scientific reports

Check for updates

OPEN

Complete mitochondrial genome characterization and exploratory mitochondrial DNA modification signals in square-head catfish *Chaca chaca*

Partha Sarathi Tripathy[1], Lekiningroy Dann[2], Satya Narayan Parida[1], Ajaya Kumar Rout[1]✉, Janmejay Parhi[3], Bijay Kumar Behera[4]✉ & Pramod Kumar Pandey[1]✉

Complete mitochondrial genomes provide powerful insights into evolutionary relationships, phylogeography, and emerging epigenetic regulation, yet remain poorly characterized for many fishes. Here, we report the first complete mitogenome of the square-head catfish *Chaca chaca* and integrate mitogenomic evolution with mitochondrial DNA (mtDNA) methylation profiling. High-molecular-weight DNA was sequenced using Oxford Nanopore adaptive sampling, followed by de novo assembly, annotation, and comparative analyses. The assembled mitogenome is 16,591 bp long and contains the standard complement of 13 protein-coding genes, 22 tRNAs, two rRNAs, and a complete control region. Phylogenetic reconstruction based on concatenated alignments of the 13 mitochondrial protein-coding genes (PCGs) recovered *C. chaca* within Chacidae and adjacent to Pimelodidae + Pseudopimelodidae. Genome-wide methylation analysis detected signals consistent with 5-methylcytosine (5mC) and N6-methyladenine (6mA) across the mitogenome, with differences in modification frequency observed between strands after normalization for base composition within this dataset. The separation was more pronounced for 5mC compared to 6mA and is consistent with strand-associated asymmetry observed in this dataset. In addition to this, two cytosine positions within an intergenic spacer between *tRNA-Asp* and *COXII* showed detectable 5mC-associated signal. This mitogenome provides a foundational genomic resource for *C. chaca*, enabling refined phylogenetic resolution, conservation genetics, and future investigations into the potential relevance of mtDNA modification signals, pending further validation in fishes.

Keywords *Chaca chaca*, Mitogenome, Phylogenomics, DNA methylation, Strand bias, Conservation genetics

Mitogenomes, owing to their distinct characteristics, including maternal inheritance, high mutation rates, and conserved gene content, have become indispensable tools for investigating phylogenetic relationships and evolutionary histories across diverse taxa[1]. The application of long PCR and fish-versatile primers has significantly advanced the sequencing of complete mitogenomes in fishes, contributing extensively to molecular phylogenetics over the past fifteen years[2]. This robust methodology facilitates comprehensive analyses of genetic diversity within fish populations and elucidates the genetic characteristics of hybrid fish species[3]. For instance, complete mitogenomes have been successfully sequenced for hybrid groupers and hybrid carp, providing critical insights into their genetic makeup and taxonomic placements[4–6]. Furthermore, mitochondrial DNA (mtDNA) polymorphisms have proven instrumental in analyzing hybridization and introgression among various fish species, elucidating complex genetic interactions[7]. Modern sequencing technologies, such as those employed in the Vertebrate Genomes Project, have further streamlined the assembly of mitogenomes, enabling rapid

[1]College of Fisheries, Rani Lakshmi Bai Central Agricultural University, Datia, Madhya Pradesh 475686, India. [2]Department of Fisheries Resource Management, Kerala University of Fisheries and Ocean Studies, Kochi, Kerala 682506, India. [3]ICAR-Central Institute of Freshwater Aquaculture, Kausalyaganga, Bhubaneswar, Odisha 751002, India. [4]National Fisheries Development Board, Department of Fisheries, Government of India, Hyderabad, Telangana 500052, India. ✉email: ajayabi2012@gmail.com; beherabk18@yahoo.co.in; pkpandey_in@yahoo.co.uk

and accurate reconstruction even from low-depth sequencing data[8]. This technical advancement has profound implications for understanding the evolutionary dynamics of fish, especially for species with limited genomic resources or those of conservation concern. Despite this, comprehensive mitochondrial genomic research in fishes remains limited, with much of the current focus centered on understanding how mitochondrial mutations contribute to detrimental phenotypes in humans rather than exploring the broader evolutionary implications in piscine lineages[9]. However, the burgeoning field of comparative genomics in teleost fishes now enables comprehensive analyses of mitogenome evolution, revealing selective pressures related to energetic demands and environmental adaptations[10].

Consequently, the sequencing and analysis of complete mitogenomes, such as that of the squarehead catfish, *Chaca chaca*, offer a crucial foundation for deciphering the phylogenetic relationships within the order Siluriformes and assessing its evolutionary status[11]. This study presents the first complete mitogenome sequence of *C. chaca*, further illuminating its taxonomic placement and evolutionary trajectory within this diverse fish order. This genomic information provides a vital resource for conservation genetics, enabling precise identification and monitoring of distinct evolutionary lineages within the species[12]. The availability of this complete mitogenome enables detailed comparative analyses with other Siluriformes, aiding the resolution of ambiguities in their phylogenetic tree[13]. Furthermore, the unique genetic markers within this mitogenome can serve as a barcode for species identification, offering a rapid and reliable method for distinguishing *C. chaca* from closely related species[14]. Mitogenomes, though relatively small, contain essential genes crucial for cellular respiration and are thus valuable for biotechnological applications and phylogenetics[15,16]. The use of advanced sequencing technologies, such as de novo assembly on Illumina platforms, facilitates comprehensive annotation of gene content, genetic makeup, and gene order in these mitogenomes, thereby enhancing phylogenetic resolution[17]. This approach has been successfully applied to other taxa, where complete mitogenomes have elucidated complex evolutionary relationships and informed conservation strategies[18]. For example, comparative mitochondrial analyses have been instrumental in understanding hybridization events and population dynamics in various fish species[7]. Such investigations provide critical data for assessing the conservation status of endangered species and identifying significant evolutionary units within their populations. The increasing availability of complete mitogenomes across diverse species, facilitated by tools such as CamlTree, enables more robust phylogenetic reconstructions, thereby refining our understanding of evolutionary divergence and speciation events within and between orders[13]. This refined phylogenetic resolution, in turn, offers a robust framework for assessing biodiversity and informing targeted conservation strategies for species like *C. chaca*.

Moreover, DNA barcoding, often using mitochondrial genes such as cytochrome c oxidase subunit I and cytochrome b, offers a pragmatic and universal approach to species identification and authentication, particularly for combating mislabeling in the fish industry[19]. The complete mitogenome provides a comprehensive set of genetic markers, surpassing the utility of individual gene segments for resolving intricate phylogenetic relationships and understanding deep evolutionary divergences[20]. This method, by leveraging the entire mitogenome, yields higher resolution and reliability for species-level identification compared to single-gene approaches[3]. This detailed genetic information is particularly valuable for species that are difficult to identify morphologically, providing an unambiguous genetic signature for *C. chaca*. This comprehensive dataset enables the identification of unique genetic variations crucial for distinguishing intraspecific lineages, thereby supporting targeted conservation efforts. Moreover, the complete mitogenome provides substantial insights into population genetics, enabling assessment of genetic diversity and population structure, which are essential for effective conservation management[21]. The application of mtDNA in such studies is widespread, offering insights into species identification, ploidy, clonal lineages, and phylogenetic relationships[22].

This extensive genetic resource provides an invaluable foundation for understanding the adaptive evolution of *C. chaca* in response to environmental pressures and contributes significantly to the broader phylogenetic understanding of the Siluriformes order. Such comprehensive mitogenomic data is also instrumental in developing novel DNA markers for forensic identification, which can further aid in combating illegal trade and ensuring the protection of species[23]. This is particularly relevant for species like *C. chaca*, where accurate identification is crucial for monitoring wildlife crime and preventing illegal trade[24]. Furthermore, the rich genetic information contained within the complete mitogenome provides a powerful tool for authenticating fish products and combating species substitution in commercial markets, addressing a critical need in food safety and consumer protection[25]. The comprehensive mitogenomic data from *C. chaca* can thus serve as a foundational resource for developing advanced molecular tools, ensuring accurate species identification and contributing to both conservation efforts and food security initiatives. This study, therefore, aims to characterize the complete mitogenome of *C. chaca* by analyzing its gene content, organization, and phylogenetic placement within the Siluriformes order. This characterization will involve detailed comparisons with existing mitogenomes of related species to elucidate evolutionary relationships and identify unique genetic features. This in-depth analysis will provide crucial insights into the molecular adaptations and evolutionary history of *C. chaca*, thereby strengthening conservation strategies[26,27]. Ultimately, this research aims to provide a robust genetic baseline for *C. chaca*, enabling the development of targeted genetic monitoring programs and informing in-situ and ex-situ conservation initiatives[28,29]. The resulting mitogenomic data will also facilitate the development of novel genetic markers for population genetic studies, which are crucial for understanding genetic diversity and population structure in this species. Moreover, the established complete mitogenome will serve as a reference for future ecological and evolutionary studies of the squarehead catfish, providing a foundation for understanding its adaptive capacity and resilience to environmental changes. It is worth mentioning that this study is primarily intended as a mitogenomic resource, with methylation analyses presented as exploratory observations from a single individual.

Materials and methods

Ethical statement

No live animals were captured or sacrificed specifically for this study. The specimen used for DNA analysis was obtained post-mortem from a local fish market, where it had been landed as part of routine commercial fisheries. Therefore, the study did not involve animal experimentation, handling of live specimens, or invasive procedures, and formal ethical approval was not required in accordance with national guidelines for the use of animals in research. The work relied solely on tissue collected from an already deceased specimen.

