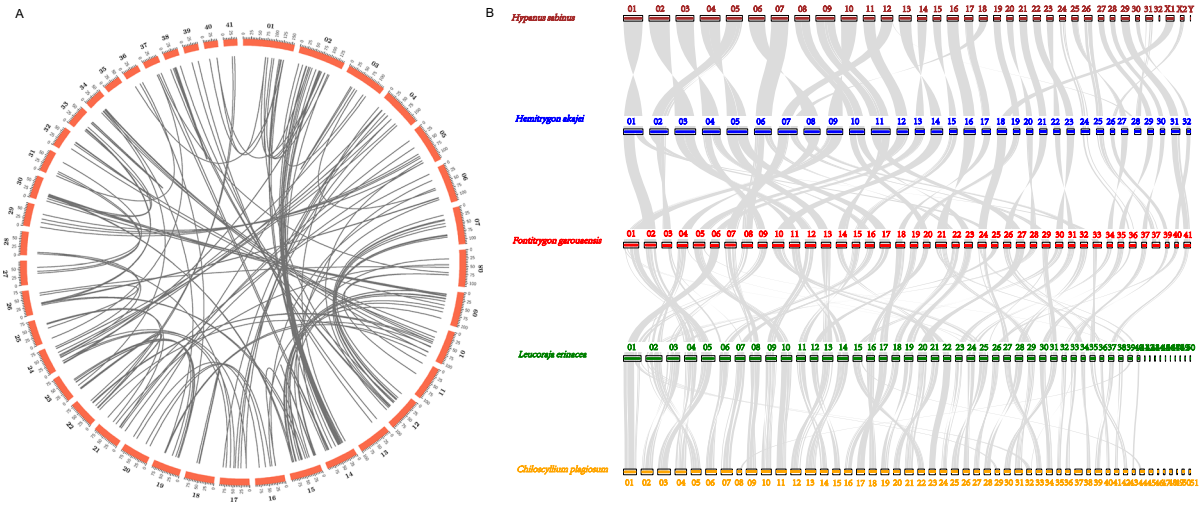

Fig. 5 Chromosome-level genomic structure in the genomes of *Fontitrygon garouaensis*. (**A**) intragenomic rearrangements and (**B**) Genomic synteny among the Elasmobranchs.

fissions, fusions, and translocations within the genome[61,62] (Fig. 5A). Inter-species synteny comparisons showed that while chromosomes are relatively conserved between *F. garouaensis* and closely related species, divergence from more distantly related lineages was characterized by pervasive chromosomal fusions and fissions (Fig. 5B). We also observed distinct chromosomal evolution within elasmobranchs, suggesting lineage-specific karyotype remodeling driven by recurrent fusion and fission events. These findings are consistent with evolutionary scenarios in which both chromosomal fragmentation and fusion contribute to genomic divergence[63].

The combination of extensive intragenomic structural variation and non-conserved synteny across elasmobranchs underscores the rapid and dynamic nature of karyotype evolution in this lineage. While inter-synteny is often conserved to maintain critical regulatory functions, disruptions that have been increasingly documented in other taxa, such as bryozoans and cephalopods, are thought to contribute to speciation and adaptation[64–66]. This genomic plasticity may therefore play an important role in shaping ecological adaptations and evolutionary trajectories across sharks and rays[67].

Declarations. The specimen used in this study was obtained from a freshly deceased stingray provided by a local fisherman. No animal experimentation was involved.

Ethics approval. All procedures in this study were conducted in accordance with Howard University's Institutional Animal Care and Use Committee (IACUC) guidelines (Protocol No. IACUC-2025-0003), and all aspects of the research complied with approved protocols, relevant institutional guidelines, and applicable national regulations.

Data availability

All sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1302590. The DNA sequencing data from the PacBio HiFi library are available under the SRA accession number SRR35772892, the Illumina short-read genomic sequencing data used for assembly polishing under SRR35557154, the RNA-seq data used for genome annotation under SRR35772888 & SRR35772889. The chromosome-level genome assembly of F. garouaensis has been deposited in NCBI GenBank under the Whole Genome Shotgun (WGS) accession number JBSTRW000000000.

Received: 7 October 2025; Accepted: 19 January 2026;
Published online: 26 January 2026

References

1. Rhie, A. *et al*. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746, https://doi.org/10.1038/s41586-021-03451-0 (2021).
2. Tan, M. *et al*. The whale shark genome reveals patterns of vertebrate gene family evolution. *eLife* **10**, Article e65394, https://doi.org/10.7554/eLife.65394 (2021).
3. Venkatesh, B. *et al*. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179, https://doi.org/10.1038/nature12826 (2014).
4. Carrier, J. C., Heithaus, M. R. & Simpfendorfer, C. A. (Eds.). *Shark Research*. 1st ed, https://doi.org/10.1201/b21842 (CRC Press, 2018).
5. Symonová, R. & Howell, W. M. Vertebrate genome evolution in the light of fish Cytogenomics and rDNAomics. *Genes* **9**, Article 96, https://doi.org/10.3390/genes9020096 (2018).
6. Zhou, J. *et al*. Draft genome of white-blotched river stingray provides novel clues for niche adaptation and skeleton formation. *Genomics Proteomics Bioinformatics* **21**, 501–514, https://doi.org/10.1016/j.gpb.2022.11.005 (2023).
7. Compagno, L. J. & Roberts, T. R. Marine and freshwater stingrays (Dasyatidae) of West Africa, with a description of a new species. *Proc., Calif. Acad. Sci.* **43**, 283–300 (1984).

