Integrative Zoology

WILEY

ISZS International Society of Zoological Sciences

LETTER TO THE EDITOR OPEN ACCESS

Refining the Vertebrate Mitochondrial 12S rRNA Secondary Structure by Comparative Analysis

Fengxia Li[1,3] | Gengyun Niu[2] | Renzeng Du[1,4] | Kun Wan[1,5] | David Roy Smith[6] | Meicai Wei[2] | Haihe Shi[1]

[1]School of Artificial Intelligence, Jiangxi Normal University, Nanchang, China | [2]School of Mathematics and Intelligence, Nanchang Vocational University, Nanchang, China | [3]College of Life Sciences, Jiangxi Normal University, Nanchang, China | [4]School of Software, Nanchang Hangkong University, Nanchang, China | [5]School of Economic Management, Jiangxi Youth Vocational College, Nanchang, China | [6]Department of Biology, Western University, London, Canada

Correspondence: Meicai Wei (weim@jxnu.edu.cn) | Haihe Shi (haiheshi@jxnu.edu.cn)

Received: 5 April 2025 | **Revised:** 6 August 2025 | **Accepted:** 3 October 2025

Keywords: analytical framework | comparative genomics | evolutionary dynamics | mitochondrial 12S rRNA | secondary structure prediction | vertebrate

1 | Introduction

Ribosomal RNA (rRNA) is ubiquitous across all extant life forms. Its functional characteristics are largely due to its structure (Vicens and Kieft 2022). The rRNAs found and encoded within mitochondria (mt-rRNA) often have a reduced and streamlined structure, providing an excellent framework to study rRNAs (Ganser et al. 2019; Ramrath et al. 2018). Metazoan mt-RNAs, despite their architectural reduction and variation in size (368–1833 bp) (Pett et al. 2011; Lukić-Bilela et al. 2008), maintain operational integrity while still demonstrating pronounced diversity and plasticity (Noller et al. 2022; Petrov et al. 2019). Understanding the structural basis of this functional diversity demands precise structural predictions, which at present involve fundamental methodological challenges across biological scales.

Modeling RNA structural diversity typically relies on complementary approaches. Experimentally derived models, although regarded as the "gold standard," are applicable only to a limited set of model organisms (Zhang, Li et al. 2024). For the overwhelming majority of species, computational methods provide the sole means of structural inference and are typically achieved through two strategies. The first focuses on detailed comparisons within phylogenetically constrained groups, achieving high resolution for specific RNA families (De Los Monteros 2003; Mallatt et al. 2010; Page 2000; Satoh et al. 2016; Xie et al. 2008). Such taxonomically restricted analyses, however, can overlook structural patterns conserved across broader evolutionary scales. An alternative strategy establishes large-scale RNA databases to systematically investigate structural–functional correlations across diverse lineages (Andrews et al. 2017; Andronescu et al. 2008; Cannone et al. 2002; Kalvari et al. 2021; Szymanski et al. 2016). This top-down approach can reveal pan-taxonomic structural patterns, but its coarse granularity can also obscure lineage-specific features, creating resolution trade-offs between phylogenetic breadth and structural detail. This micro–macro dichotomy reflects the fundamental challenge of reconciling universal structural principles with taxon-specific adaptations in RNA modeling.

Although AI-driven computational advances have mitigated certain limitations, they introduce new methodological complexities, such as deep-learning-based methods, which commonly suffer from overfitting, especially when dealing with long RNA sequences (Zhao et al. 2021). The parameter space of these models expands with sequence length, whereas available structural data remain sparse (Bugnon et al. 2022; Zhang, Lang et al. 2024). In this context, homology comparisons (HCs), which leverage evolutionary principles and covariation signals, continue to stand out as the most robust approach for RNA secondary structure prediction (Gutell et al. 2002; Gutell 2015). HCs operate under the empirically validated principle that structural constraints drive compensatory mutations in evolu-

Fengxia Li and Gengyun Niu contributed equally to this work.

tionarily related RNAs, enabling detection of conserved motifs through covariation analysis (Morandi et al. 2022; Nawrocki and Eddy 2013). However, HC accuracy depends critically on phylogenetic sampling: Insufficient taxonomic coverage reduces signal detection, while overly conserved sequences prevent covariation resolution, both of which can introduce systematic errors.