Sample collection, DNA extraction, and quality assessment

A single specimen of *C. chaca* was procured from the local fish market of Narayanpur, Assam, India (Lat: lon-26.955880428390916, 93.85502385285048), where it had been landed through routine fisheries harvest. The specimen was already deceased at the time of purchase and was not collected as part of targeted sampling. Species identification was performed based on external morphological characters following published taxonomic descriptions of Chacidae, including diagnostic features such as the depressed head, wide terminal mouth, dermal ornamentation, and fin ray configuration. A photographic record of the specimen was retained as a reference voucher, and tissue samples used for molecular work are archived at the College of Fisheries, RLBCAU, Datia, India, under laboratory reference ID: CC-MITO-01-2025. White muscle tissue was excised and immediately preserved in absolute ethanol to prevent DNA degradation[30]. High-molecular-weight genomic DNA, essential for long-read sequencing technologies, was extracted from the preserved tissue using Monarch® HMW DNA Extraction Kit for Tissue (NEB, USA) following manufacturer's protocol. DNA quantity was measured using a Qubit fluorometer (Thermofisher Scientific, USA). DNA purity was assessed using a NanoDrop spectrophotometer (Thermofisher Scientific, USA). The integrity and fragment size distribution of the extracted HMW DNA were evaluated through agarose gel electrophoresis.

Nanopore library preparation

Library preparation was performed using the Ligation Sequencing Kit SQK-LSK114 (Oxford Nanopore Technologies, UK) following manufacturer's protocol. DNA repair and end-preparation steps were conducted to ensure that DNA fragments have blunt ends and 5'-phosphorylated termini. Following this, dA-tailing was performed to add an 'A' overhang to the 3' ends of the DNA fragments. Subsequently, specific sequencing adapters are ligated to these prepared DNA fragments. Cleanup steps using AMPure XP beads were then performed to remove unligated adapters and short DNA fragments, ensuring the enrichment of the desired library molecules. The final library was quantified using a Qubit fluorometer.

Nanopore sequencing with adaptive sampling on P2 Solo

Sequencing was conducted on the Oxford Nanopore Technologies P2 Solo device, using R10.4.1 flow cell chemistry. The sequencing run was initiated using the MinKNOW v25.09.16 (https://nanoporetech.com/document/experiment-companion-minknow) and adaptive sampling (also known as "Read Until") was implemented to selectively enrich for mtDNA reads. The reference mitogenome of *Chaca bankanensis* (Accession number: AP012008) from NCBI was uploaded to the sequencer and this technique operates in real-time, where MinKNOW performed rapid alignment of reads against this provided reference sequence. Although the reference mitogenome of *Chaca bankanensis* (AP012008) used for adaptive sampling does not include a fully resolved control region (D-loop), this did not prevent recovery of the corresponding region in *C. chaca*. Adaptive sampling was used only to enrich reads showing similarity to the reference sequence and does not strictly exclude non-matching regions. Consequently, reads originating from the control region of *C. chaca* were retained during sequencing. The sequencing run was allowed to proceed for 48 h to achieve sufficient depth of coverage for the mitogenome.

Data processing and mitogenome assembly

Raw Nanopore signal data (in POD5 format) were basecalled using Dorado v1.3.0 (https://github.com/nanoporetech/dorado) with the high-accuracy model to generate FASTQ reads. Because adaptive sampling was employed during sequencing to enrich mtDNA, the resulting dataset was inherently enriched for mitochondrial reads through real-time selection performed by the MinKNOW v25.09.16 (https://nanoporetech.com/document/experiment-companion-minknow). No additional post-basecalling read filtering, subsampling, or selection step was applied. Accordingly, the term "filtered mitochondrial read set" used here refers solely to the enriched read pool generated through adaptive sampling, rather than to any subsequent bioinformatic filtering procedure. Adaptive sampling operates by selectively retaining reads that show similarity to a provided reference during sequencing, but it does not impose strict exclusion boundaries. As a result, reads corresponding to divergent mitochondrial regions, including hypervariable segments such as the control region (D-loop), as well as partially matching fragments, are retained within the dataset. Consequently, all basecalled reads produced under adaptive sampling conditions were included in downstream assembly without exclusion based on alignment identity, read length, or mapping criteria. The *de novo* assembly of the enriched read dataset was performed using Flye v2.9.6 with parameters optimized for circular genome reconstruction (--nano-hq, --meta disabled). Flye identifies candidate circular contigs by detecting terminal repeat overlaps and resolving circular structures internally. Assembly polishing was conducted to improve consensus accuracy. First, two rounds of Racon v1.5.0 (https://github.com/lbcb-sci/racon) were applied using the original Nanopore reads mapped back to the assembly to correct systematic long-read errors. This was followed by polishing with Medaka v2.2.0 (https://github.com/nanoporetech/medaka) using the r10.4.1_e8.2_400bps_sup model to obtain a high-quality consensus sequence. To validate assembly completeness and structural correctness, all basecalled reads were remapped to the polished assembly using minimap2. Continuous read alignment across the circularization junction, uniform

depth-of-coverage across the genome, and the absence of coverage breakpoints supported successful recovery of a complete circular mitochondrial genome. Overlapping terminal regions identified during assembly were trimmed to generate a single continuous circular sequence. Assembly quality was further evaluated based on (i) uniformity of sequencing depth across the mitogenome, (ii) absence of conflicting or chimeric alignments upon read remapping, and (iii) concordance with the expected vertebrate mitochondrial gene order as determined during annotation with MitoAnnotator. The finalized circular mitogenome was used for all downstream analyses and deposited in the NCBI GenBank database.

DNA methylation analysis

The final assembled mitogenome was used as a reference to detect 5-methylcytosine (5mC) and N6-methyladenine (6mA) signals via basecalling with Dorado v1.3.0 using the R10.4.1 SUP modified-base model. The resulting modification calls were processed using modkit v0.6.0 to obtain per-read, per-site modification probabilities. To ensure reliability of modification signals, only genomic positions with a posterior modification probability ≥ 0.8 and minimum read coverage ≥ 100× were retained. Because individual nucleotide positions within mtDNA cannot be considered independent biological replicates and given that the study is based on a single individual, all downstream analyses were designed as descriptive, site-level comparisons within this dataset, rather than inferential statistical testing. Methylation frequency was calculated at each callable site as the proportion of reads supporting a modification at that position. These site-level methylation frequencies were used as the primary observational units for all downstream comparisons, including strand-specific analyses. Distributions of methylation frequencies were compared between heavy- and light-strand–encoded positions using rank-based approaches. The Wilcoxon rank-sum statistic was used descriptively to assess distributional shifts, and Cliff's delta was reported as a nonparametric measure of effect size. These statistics are presented to quantify differences within this dataset and are not interpreted as population-level inference. Feature-level summaries (e.g., protein-coding genes, rRNAs, control region) were generated by averaging site-level methylation frequencies within each genomic feature. These aggregated values were used only for descriptive visualization and interpretation and were not treated as independent observational units for statistical comparison. To reduce potential false-positive signals associated with Nanopore-based modification detection, several internal filters were applied: (i) exclusion of low-complexity and homopolymeric regions prone to elevated false-positive rates, (ii) enforcement of a high coverage threshold (≥ 100×) to stabilize frequency estimates, and (iii) In addition, modification levels in genomic regions such as *tRNA* clusters were examined as an empirical sanity check to contextualize background signal levels within the dataset. However, this comparison does not constitute a formal calibration strategy and cannot substitute for experimental controls or orthogonal validation methods. Accordingly, all detected modification signals are interpreted as dataset-specific observations rather than validated biochemical modifications. All analyses and visualizations were conducted in R using ggplot2, dplyr, and effsize. Proportion estimates were calculated using only callable positions (coverage ≥ 100×), ensuring that denominators represent confidently observed bases rather than total genomic length.

Because the study was conducted on a single opportunistically obtained specimen, downstream analyses, particularly mtDNA methylation profiling, were interpreted as descriptive at the level of the sampled individual rather than as population- or species-level estimates. Factors such as tissue specificity, unknown age, post-mortem interval, and preservation conditions may influence methylation detection in long-read sequencing datasets. Therefore, the present work is intended to establish a first reference framework for *C. chaca* mitogenomics and to generate testable hypotheses for future population-scale validation rather than to infer functional or ecological significance.

Phylogenetic analysis

The complete mitogenome sequence of *C. chaca*, newly sequenced in this study, was integrated with complete mitogenome sequences of various Siluriformes species retrieved from the MitoFish database[31]. *Danio rerio* was selected as an outgroup to root the phylogeny, consistent with broader teleost phylogenetic studies. Phylogenetic analysis was conducted using the streamlined CamITree software[13]. To minimize potential bias associated with highly variable non-coding regions and compositional heterogeneity typical of mitochondrial genomes, multiple phylogenetic datasets were constructed. The primary dataset consisted of concatenated alignments of the 13 mitochondrial protein-coding genes (PCGs), excluding the control region and other non-coding segments. Nucleotide sequences were aligned gene-wise using MAFFT and concatenated, followed by trimming with trimAl to remove poorly aligned regions. To evaluate robustness of the inferred topology, three alternative data treatments were analyzed in parallel: (i) amino-acid translations of all PCGs to reduce nucleotide compositional bias, (ii) nucleotide alignments excluding third codon positions (retaining 1st+2nd positions only) to limit substitutional saturation, and (iii) a full PCG nucleotide matrix with partition-specific models selected by ModelFinder[32]. Maximum-likelihood phylogenies were reconstructed in IQ-TREE using partition-aware models with 1000 ultrafast bootstrap replicates and SH-aLRT support[32]. Concordance among these datasets was used to assess stability of the phylogenetic placement of *C. chaca*. The final tree was then visualized in FigTree inside CamITree.