8. Jabado, R. W., Keith Diagne, L., Sayer, C., Tamo, A. & Williams, A. B. Fontitrygon garouaensis. *IUCN Red List Threat Species*: e.T39406A104171509;2. https://doi.org/10.2305/IUCN.UK.2021 (2021).
9. Oladipo, S. O. *et al*. Morphology, DNA barcoding and range extension of a poorly known freshwater stingray Fontitrygon garouaensis Stauch & Blanc, 1962 from Nigerian inland Water. *Integr. Comp. Biol.* **65**, Article icaf125, https://doi.org/10.1093/icb/icaf125 (2025).
10. Oladipo, S. O. *et al*. Patterns of ichthyofaunal diversity and distribution across Jebba Hydro-Electric Power (HEP) Dam, Jebba, north-central Nigeria. *Braz. J. Biol.* **81**, 258–267, https://doi.org/10.1590/1519-6984.222952 (2021).
11. Thorson, T. B. & Watson, D. E. Reassignment of the African freshwater stingray, Potamotrygon garouaensis, to the genus Dasyatis, on physiologic and morphologic grounds. *Copeia* **1975**, 701–712, https://doi.org/10.2307/1443322 (1975).
12. Ballantyne, J. S. & Robinson, J. W. Freshwater elasmobranchs: a review of their physiology and biochemistry. *J. Comp. Physiol., B* **180**, 475–493, https://doi.org/10.1007/s00360-010-0447-0 (2010).
13. Hammerschlag, N. Osmoregulation in elasmobranchs: a review for fish biologists, behaviourists and ecologists. *Mar. Freshw. Behav. Physiol.* **39**, 209–228, https://doi.org/10.1080/10236240600815820 (2006).
14. Kuraku, S. Shark and ray genomics for disentangling their morphological diversity and vertebrate evolution. *Dev. Biol.* **477**, 262–272, https://doi.org/10.1016/j.ydbio.2021.06.001 (2021).
15. Marra, N. J. *et al*. Comparative transcriptomics of elasmobranchs and teleosts highlight important processes in adaptive immunity and regional endothermy. *BMC Genomics* **18**, Article 87, https://doi.org/10.1186/s12864-016-3411-x (2017).
16. Petean, F. F., Yang, L., Corrigan, S., Lima, S. M. Q. & Naylor, G. J. P. How many lineages are there of the stingrays genus *Hypanus* (Myliobatiformes: Dasyatidae) and why does it matter? *Neotrop. Ichthyol.* **22**, Article e20230046, https://doi.org/10.1590/1982-0224-2023-0046 (2024).
17. Totikov, A. *et al*. Chromosome-level genome assemblies expand capabilities of genomics for conservation biology. *Genes* **12**, Article 1336, https://doi.org/10.3390/genes12091336 (2021).
18. Yamaguchi, K. *et al*. Elasmobranch genome sequencing reveals evolutionary trends of vertebrate karyotype organization. *Genome Res.* **33**, 1527–1540, https://doi.org/10.1101/gr.276840.122 (2023).
19. Olaosebikan, B. D. & Raji, A. *Field Guide to Nigerian Freshwater Fishes* (Federal College of Freshwater Fisheries Technology, 2013).
20. Healey, A., Furtado, A., Cooper, T. & Henry, R. J. Protocol. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **10**, 21, https://doi.org/10.1186/1746-4811-10-21 (2014).
21. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, https://doi.org/10.1093/bioinformatics/btu170 (2014).
22. Andrews, S. FastQC: A quality control tool for high throughput sequence data, 2010, Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
23. NCBI Assembly. Hypanus sabinus genome assembly (GenBank GCF_030144855.1). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/assembly/GCF_030144855.1.
24. 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).
25. Luo, R. *et al*. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18, https://doi.org/10.1186/2047-217X-1-18 (2012).
26. Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, Article giab008, https://doi.org/10.1093/gigascience/giab008 (2021).
27. 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).
28. Read, T. D. *et al*. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: Rhincodon typus Smith 1828. *BMC Genomics* **18**, Article 532, https://doi.org/10.1186/s12864-017-3926-9 (2017).
29. Marra, N. J. *et al*. White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proc. Natl Acad. Sci. USA.* **116**, 4446–4455, https://doi.org/10.1073/pnas.1819778116 (2019).
30. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175, https://doi.org/10.1038/s41592-020-01056-5 (2021).
31. Cheng, X., Yan, J., Liu, Y., Wang, J. & Taubert, S. eVITTA: A web-based visualization and inference toolbox for transcriptome analysis. *Nucleic Acids Res.* **49**, W207–W215, https://doi.org/10.1093/nar/gkab366 (2021).
32. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845, https://doi.org/10.1038/s41477-019-0487-8 (2019).
33. Kearse, M. *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649, https://doi.org/10.1093/bioinformatics/bts199 (2012).
34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780, https://doi.org/10.1093/molbev/mst010 (2013).
35. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274, https://doi.org/10.1093/molbev/msu300 (2015).
36. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235, https://doi.org/10.1093/nar/gkw256 (2016).
37. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589, https://doi.org/10.1038/nmeth.4285 (2017).
38. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522, https://doi.org/10.1093/molbev/msx281 (2018).
39. Rambaut, A. FigTree v1.4.3, 2024, Available from: http://tree.bio.ed.ac.uk/software/figtree/.
40. NCBI BioProject. *Potamotrygon leopoldi* genome project (BioProject PRJCA003510). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJCA003510.
41. NCBI Assembly. *Oreochromis niloticus* genome assembly (GenBank GCF_001858045.2). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/assembly/GCF_001858045.2.
42. NCBI Assembly. *Gasterosteus aculeatus* genome assembly (GenBank GCF_016920845.1). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/assembly/GCF_016920845.1.
43. NCBI Assembly. *Danio rerio* genome assembly (GenBank GCF_049306965.1). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/assembly/GCF_049306965.1.
44. Haas, B. J. *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512, https://doi.org/10.1038/nprot.2013.084 (2013).
45. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915, https://doi.org/10.1038/s41587-019-0201-4 (2019).
46. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, Article 278, https://doi.org/10.1186/s13059-019-1910-1 (2019).
47. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7, https://doi.org/10.1186/gb-2008-9-1-r7 (2008).
48. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628, https://doi.org/10.1038/nmeth.1226 (2008).