Massive, publicly available genomic datasets now enable HC applications at unprecedented scales, particularly when integrated with robust phylogenomic frameworks (Blair 2023). Paradoxically, the very data that empower HC analyses also complicate their execution: An overwhelming array of specialized tools with varying limitations forces researchers into time-consuming trial-and-error evaluations, coupled with non-reproducible analytical steps that hinder result verification. These compounding difficulties highlight the critical demand for implementing standardized workflow systems with transparent operational protocols (Lott et al. 2017).

Here, we address these limitations by integrating HC within evolutionary frameworks to refine long RNA structures on a large scale. Using vertebrate 12S mt-rRNAs as a model system, we have predicted 21 058 secondary structures and classified them into nine structurally conserved types. Our study provides three major advancements: (1) It reveals the dynamic balance between structural conservation and variation throughout vertebrate evolution from the perspective of mitochondrial ribosomes, giving new evidence for the molecular mechanisms underlying gene streamlining and functional maintenance; (2) it develops a standardized bioinformatics workflow that strictly regulates data acquisition, covariation analysis, and phylogenetic modeling steps, ensuring the reproducibility and scalability of the research process; and (3) it presents an HC paradigm that can be seamlessly extended to the entire biological realm, laying a methodological foundation for decoding the complex relationships between rRNA structure and function, and provides a structural database to support the mechanistic elucidation of disease-associated mutations. This study not only advances the innovation of long RNA structural prediction technologies but also opens new avenues for systematically exploring the functional realization of non-coding RNAs from an evolutionary perspective.

2 | Materials and Methods

2.1 | Data Acquisition and Standardization

We retrieved 27 772 vertebrate mitochondrial genome sequences from the NCBI Genome Database (retrieval detailed in Supporting Information S1). Data cleaning and reannotation (n = 5593) were performed using the NCBI2GO (unpublished) pipeline, including taxonomic reconciliation. Structured metadata tables (Supporting Information S2) were constructed by parsing the LOCUS, SOURCE, FEATURES, and ORIGIN fields from GenBank files, incorporating *rrnS* genomic coordinate mapping and taxonomic hierarchy, ultimately yielding 22 179 standardized records.

2.2 | Precision Extraction of Mitochondrial 12S rRNA Sequences

Gene boundary ambiguity remains a persistent challenge in mitochondrial genome annotation. As illustrated in Figure 1, metadata frequently contain inconsistent annotations, including erroneous start/end coordinates and incomplete 5′ or 3′ termini, particularly for sequences flanked by the control region (CR) or Mitochondrial 16S rRNA gene. To address this, we implemented a four-tier validation protocol for mitochondrial 12S rRNA (mt 12S rRNA) demarcation: (1) verification of co-localization signatures with tRNA-Phe (F) and tRNA-Val (V); (2) cross-validation between MITOS (Bernt et al. 2013) and GenBank annotations; (3) taxon-specific length thresholds (≥900 bp for teleosts, ≥800 bp for other vertebrates); and (4) exclusion of CR-proximal sequences. Following manual inspection and exclusion of 279 anomalous records, we extracted 21 900 high-confidence *rrnS* sequences. The genomic coordinates (start/end positions) of these sequences were documented in the "Location" column of the master table, with corresponding sequences programmatically extracted from ORIGIN fields and recorded in the "Sequence" column. Subsequent filtering retained sequences within 800–1300 bp, and duplicate entries were removed based on sequence identity (with unique identifiers assigned in the "Seq_Type" column), yielding a non-redundant dataset of 14 048 sequences for downstream analyses.