Results and discussion

Sequencing statistics

The complete mitogenome sequence of *C. chaca* was successfully determined using nanopore sequencing (Fig. 1). Nanopore sequencing generated 2,556,478 raw reads (basecalled from POD5 files), corresponding to approximately 2.5 Gb of total sequence yield. The coverage was based on the total number of mapped mitochondrial bases rather than the number of alignment segments. Based on a mean sequencing depth of ~ 1500× and a genome length of 16,591 bp, the total mapped mitochondrial bases were approximately ~ 24.9 Mb.

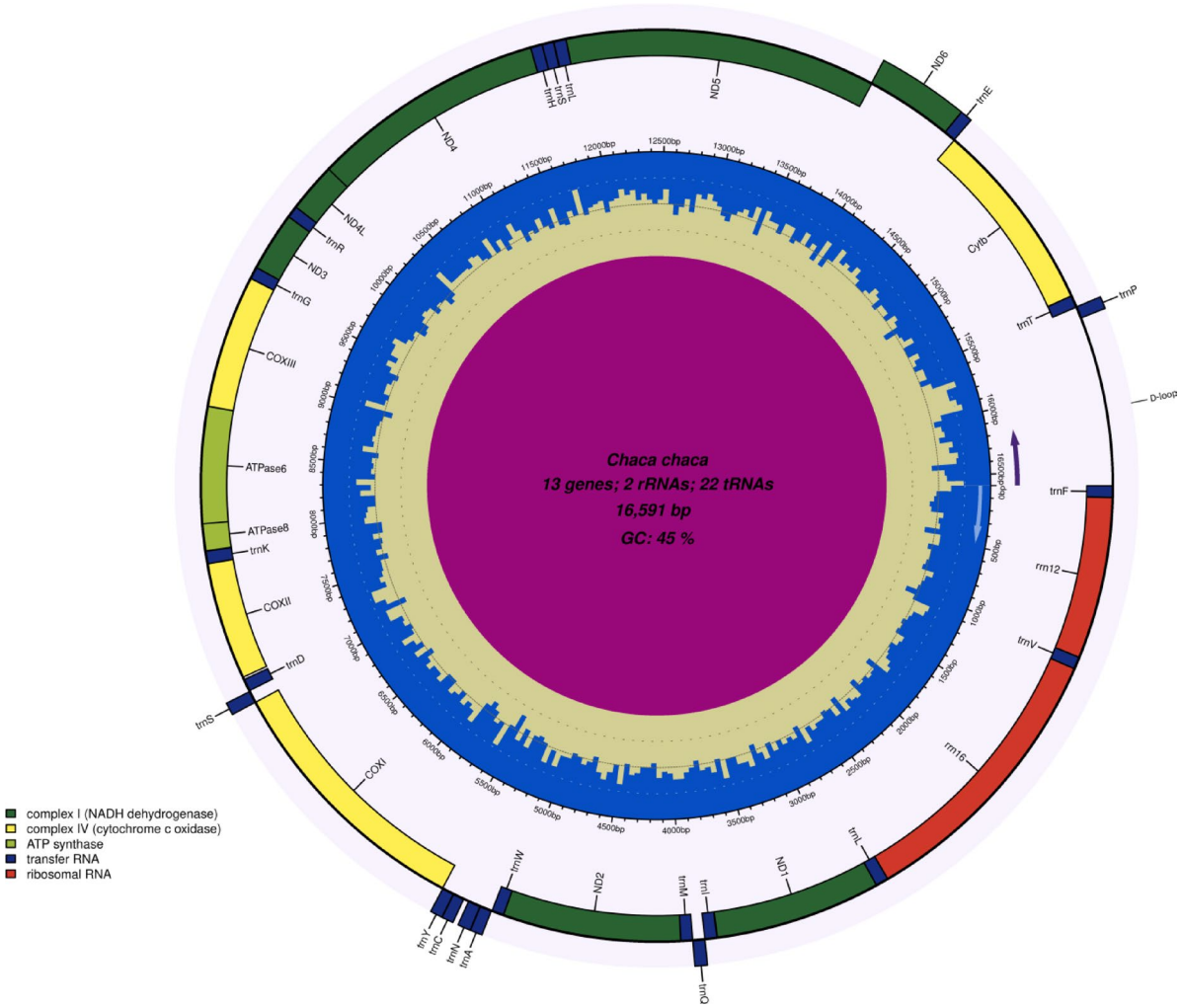

Fig. 1. Circular map of the complete mitogenome of *Chaca chaca* (16,591 bp). The genome consists of 13 protein-coding genes, 22 tRNA genes, two ribosomal RNA genes and a control region/D-loop. Gene orientations are indicated by arrow direction. The inner rings display GC content variation and GC skew, highlighting compositional patterns across the mitogenome. Annotation was generated and visualized using MitoAnnotator.

This value represents effective mitochondrial coverage after enrichment and read selection, not whole-genome sequencing depth. However, slightly lower coverage was observed in certain regions of the control region, likely attributable to the presence of repeat sequences[33]. Remapping of raw reads to the final assembly showed continuous alignment across the circularization junction with no evidence of structural breaks or alternative haplotypes, supporting recovery of a single complete mitochondrial genome. Nanopore sequencing has proven to be an effective method for obtaining complete and accurate mitogenomes[34,35].

The assembled mitogenome, constructed using Flye v2.9.6, measured 16,591 bp. This size is consistent with the typical range observed for closely related species within the order Siluriformes, where mitogenomes often fall within 16,500 to 17,000 bp[36]. Analysis of the overall mitogenome's nucleic acid composition revealed percentages of 31.80% A, 23.01% T, 15.20% C, and 29.99% G, resulting in a G-C content of 45.20%. Such nucleotide compositions and GC content are characteristic of fish mitogenomes[37,38]. The annotated genes from the complete mitogenome has been mentioned in Table 1.

The D-loop, also known as the displacement loop or control region, is a non-coding segment containing transcription and replication start sites[39]. In *C. chaca*, this region spanned positions 15,598 to 16,590 bp, with an exact size of 993 bp. This region constitutes approximately 5.98% of the total mitogenome. The D-loop is recognized for its significant genetic variation, including numerous polymorphisms such as insertions and deletions, making it a valuable marker for tracing ancestry and population genetic studies[39,40]. In the previously reported mitogenome of *Chaca bankanensis* (GenBank accession: AP012008), the control region was incomplete or not fully resolved, likely due to technical limitations of earlier sequencing approaches. The present study provides the first fully assembled control region for the genus *Chaca*, enabled by long-read sequencing, which allows resolution of repetitive and structurally complex regions.

Gene	Start	End	Strand	Length (bp)
ND1	2821	3795	+	975
ND2	4009	5053	+	1045
COXI	5439	6989	+	1551
COXII	7151	7841	+	691
ATPase8	7917	8084	+	168
ATPase6	8075	8757	+	683
COXIII	8758	9541	+	784
ND3	9613	9961	+	349
ND4L	10032	10328	+	297
ND4	10322	11702	+	1381
ND5	11914	13740	+	1827
ND6	14249	13737	–	513
Cytb	14320	15457	+	1138
12S rRNA	70	1018	+	949
16S rRNA	1091	2745	+	1655
tRNA-Phe	1	69	+	69
tRNA-Val	1019	1090	+	72
tRNA-Leu (UAA)	2746	2820	+	75
tRNA-Ile	3798	3869	+	72
tRNA-Gln	3939	3869	–	71
tRNA-Met	3939	4008	+	70
tRNA-Trp	5054	5124	+	71
tRNA-Ala	5196	5128	–	69
tRNA-Asn	5270	5198	–	73
tRNA-Cys	5366	5301	–	66
tRNA-Tyr	5437	5368	–	70
tRNA-Ser (UGA)	7060	6990	–	71
tRNA-Asp	7065	7136	+	72
tRNA-Lys	7842	7915	+	74
tRNA-Gly	9542	9612	+	71
tRNA-Arg	9962	10031	+	70
tRNA-His	11703	11771	+	69
tRNA-Ser (GCU)	11772	11838	+	67
tRNA-Leu (UAG)	11841	11913	+	73
tRNA-Glu	14318	14250	–	69
tRNA-Thr	15458	15529	+	72
tRNA-Pro	15597	15528	–	70
Origin of light-strand replication (O_L)	5300	5271	–	30
D-loop (Control region)	15598	16590	+	993

Table 1. Annotation of complete mitogenome of *Chaca chaca*.

A comparative analysis of the *C. chaca* mitogenome with those of related Siluriformes species was performed. Visualization tools are commonly employed in comparative genomics to facilitate the identification of conserved and variable regions across species[41,42]. The results of this comparative analysis, presented as a BLAST Atlas using the CGview Comparison Tool, are provided in Fig. 2.

Phylogenetic analysis

Clarified familial placement of chacidae

The phylogenetic dataset comprised 200 complete mitochondrial genomes representing major lineages within Siluriformes, with *Danio rerio* as the outgroup. The final concatenated alignment of 13 mitochondrial protein-coding genes had a total length of 21,381 bp after trimming. Poorly aligned and ambiguous regions were removed using trimAl to improve alignment quality.