49. Jones, P. *et al*. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240, https://doi.org/10.1093/bioinformatics/btu031 (2014).
50. UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515, https://doi.org/10.1093/nar/gky1049 (2019).
51. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60, https://doi.org/10.1038/nmeth.3176 (2015).
52. Feng, J. *et al*. GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* **28**, 2782–2788, https://doi.org/10.1093/bioinformatics/bts515 (2012).
53. *NCBI BioProject accession* https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1302590 (2026).
54. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR35772892 (2026).
55. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR35557154 (2026).
56. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR35772888 (2026).
57. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR35772889 (2026).
58. Nneji, L. M. *et al*. The first chromosome-level genome assembly and annotation of freshwater stingray (*Fontritrygon garouaensis*) from an African Inland Water. *GenBank* https://identifiers.org/ncbi/insdc:JBSTRW000000000 (2026).
59. Camacho, C. *et al*. Blast+: architecture and applications. *BMC Bioinform.* **10**, 421, https://doi.org/10.1186/1471-2105-10-421 (2009).
60. Tang, H. *et al*. Synteny and collinearity in plant genomes. *Science* **320**, 486–488, https://doi.org/10.1126/science.1153917 (2008).
61. Duran, C., Edwards, D. & Batley, J. Genetic maps and the use of synteny. *Methods Mol. Biol.* **513**, 41–55, https://doi.org/10.1007/978-1-59745-427-8_3 (2009).
62. Höök, L., Näsvall, K., Vila, R., Wiklund, C. & Backström, N. High-density linkage maps and chromosome level genome assemblies unveil direction and frequency of extensive structural rearrangements in wood white butterflies (Leptidea spp.). *Chromosome Res.* **31**, Article 2, https://doi.org/10.1007/s10577-023-09713-z (2023).
63. Kemkemer, C. *et al*. Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. *BMC Evol. Biol.* **9**, 84, https://doi.org/10.1186/1471-2148-9-84 (2009).
64. Hane, J. K. *et al*. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* **12**, R45, https://doi.org/10.1186/gb-2011-12-5-r45 (2011).
65. Lewin, T. D. *et al*. Fusion, fission, and scrambling of the bilaterian genome in Bryozoa. *Genome Res.* **35**, 78–92, https://doi.org/10.1101/gr.279636.124 (2025).
66. Lewin, T. D., Liao, I. J.-Y. & Luo, Y.-J. Annelid comparative genomics and the evolution of massive lineage-specific genome rearrangement in Bilaterians. *Mol. Biol. Evol.* **41**, Article msae172, https://doi.org/10.1093/molbev/msae172 (2024).
67. Lee, S.-H. *et al*. Insights into the evolution of ancient shark and ray sex chromosomes. *bioRxiv* https://doi.org/10.1101/2025.02.26.637739 (2025).

Acknowledgements

This study was funded by the Office of the Provost Faculty Start-Up Grant, Howard University, Washington, DC, awarded to LMN.

Author contributions

L.M.N. conceived the study. S.O.O. and K.A. collected samples. Bioinformatics analysis was performed by O.O.O., A.O.A., S.O.O and O.O., O.O.O., S.O.O., A.O.A., O.O. and I.C.N. wrote and revised the original manuscript. L.M.N. revised the final draft. All authors have read and approved the final 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/s41597-026-06646-0.

Correspondence and requests for materials should be addressed to L.M.N.