2.3 | Secondary Structure Prediction and Refinement

Mitochondrial 12S rRNA secondary structures were predicted using an automated Python pipeline leveraging SSU-align (v0.1.1) under default parameters. The workflow implemented batch processing through phylogenetically stratified iterative template refinement, employing default covariance models configured with a gap retention threshold of 0.8 (-gapthresh 0.8) and disabled entropy weighting (-enone). Structural outputs were standardized through bracket notation conversion and gap removal procedures, followed by compilation of predictions from Stockholm-format alignments (.stk) into unified structural files. The pipeline source code is publicly available at https://github.com/lifengxia0110-lab/ssu-predict12S. Initial template derivation combined mammalian consensus structures from the CRW database. A multi-sequence alignment consensus template was constructed using MARNA (Siebert and Backofen 2005) under parameters: gap penalty = 5, conservation weight = 0.8, with conflict resolution via majority voting. For results that deviate from the consensus template, elongated sequences (>1100 bp) were reanalyzed using avian templates derived from structures in CRW, and shortened sequences (<850 bp) utilized templates generated from manually curated homologs. Quality assessment was conducted by referencing both the sequence length and the structural and domain segmentation results provided by bpRNA (Danaee et al. 2018). Predicted structures failing quality criteria underwent reprediction based on branch-specific templates, followed by manual validation through the Forna (Kerpedjiev et al. 2015) platform (http://rna.tbi.univie.ac.at/forna/) for structural visualization using default parameters (Color Scheme = "Struc-