Historically, the family-level placement of Chacidae within Siluriformes has been largely problematic, frequently remaining unresolved or being treated as an isolated lineage in higher-level summaries. This ambiguity is underscored by the fact that the phylogenetic affinities of Chacidae to other catfishes remain unclear despite numerous studies, with various hypotheses proposing relationships with diverse families such as Clariidae, Sisoroidei, Loricarioidei, Plotosidae, and Bagridae[43]. All alternative datasets (amino-acid matrix, exclusion of third codon positions, and PCG-only nucleotide dataset) recovered the same overall topology and consistently

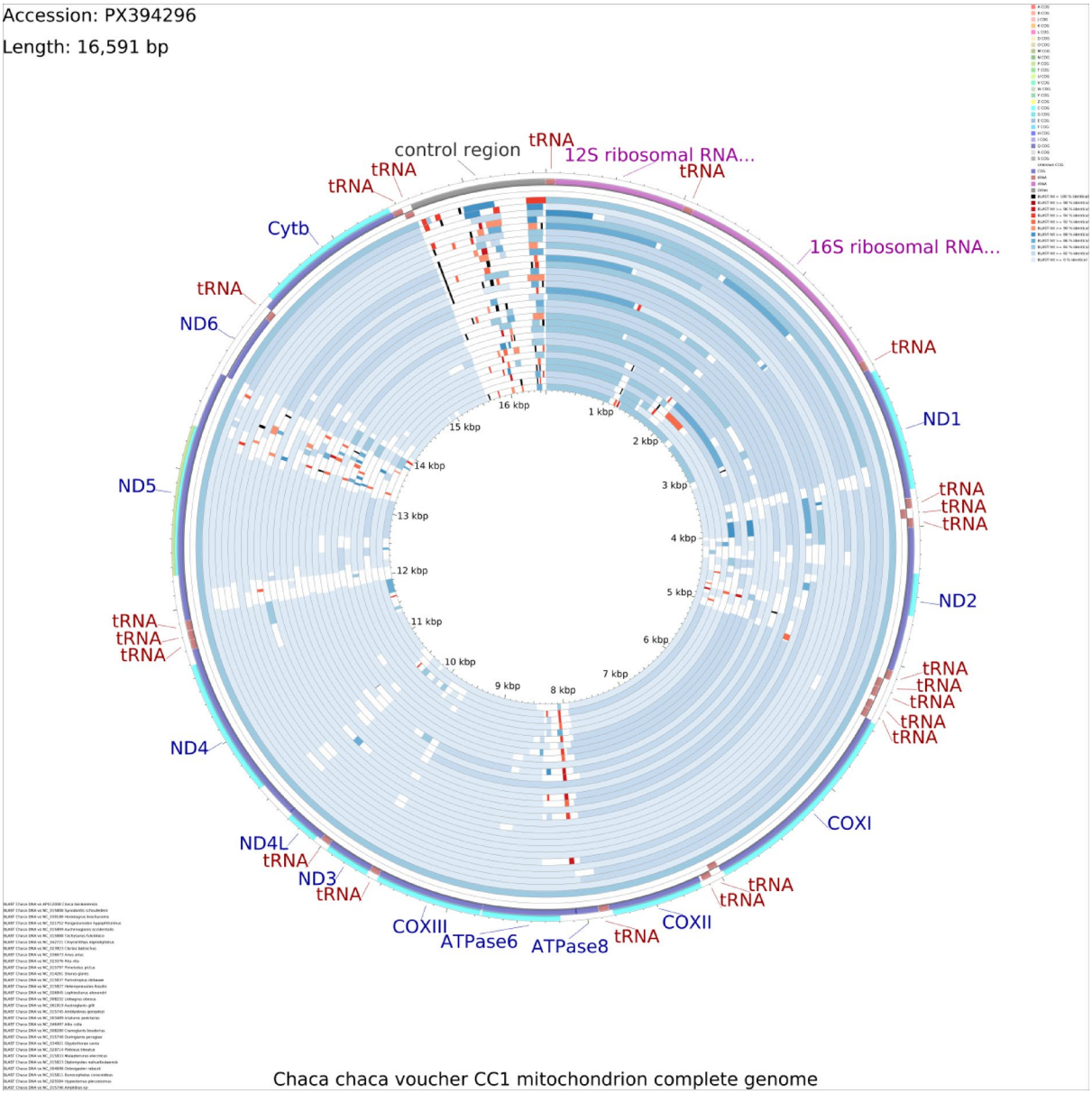

Fig. 2. Comparative BLAST atlas of the *Chaca chaca* mitogenome (accession: PX394296; length: 16,591 bp) generated using the CGView Comparison Tool (CCT). The outer ring displays annotated mitochondrial genes. Inner concentric rings represent BLAST similarity profiles of *C. chaca* against multiple Siluriformes mitogenomes, where darker blue shading indicates higher sequence similarity and white regions indicate lower or no alignment. Red blocks denote regions of divergence or reduced homology.

placed *C. chaca* as sister to the pimelodoid clade. Branch support values varied slightly among analyses, but no conflicting deep relationships were observed, indicating that the inferred placement is not driven by control-region variability or nucleotide compositional bias.

Early osteological work on *C. bankanensis* highlighted numerous autapomorphies of its cephalic and pectoral girdle structures[44,45]. These unique characteristics, however, often complicated efforts to unambiguously link Chacidae to any of the major catfish superfamilies, further contributing to the uncertainty surrounding its classification[45]. Similarly, early molecular studies employing a limited number of nuclear markers (e.g., rag1, rag2)[46] or restricted mitochondrial regions often suffered from incomplete taxon sampling of Chacidae, leading to a persistent lack of resolution in its deeper phylogenetic relationships[36,43].

This study provides a mitogenome-scale perspective on the placement of Chacidae by incorporating the first complete mitogenome of *C. chaca* alongside the previously available mitogenome of *C. bankanensis*. Our analysis (Fig. 3) provides strong and consistent support, within the present mitochondrial dataset, for a clade comprising Chacidae + (Plotosidae + Pimelodidae + Pseudopimelodidae), placing *C. chaca* within the broader pimelodoid radiation. This placement was recovered consistently across all alternative dataset treatments (amino-acid matrix, exclusion of third codon positions, and PCG-only nucleotide dataset), indicating that the

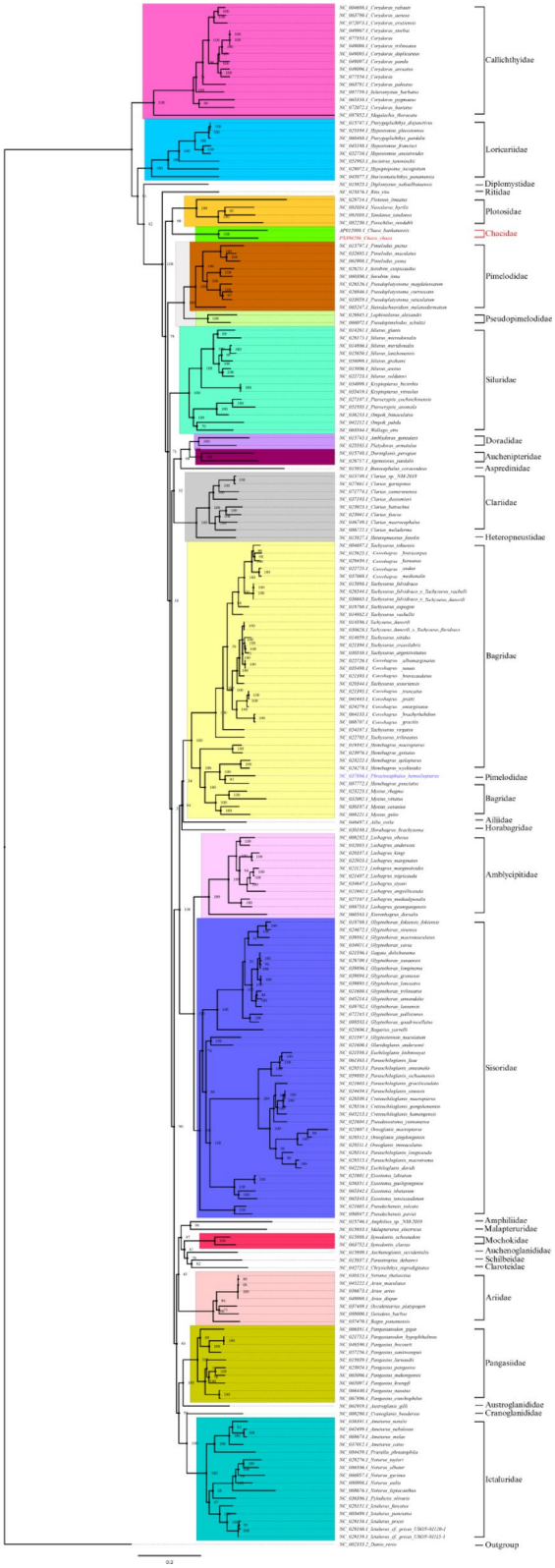

Fig. 3. Maximum-likelihood phylogenetic tree of *Chaca chaca* and representative Siluriformes based on concatenated alignments of the 13 mitochondrial protein-coding genes (PCGs). The tree was constructed using concatenated mitogenomic sequences aligned with MAFFT and analyzed in IQ-TREE under the best-fit partition models, with node support assessed through 1,000 ultrafast bootstrap replicates and SH-aLRT tests in CamlTree. *C. chaca* (highlighted in red color) forms a strongly supported monophyletic Chacidae clade with *C. bankanensis*, positioned within the pimelodoid assemblage and sister to Pimelodidae + Pseudopimelodidae. *Danio rerio* serves as the outgroup.

inferred topology is not driven solely by compositional bias or control-region variability. However, because this study relies exclusively on mitochondrial genome data, the resulting topology should be interpreted as a mitochondrial gene-tree signal rather than a definitive resolution of higher-level phylogenetic relationships. Alternative familial hypotheses proposed in earlier morphological or limited-marker studies were not explicitly tested using constrained-topology or Approximately Unbiased (AU) tests in the present analysis and therefore cannot be statistically excluded[43]. Accordingly, rather than resolving longstanding ambiguities in the familial placement of Chacidae, the current results demonstrate that the mitochondrial dataset provides a consistent and reproducible phylogenetic signal under multiple analytical treatments, supporting its placement within the pimelodoid radiation in this dataset. Further confirmation using multilocus nuclear data and formal topology testing will be necessary to evaluate the robustness of this relationship at the species-tree level.

The inferred sister-group relationship between Chacidae and Pimelodidae is particularly intriguing given their striking ecological and morphological divergences. Many pimelodid species are characterized as large, active predators inhabiting extensive river channels in South America[46,47]. In stark contrast, *Chaca* species, commonly referred to as "angler catfishes" or "frogmouth catfishes"[43], are typically small, sedentary, benthic ambush predators, exhibiting epidermal morphological specializations indicative of this ecological niche[48]. Because this study relies on mitochondrial genome data alone, the resulting topology should be interpreted as a single line of molecular evidence rather than as a definitive test of competing higher-level siluriform hypotheses. Future phylogenomic analyses integrating independent nuclear loci will be necessary to evaluate whether this relationship reflects organismal history or the known limitations of mitochondrial-only inference.

Phylogenetic caveats and dataset limitations

Interpretation of mitochondrial phylogenies at deeper taxonomic levels requires caution because several well-recognized factors can influence inferred relationships. First, mitochondrial genomes exhibit lineage-specific compositional biases that may contribute to long-branch attraction, particularly when comparing distantly distributed taxa. Second, public database sequences rely on prior identifications that cannot be independently verified here, and potential misidentification cannot be entirely excluded. Third, as mtDNA represents a single, non-recombining locus, the recovered gene tree may not fully reflect the species tree due to processes such as incomplete lineage sorting or historical introgression. To mitigate these effects, alternative dataset treatments (amino-acid translation and exclusion of third codon positions) were examined and yielded congruent topologies, suggesting that the observed placement is not solely driven by nucleotide compositional signal. Nevertheless, the relationships presented here should be regarded as mitochondrial in origin and require confirmation from multilocus nuclear datasets.

Monophyly and internal diversification of Chaca

Within Chacidae, the two sampled species, *C. chaca* and *C. bankanensis*, consistently form a maximally supported sister-pair. This finding indicates that the genus *Chaca* is monophyletic, at least among the taxa sampled, and that the profound morphological similarities observed across described species reflect a shared ancestry[49]. Combined with taxonomic compilations that recognize *C. chaca*, *C. bankanensis*, *C. burmensis*, and *C. serica* as valid members of a single genus[43], our phylogenetic tree supports the current familial concept of Chacidae as a small but evolutionarily distinct Asian lineage. The relatively long stem branch leading to Chacidae, in comparison to branches within Pimelodidae or Pseudopimelodidae, suggests an extended period of independent evolution for the frogmouth catfishes. Within this framework, the divergence between *C. chaca* and *C. bankanensis* appears to be well supported in the mitochondrial phylogeny; however, the timing of this divergence cannot be inferred in the absence of molecular clock analyses or calibrated phylogenetic models. The relatively long branch separating these lineages in the tree indicates substantial genetic differentiation within the mitochondrial dataset but does not provide a direct estimate of divergence time. Similarly, although the present-day geographic distribution of *Chaca* species spans South and Southeast Asia[50], no explicit historical biogeographic reconstruction was performed in this study. Therefore, interpretations invoking specific dispersal or vicariant scenarios, including links to regional geological history such as Sundaland, remain speculative and are not directly supported by the current data. Accordingly, the results are best interpreted as demonstrating phylogenetic distinctiveness and geographic structuring within the genus *Chaca* in the mitochondrial dataset, without inferring the timing or mechanisms underlying lineage divergence. Future studies incorporating time-calibrated phylogenies and formal biogeographic modeling will be required to test these hypotheses.

Biogeographic and ecological implications

The genus *Chaca* is restricted to tropical South and Southeast Asia, inhabiting sluggish, often turbid lowland waters from the Ganga-Brahmaputra system westward to peninsular India and eastward to Borneo and Sumatra[43,49]. In the phylogeny constructed for this study, *C. chaca* (from the Indian region) and *C. bankanensis* form the most strongly supported sister lineages. This phylogenetic relationship, combined with their disjunct distributions, indicates the coexistence of a sister relationship and geographic separation between the two species in the mitochondrial dataset, without implying a specific historical biogeographic scenario. The relatively long branches separating these species reflect genetic divergence within the mitochondrial dataset but do not provide direct evidence of the timing or mechanisms underlying lineage separation. Although geological processes have been shown to influence freshwater faunal diversification[50,51], such processes are not explicitly inferred in the present study.

Ecologically, both *C. chaca* and *C. bankanensis* are highly specialized benthic predators[48]. They exhibit a unique "frogmouth" morphology[43] and spend much of their time camouflaged, buried in soft substrates, from which they rapidly engulf unsuspecting prey using their capacious mouths[48,49]. Their phylogenetic placement within an otherwise more morphologically "typical" pimelodoid radiation underscores the remarkable extent to

which feeding mode and body shape can diverge, while the overall phylogenetic signal in the mitogenome still clearly tracks shared ancestry. This striking morphological and ecological specialization, despite strong molecular affinity to less specialized groups, highlights Chacidae as a compelling model for future studies investigating the genomic underpinnings of extreme trophic specialization within catfishes.

Concordance with previous molecular and morphological work

Our phylogenetic topology is broadly congruent with recent multilocus and mitogenomic reconstructions of Siluriformes, which recover loricarioids as sister to all remaining catfishes or as an early diverging lineage, and Diplomystidae as another basal siluroid lineage[52,53]. These broader phylogenies also consistently identify a coherent pimelodoid assemblage, which includes families such as Pimelodidae, Pseudopimelodidae, and Heptapteridae[52,53]. The novel contribution of our work is the dense mitochondrial placement of *C. chaca*, which provides an additional mitochondrial line of evidence for the placement of Chacidae within the pimelodoid assemblage.

Morphological analyses of *C. bankanensis* have identified several distinct features in its cephalic region and pectoral girdle, including the configuration of the palatine-maxillary mechanism, cranial musculature, and pectoral girdle[44]. Historically, these morphological characters led to *suspected* affinities with pimelodoid families, although definitive links were not unambiguously established through osteological work alone[43,44]. The fact that our mitogenome phylogeny independently recovers Chacidae within the pimelodoid radiation increases confidence that both molecular data and these previously considered morphological characters are reflecting the same underlying evolutionary history, rather than being methodological artifacts.

However, given that the present study is based solely on mitochondrial genome data, the observed concordance between mitochondrial topology and previously described morphological characters should be interpreted as provisional and dataset-specific, rather than as confirmation of shared organismal evolutionary history. Accordingly, while the mitochondrial dataset provides a coherent signal placing Chacidae within the pimelodoid radiation in this analysis, independent validation using multilocus nuclear datasets will be required to assess whether this relationship reflects species-level phylogeny.

DNA methylation patterns across the C. chaca mitogenome

The *C. chaca* mitogenome showed strand-specific nucleotide compositional asymmetry, a characteristic vertebrate trait driven by asymmetric replication and mutational processes[54,55]. Under the strand-displacement model, the parental majority strand remains single-stranded for extended periods during replication, increasing its susceptibility to spontaneous hydrolytic deamination of cytosine and adenine[54,56]. Consequently, GC- and AT-skew metrics were used to assess this bias, as they more accurately reflect directional mutational pressure than raw base counts[54,57].

Analysis of the *C. chaca* mitogenome revealed specific nucleotide counts on the light (L) strand (A: 5,276; T: 3,817; G: 2,522; C: 4,976) and heavy (H) strand (A: 3,817; T: 5276; G: 4,976; C: 2522) (Supplementary file: meth.xlsx). This distribution results in a positive AT-skew (≈ 0.16) and a negative GC-skew (≈ − 0.32) on the L-strand, consistent with the guanine-poor and adenine-rich architecture typical of vertebrates[54,57]. Accounting for these baseline asymmetries was essential during the normalization of modification frequencies to prevent inherent compositional bias from confounding the epigenetic data. To ensure high-quality signals, only positions with ≥ 100x coverage were analyzed (Fig. 4a). This threshold was selected because Nanopore-based mitochondrial modification frequencies require a depth of at least 100× to achieve statistical stability and minimize stochastic noise[58]. Additionally, a visualization threshold of ≥ 25% modification was applied to distinguish robust signals from the low background methylation (typically < 7%) observed in various vertebrate mitochondrial contexts[59,60]. This threshold serves as a data-specific filter for high-confidence signals rather than a universal biological marker[54,60].

As these findings are derived from a single specimen, they represent an initial descriptive profile of mtDNA modification in *C. chaca* rather than a definitive species-level estimate. Nonetheless, these results align with recent single-molecule sequencing studies and provide a foundational baseline for understanding the mitogenomic epigenome in siluriform fishes[60,61].

Overall detection of mtDNA modification signals

Across all callable positions, signals consistent with 5-methylcytosine (5mC) and N6-methyladenine (6mA) were detected at subsets of cytosine and adenine residues, respectively. When calculated relative to total callable bases (coverage ≥ 100×), 8.86% of cytosines and 7.29% of adenines showed detectable modification signal prior to application of the ≥ 25% visualization threshold. These values represent overall detection frequencies within this dataset and do not distinguish between high- and low-confidence biological modifications. Reports of 5mC and 6mA signals in mtDNA have been described in multiple taxa, although their accurate detection and biological interpretation remain subjects of ongoing methodological evaluation. Accordingly, the results presented here are limited to a quantitative description of observed signal patterns without inferring functional or regulatory roles.

Strand-wise distribution of modification signals

The distribution of DNA modification signals across the *C. chaca* mitogenome was markedly uneven between the two strands (Fig. 4b). For 5-methylcytosine (5mC), 345 sites were identified on the light (L) strand, compared to 319 sites on the heavy (H) strand. This distribution became more polarized for N6-methyladenine (6mA), where 437 sites were located on the L-strand, while only 226 sites were detected on the H-strand. Even after normalization by the total number of callable bases per strand, these strand-associated differences remained distinct. Such asymmetry is consistent with the inherent mutational and structural biases of the vertebrate mitogenome, including the extended single-stranded state of the L-strand during strand-displacement

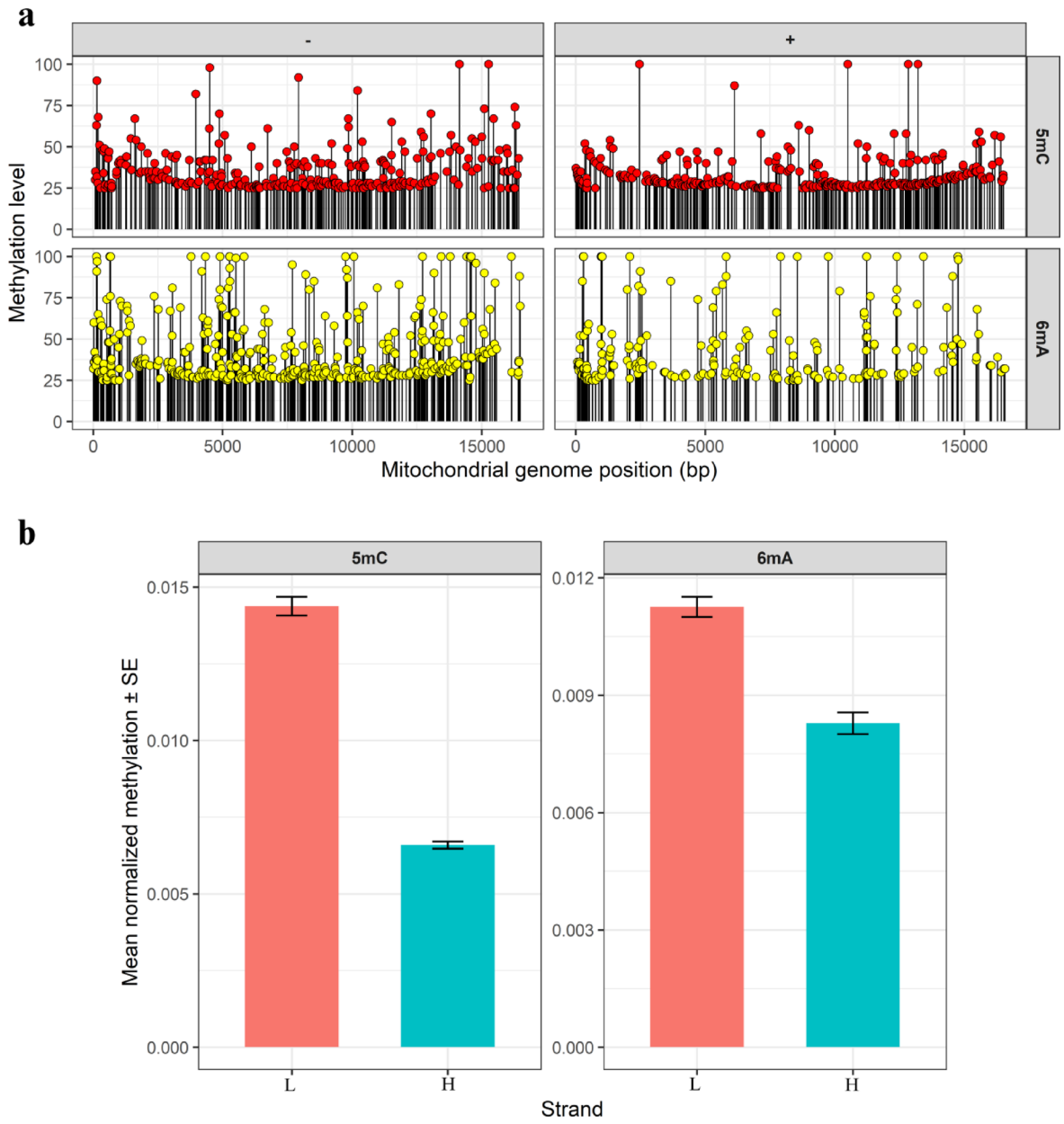

Fig. 4. **a** Strand-specific genome-wide distribution of mtDNA methylation. Lollipop plots show the genomic positions of individual methylation sites along the mitogenome, plotted separately for the H and L-strands. **b** Strand-wise normalized mtDNA methylation levels of 5mC and 6mA. Bar plots represent the mean normalized methylation values for the H and L-strands, with error bars indicating the standard error (SE). Normalization accounts for differences in the total number of methylated sites between strands, enabling direct comparison of strand-specific methylation bias for each modification type.

replication[54,55]. However, these findings should be interpreted with caution; while similar strand-specific patterns have been documented in other vertebrate studies[61,62], direct comparisons are often confounded by differences in sequencing platforms and bioinformatic pipelines. Furthermore, Nanopore-based detection is sensitive to local sequence context, coverage fluctuations, and potential model-specific biases that may disproportionately affect one strand[58,63]. Consequently, the observed asymmetry is presented here as a dataset-specific profile rather than a definitive, bias-corrected estimate of the *C. chaca* mitogenome.

Regions with elevated modification signal density

Genomic regions contributing ≥ 5% of total detected modification sites were identified as areas with relatively higher signal density. For 5mC, these included *ND5, ND4, ND2, ND1, COXI, COXIII, Cytb, 12S rRNA, 16S rRNA*, and the control region. For 6mA, the corresponding regions were *12S rRNA, ND5, COXI, ND4, ND2, ATPase6*, and *Cytb*. The clustering of modifications in these specific genes highlights them as focal points for epigenetic study in the *C. chaca* mitogenome. While these regions represent locations where modification signals

were most frequently called in this dataset, no broader inferences regarding functional enrichment or biological significance are made at this stage, consistent with the descriptive nature of initial mitogenomic modification profiles[60,61].

Strand-associated differences in modification frequency

The evaluation of strand-associated differences in modification frequency was conducted by comparing site-level distributions across all callable positions. Because these observations are derived from a single individual, genomic sites cannot be treated as independent biological replicates; therefore, the results are presented as descriptive summaries rather than inferential statistical outcomes[61]. To quantify the magnitude of distributional differences, Cliff's delta (δ) was used as a non-parametric effect size measure appropriate for non-normal methylation data[58,60]. The observed values (5mC: $\delta = 0.94$; 6mA: $\delta = 0.48$) indicate strong separation for 5mC and moderate separation for 6mA between strand-specific distributions within this dataset. These values describe the magnitude of difference and are not interpreted as indicators of statistical significance or population-level inference. While strand-associated asymmetry is a known feature of vertebrate mitochondrial genomes, the observed differences may also reflect factors such as sequence context, coverage variation, and model-specific biases in Nanopore-based modification calling. Therefore, the reported patterns are best interpreted as dataset-specific descriptive observations without assigning a causal or mechanistic explanation.

Modification signal in an intergenic spacer

A short intergenic spacer was identified between the *tRNA-Asp* and *COXII* genes, characterized by the sequence AGCTCTATAGCTTA. Within this non-coding region, two cytosine residues showed detectable 5mC signals, located at positions 7139 and 7147 (1-based reference coordinates). These positions correspond to cytosines on opposite strands of the mitogenome. Interestingly, no 6mA signals were detected in this specific spacer. While most mitochondrial modifications are typically concentrated within protein-coding genes like *ND5* or ribosomal RNA genes such as *12S rRNA*[64–66], the detection of signals in intergenic regions highlights the high resolution of single-molecule sequencing approaches[58]. This observation is reported as a descriptive component of the *C. chaca* modification profile. Consistent with current standards in mitochondrial epigenomics, no functional or regulatory significance is attributed to these specific sites without further experimental validation of their biological role[60,61].

Because methylation detection was performed on a single individual, the analysis was designed as a descriptive characterization of modification signals within this mitogenome rather than a statistical comparison across biological replicates. No inference regarding population-level methylation patterns or functional regulation was intended. The methylated positions detected in this study can serve as a starting point for mitoepigenetic studies in *C. chaca* in the future (Table 1).

Conclusion

In this study, we present the first complete mitogenome of the square-head catfish, *C. chaca*, and provide an integrated description of its genomic features, including mtDNA modification signals. The assembled mitogenome exhibits the typical vertebrate mitochondrial gene complement and organization, serving as a valuable genomic resource for this poorly represented taxon and supporting future evolutionary and conservation-genetic studies. Phylogenetic analysis based on mitochondrial genome data places *C. chaca* within the pimelodoid lineage in the present dataset, contributing mitochondrial evidence relevant to its taxonomic position. Genome-wide analysis detected signals consistent with 5mC and 6mA modifications across the mitogenome. Differences in modification frequency between strands were observed after normalization for base composition within this dataset, with a stronger separation for 5mC compared to 6mA. These observations are derived from a single individual and are presented as a descriptive profile of mtDNA modification signals without inferring functional, regulatory, or mechanistic significance. Overall, this study provides a mitogenomic resource for *C. chaca* and reports preliminary observations of mtDNA modification patterns, which require further validation using independent methods and broader sampling to assess their biological relevance.

Data availability

The sequence data supporting the findings of this study are publicly available in the Zenodo repository at https://doi.org/10.5281/zenodo.18029752 under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. The complete mitochondrial genome sequence of C. chaca has also been deposited in GenBank under the accession number PX394296.

Received: 20 December 2025; Accepted: 20 April 2026
Published online: 24 April 2026

References

1. Zhang, C., Zhang, S., Tian, Z., Wang, D. & Xu, S. Identification of optimal codons and analysis of phylogenetic relation ship in *Osteochilus salsburyi* (Teleostei: Cypriniformes) based on complete mitogenome. *Front. Ecol. Evol.* **11** https://doi.org/10.3389/fevo.2023.1170927 (2023).
2. Miya, M. & Nishida, M. The mitogenomic contributions to molecular phylogenetics and evolution of fishes: A 15-year retrospect. *Ichthyol. Res.* **62** (1), 29–71. https://doi.org/10.1007/s10228-014-0440-9 (2014).
3. Zhang, W-Z. et al. Mitochondrial genome variation after hybridization and differences in the first and second generation hybrids of bream fishes. *PLoS ONE.* **11** (7). https://doi.org/10.1371/journal.pone.0158915 (2016).
4. Cheng, M. L. et al. The complete mitochondrial genome of the hybrid offspring *Epinephelus fuscoguttatus* ♀ × *Epinephelus tukula* ♂. *Mitochondrial DNA Part. B.* **4** (2), 2717–2718. https://doi.org/10.1080/23802359.2019.1644554 (2019).

5. Li, Z. et al. The complete mitochondrial genome of the hybrid offspring Epinephelus awoara♀ × Epinephelus tukula♂. *Mitochondrial DNA Part. B.* **5** (1), 1025–1026. https://doi.org/10.1080/23802359.2020.1721356 (2020).
6. Liu, K., Ma, H., Feng, X. & Xie, N. Complete mitochondrial genome of the hybrid of *Megalobrama terminalis* (♀) × *Culter alburnus* (♂). *Mitochondrial DNA Part. B.* **5** (1), 841–842. https://doi.org/10.1080/23802359.2020.1716638 (2020).
7. Zhang, X., Jiang, Q., Din, Y. & Zhu, C. The application of DNA barcodes in the authenticity identification of marine fishes. *Adv. J. Food Sci. Technol.* **11** (9), 605–610. https://doi.org/10.19026/ajfst.11.2735 (2016).
8. Das, S. P. et al. Low depth sequencing reveals the critically endangered *Batagur kachuga* (Red-crowned roofed turtle) mitochondrial genome and its evolutionary implications. *Gene* **927**, 148671. https://doi.org/10.1016/j.gene.2024.148671 (2024).
9. Brown, K. H. Fish mitochondrial genomics: Sequence, inheritance and functional variation. *J. Fish Biol.* **72** (2), 355–374. https://doi.org/10.1111/j.1095-8649.2007.01690.x (2008).
10. Sun, Y., Shen, Y., Irwin, D. M. & Zhang, Y. P. Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution p. 39–44 (Oxford University Press, 2010).
11. Miya, M. & Nishida, M. Use of mitogenomic information in teleostean molecular phylogenetics: A tree-based exploration under the maximum-parsimony optimality criterion. *Mol. Phylogenet. Evol.* **17** (3), 437–455. https://doi.org/10.1006/mpev.2000.0839 (2000).
12. Sahoo, L. et al. Population structure and genetic diversity of hatchery stocks as revealed by combined mtDNA fragment sequences in Indian major carp, *Catla catla*. *Mitochondrial DNA Part. A.* **30** (2), 289–295. https://doi.org/10.1080/24701394.2018.1484120 (2018).
13. Sun, P., Yang, Y., Yuan, M. & Tang, Q. CamITree: A streamlined software for phylogenetic analysis of viral and mitochondrial genomes. *BMC Bioinform.* **26** (1), 53. https://doi.org/10.1186/s12859-025-06034-2 (2025).
14. Naeem, Z., Masud, S., Hassan, S. & Naeem, M. Molecular approach for identification of *Catla catla* using mitochondrial CO1 from Pakistan. *Mitochondrial DNA Part. B.* **5** (3), 3000–3003. https://doi.org/10.1080/23802359.2020.1768913 (2020).
15. Du, W. et al. Equus mitochondrial pangenome reveals independent domestication imprints in donkeys and horses. *Sci. Rep.* **15** (1), 6803. https://doi.org/10.1038/s41598-025-91564-1 (2025).
16. Mahar, N. S., Satyam, R., Sundar, D. & Gupta, I. A systematic comparison of human mitochondrial genome assembly tools. *BMC Bioinform.* **24** (1), 341. https://doi.org/10.1186/s12859-023-05445-3 (2023).
17. Patil, M. P., Kim, J. O., Kim, Y-R., Yoon, S. J. & Kim, K. Complete mitogenome sequencing, annotation, and phylogeny of *Grateloupia turuturu*, a Red Alga with Intronic cox1 Gene. *Life* **13** (8), 1642. https://doi.org/10.3390/life13081642 (2023).
18. Yang, H. et al. Comparative characterization of the complete mitochondrial genomes of the three apple snails (Gastropoda: Ampullariidae) and the phylogenetic analyses. *Int. J. Mol. Sci.* **19** (11), 3646. https://doi.org/10.3390/ijms19113646 (2018).
19. Ahmed, S. et al. Pragmatic applications and universality of DNA barcoding for substantial organisms at species level: A review to explore a way forward. p. 1846485 (Hindawi Publishing Corporation, 2022).
20. Mu, X. et al. Mitochondrial DNA as effective molecular markers for the genetic variation and phylogeny of the family Osteoglossidae. *Gene* **511** (2), 320–325. https://doi.org/10.1016/j.gene.2012.09.087 (2012).
21. Amaral, F. R. et al. Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS ONE.* **10** (9). https://doi.org/10.1371/journal.pone.0138446 (2015).
22. Noh, J. K. E. Genetic differentiation of *Carassius auratus* and *C. cuvieri* by the cytochrome C Oxidase I gene analysis. *J. Aquaculture Res. Dev.* **3** (3). https://doi.org/10.4172/2155-9546.1000231 (2014).
23. Carmo, A. O. et al. Complete mitochondrial genomes from four species of the genus *Oxysarco dexia* (Sarcophagidae) with forensic entomology interest. *Mitochondrial DNA Part. B.* **4** (1), 430–431. https://doi.org/10.1080/23802359.2018.1545538 (2019).
24. Liu, X. et al. A multilocus DNA mini-barcode assay to identify twenty vertebrate wild life species. *iScience* **26** (11), 108275. https://doi.org/10.1016/j.isci.2023.108275 (2023).
25. Fernandes, T. J. R., Amaral, J. S. & Mafra, I. DNA barcode markers applied to seafood authentication: an updated review. Critical reviews in food science and nutrition: Taylor & Francis; pp. 3904-35. (2020).
26. Baltazar-Soares, M., Lima, A. R. A. & Silva, G. Targeted sequencing of mitochondrial genes reveals signatures of molecular adaptation in a nearly panmictic small pelagic fish species. *Genes* **12** (1), 91. https://doi.org/10.3390/genes12010091 (2021).
27. YuanYuan, L. et al. Applications of molecular markers in conserving endangered species. *Biodivers. Sci.* **28** (3), 367–375. https://doi.org/10.17520/biods.2019414 (2020).
28. Summarwar, S., Yadav, K. & Tailor, S. P. Genetic differentiation among Himalayan and local Mahseer populations. *Int. J. Fauna Biol. Stud.* **8** (1), 8–10. https://doi.org/10.22271/23940522.2021.v8.i1a.784 (2021).
29. Wong, L. L. et al. Genetic diversity and population structure of wild and domesticated black tiger shrimp (*Penaeus monodon*) broodstocks in the Indo-Pacific regions using consolidated mtDNA and microsatellite markers. *Gene Rep.* **23**, 101047. https://doi.org/10.1016/j.genrep.2021.101047 (2021).
30. Bein, B. et al. Long-read sequencing and genome assembly of natural history collection samples and challenging specimens. (Cold Spring Harbor Laboratory). (2024). https://doi.org/10.1101/2024.03.04.583385
31. Iwasaki, W. et al. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.* **30** (11), 2531–2540 (2013).
32. Tonini, J. F. R., Moore, A., Stern, D., Shcheglovitova, M. & Ortí, G. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLoS Curr.* https://doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be (2015).
33. Dettaı, A. et al. Correction: Conveniently pre-tagged and pre-packaged: extended molecular identification and metagenomics using complete metazoan mitochondrial genomes. *PLoS ONE.* **8** https://doi.org/10.1371/annotation/020848d7-0453-42df-bf94-40eb39f1b27a (2013).
34. Baeza, J. A. Yes, we can use it: a formal test on the accuracy of low-pass nanopore long-read sequencing for mitophylogenomics and barcoding research using the Caribbean spiny lobster *Panulirus argus*. *BMC Genom.* **21** https://doi.org/10.1186/s12864-020-07292-5 (2020).
35. Sahoo, B. et al. The complete mitochondrial genome of *Labeo catla* (Hamilton, 1822) using long read sequencing. *Mitochondrial DNA Part. B.* **6**, 402–403. https://doi.org/10.1080/23802359.2020.1870879 (2021).
36. Cavalcante, R. et al. The complete mitogenome of Amazonian *Brachyplatystoma filamentosum* and the evolutionary history of body size in the order Siluriformes. *Sci. Rep.* **15** https://doi.org/10.1038/s41598-025-94272-y (2025).
37. Muhala, V. et al. Comparative mitochondrial genome brings insights to slight variation in gene proportion and large intergenic spacer and phylogenetic relationship of mudskipper species. *Res. Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-3207876/v1 (2023).
38. Muhala, V. et al. Comparative mitochondrial genome brings insights to slight variation in gene proportion and large intergenic spacer and phylogenetic relationship of mudskipper species. *Sci. Rep.* **14** https://doi.org/10.1038/s41598-024-52979-4 (2024).
39. He, W. et al. Three mitochondrial markers reveal genetic diversity and structure of rock carp (*Procypris rabaudi*) Endemic to the Upper Yangtze: Implications for pre-release genetic assessment. *Front. Mar. Sci.* **9** https://doi.org/10.3389/fmars.2022.939745 (2022).
40. Winata, V. A., Kasitowati, R. D., Iranawati, F., Pranowo, W. S. & Sartimbul, A. Molecular and phylogenetic analysis of Sardinella lemuru Bleeker 1835 at fishing ground Canggu-Bali inferred D-loop mutations of mtDNA p. 12065 (IOP Publishing, 2022).
41. Cucini, C. et al. EZmito: a simple and fast tool for multiple mitogenome analyses. *Mitochondrial DNA Part. B.* **6** (3), 1101–1109. https://doi.org/10.1080/23802359.2021.1899865 (2021).

42. Meng, G., Li, Y., Yang, C. & Liu, S. MitoZ: A toolkit for mitochondrial genome assembly, annotation and visualization. bioRxiv (Cold Spring Harbor Laboratory). (2018). https://doi.org/10.1101/489955
43. Ng, H. H. & Kottelat, M. *Chaca serica*, a new species of frogmouth catfish (Teleostei: Siluriformes) from southern Borneo. *Zootaxa* **3258** https://doi.org/10.11646/zootaxa.3258.1.3 (2012).
44. Diogo, R., Chardon, M. & Vandewalle, P. On the osteology and myology of the cephalic region and pectoral girdle of *Chaca bankanensis* Bleeker 1852, with comments on the autapomorphies and phylogenetic relationships of the Chacidae (Teleostei: Siluriformes). *Anim. Biology*. **54**, 159–174. https://doi.org/10.1163/1570756041445227 (2004).
45. Teugels, G. G. Taxonomy, phylogeny and biogeography of catfishes (Ostariophysi, Siluroidei): An overview. *Aquat. Living. Resour.* **9**, 9–34. https://doi.org/10.1051/alr:1996039 (1996).
46. Hegg, J. C., Giarrizzo, T. & Kennedy, B. P. Diverse early life-history strategies in migratory amazonian catfish: Implications for conservation and management. *PLoS ONE*. **10** https://doi.org/10.1371/journal.pone.0129697 (2015).
47. Moreno-Árias, C., López-Casas, S., Rogéliz, C. A. & Jiménez-Segura, L. F. Protection of spawning habitat for potamodromous fish, an urgent need for the hydropower planning in the Andes. *Neotropical Ichthyol.* **19** https://doi.org/10.1590/1982-0224-2021-0027 (2021).
48. Mistri, A., Kumari, U., Mittal, S. & Mittal, A. Morphological specializations of the epidermis of an angler catfish *Chaca chaca* (Siluriformes, Chacidae) in relation to its ecological niche: A scanning electron microscopic investigation. *Microsc. Res. Tech.* **81**, 439–448. https://doi.org/10.1002/jemt.22996 (2018).
49. Roberts, T. R. A revision of the south and southeast Asian angler-catfishes (chacidae). *Copeia* **1982** (4), 895. https://doi.org/10.2307/1444100 (1982).
50. Sholihah, A. et al. Limited dispersal and in situ diversification drive the evolutionary history of rasborinae fishes in sundaland. *J. Biogeogr.* **48** (9), 2153–2173. https://doi.org/10.1111/jbi.14141 (2021).
51. Jeon Hb, Kim, D. Y., Lee, Y. J. & Bae Hg, Suk, H. Y. The genetic structure of *Squalidus multimaculatus* revealing the historical pattern of serial colonization on the tip of East Asian continent. *Sci. Rep.* **8** https://doi.org/10.1038/s41598-018-28340-x (2018).
52. Kappas, I., Vittas, S., Pantzartzi, C. N., Drosopoulou, E. & Scouras, Z. G. A time-calibrated mitogenome phylogeny of catfish (teleostei: Siluriformes). *PLoS One*. **11** (12), e0166988. https://doi.org/10.1371/journal.pone.0166988 (2016).
53. Sullivan, J. P., Lundberg, J. G. & Hardman, M. A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. *Mol. Phylogenet. Evol.* **41**, 636–662. https://doi.org/10.1016/j.ympev.2006.05.044 (2006).
54. Fonseca, M. M., Harris, D. J. & Posada, D. The inversion of the control region in three mitogenomes provides further evidence for an asymmetric model of vertebrate mtDNA replication. *PLoS ONE*. **9** (9). https://doi.org/10.1371/journal.pone.0106654 (2014).
55. Wei, S. J. et al. New views on strand asymmetry in insect mitochondrial genomes. *PLoS ONE*. **5** (9). https://doi.org/10.1371/journal.pone.0012708 (2010).
56. Eberhard, J. R. & Wright, T. F. Rearrangement and evolution of mitochondrial genomes in parrots. *Mol. Phylogenet. Evol.* **94**, 34–46. https://doi.org/10.1016/j.ympev.2015.08.011 (2015).
57. Gomes-dos-Santos, A. et al. Mitochondrial replication's role in vertebrate mtDNA strand asymmetry. *Open. Biol.* **13** (12). https://doi.org/10.1098/rsob.230181 (2023).
58. Lüth, T. et al. Nanopore single-molecule sequencing for mitochondrial DNA methylation analysis: Investigating parkin-associated parkinsonism as a proof of concept. *Front. Aging Neurosci.* **13** https://doi.org/10.3389/fnagi.2021.713084 (2021).
59. Faulk, C. Genome skimming with nanopore sequencing precisely determines global and transposon DNA methylation in vertebrates. *Genome Res.* **33** (6), 948–956. https://doi.org/10.1101/gr.277743.123 (2023).
60. Wanner, N. M., Larsen, P. A., McLain, A. T. & Faulk, C. The mitochondrial genome and epigenome of the golden lion tamarin from fecal DNA using Nanopore adaptive sequencing. *BMC Genom.* **22** (1). https://doi.org/10.1186/s12864-021-08046-7 (2021).
61. Goldsmith, C. et al. Low biological fluctuation of mitochondrial CpG and non-CpG methylation at the single-molecule level. *Sci. Rep.* **11** (1). https://doi.org/10.1038/s41598-021-87457-8 (2021).
62. Morris, M. J. et al. Reduced nuclear DNA methylation and mitochondrial transcript changes in adenomas do not associate with mtDNA methylation. *Biomark. Res.* **6** (1). https://doi.org/10.1186/s40364-018-0151-x (2018).
63. Low, H. C. et al. Changes in mitochondrial epigenome in type 2 diabetes mellitus. *Br. J. Biomed. Sci.* **80** https://doi.org/10.3389/bjbs.2023.10884 (2023).
64. Devall, M. et al. Genome-wide characterization of mitochondrial DNA methylation in human brain. *Front. Endocrinol.* **13** https://doi.org/10.3389/fendo.2022.1059120 (2023).
65. Iacobazzi, V., Castegna, A., Infantino, V. & Andria, G. Mitochondrial DNA methylation as a next-generation biomarker and diagnostic tool. *Mol. Genet. Metab.* **110**, 25–34. https://doi.org/10.1016/j.ymgme.2013.07.012 (2013).
66. Liu, B. et al. CpG methylation patterns of human mitochondrial DNA. *Sci. Rep.* **6** (1). https://doi.org/10.1038/srep23421 (2016).

Acknowledgements

The authors thank the org.one project from Oxford Nanopore Technology, UK for their support of sequencing reagents for the data generation of the research work. We express our gratitude to the Honorable Vice-Chancellor, Rani Lakshmi Bai Central Agricultural University, Jhansi, India, for providing the infrastructure for research.

Author contributions

PST conceived and designed the study, performed the analyses, and wrote the original manuscript. LD contributed resources and assisted with data curation and manuscript review. SNP contributed to methodology development, validation, and manuscript editing. AKR performed data curation, software-based analyses, visualization, and assisted with manuscript revision. JP contributed to investigation, supervision, and project administration. BKB and PKP supervised the study, contributed to conceptualization and methodology, and critically reviewed and edited the manuscript. All authors reviewed and approved the final manuscript.

Funding

declaration.

The present work received no specific grant from any funding agency in the public, commercial, or non-profit sectors.

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/s41598-026-50124-x.

Correspondence and requests for materials should be addressed to A.K.R., B.K.B. or P.K.P.

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

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