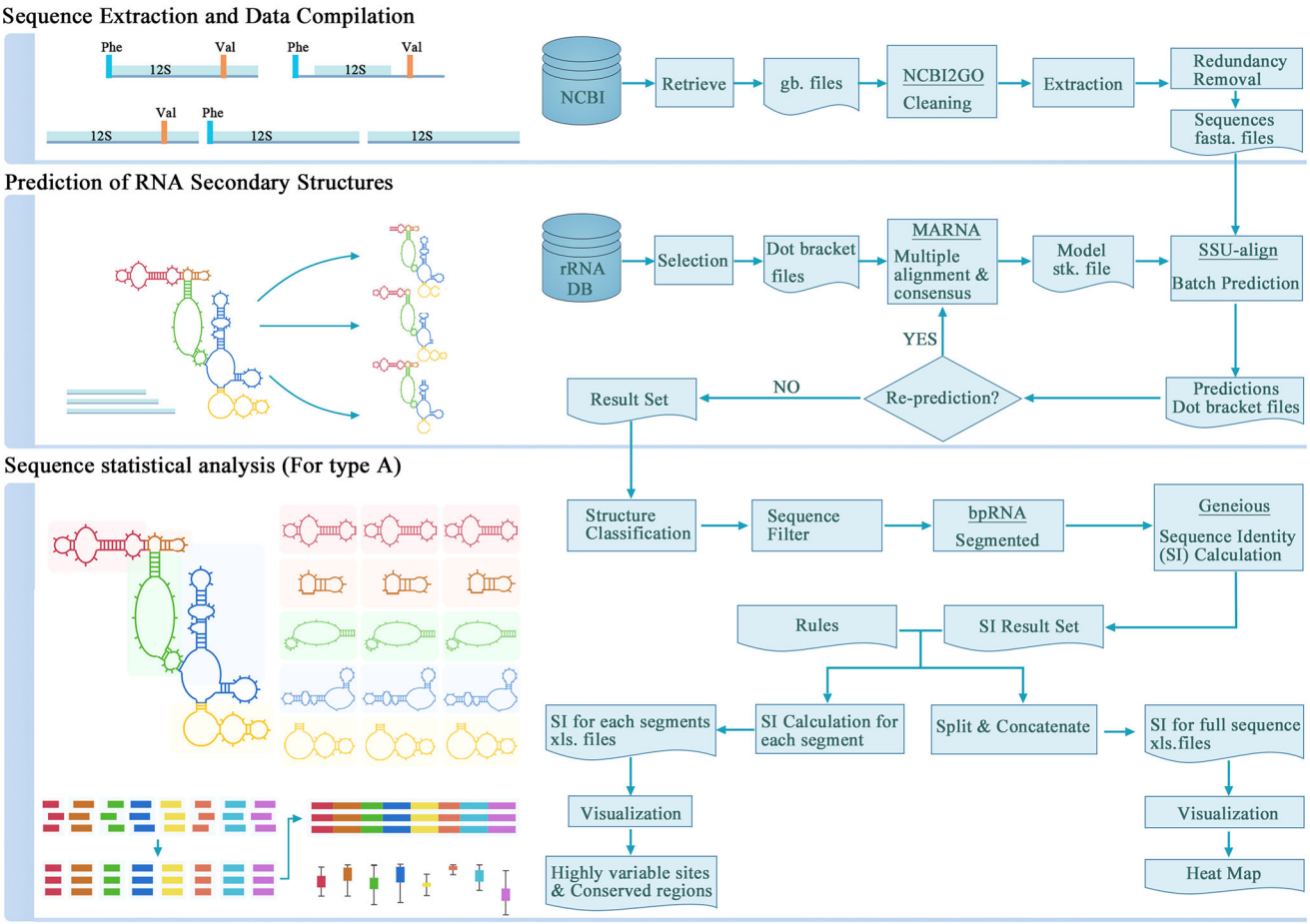

FIGURE 1 | Simplified workflow for multi-tool data processing. This flowchart outlines the comprehensive workflow of the study, integrating diverse bioinformatics tools (e.g., NCBI2GO, SSU-align, bpRNA) and their sequential interactions. Key steps, such as data preprocessing, structural prediction, and evolutionary analysis, are depicted with their respective outputs (e.g., standardized records, consensus templates, and domain annotations) listed on the right, connected via directional arrows to illustrate data flow.

ture"). All final results were converted to dot–bracket notation to standardize RNA secondary structures (structures detailed in Supporting Information S3).

2.4 | Comparative Analysis of Secondary Structures

Prediction results were subjected to quality control with classification based on the following criteria: **no12S**: failure to extract target sequences, precluding prediction attempts; **too short**: sequences extracted but below length quality thresholds, excluded from prediction; **fail**: sequences extracted with adequate length but prediction failure during processing; **Too many "N"**: sequences extracted with sufficient length but containing excessive undetermined bases ("N"), rendering predictions unreliable; **shorten str**: sequences extracted with valid length but yielding structurally incomplete predictions (insufficient domain counts). Structures passing quality control were classified into nine types (A–I) based on structural features.

To minimize structural heterogeneity impacts on evolutionary analyses, only Type A structures (n = 20 116) were retained for downstream investigations, with 12 aberrantly long sequences excluded as outliers due to sparse representation. The mt 12S rRNA sequences were decomposed into 52 functional domains (31 stem regions, 21 loop regions) using bpRNA, generating annotation files specifying structural elements including hairpin loops, bulges, and multi-branched loops. Sequence fragments corresponding to stem/internal loop regions (Segment, n = 31) and hairpin loops (H, n = 21) were systematically extracted. Domain sequences were categorized by taxonomic groups (figshare: 5_Segmented_txt) and aligned using Geneious to generate taxon-specific structural alignments (figshare: 6_Alignment_geneious). Consistency indices were calculated for 83 substructures, producing two analytical frameworks: (1) a site-level consistency matrix representing 5′→3′ concatenated positions (Supporting Information S4), and (2) domain-level consistency profiles. Heatmaps and boxplots generated in GraphPad Prism v8.0.2 employed dark gray (≥95% consistency) and light gray (≥85%) thresholds, with substructures exhibiting <85% consistency across all taxa designated as hypervariable regions.

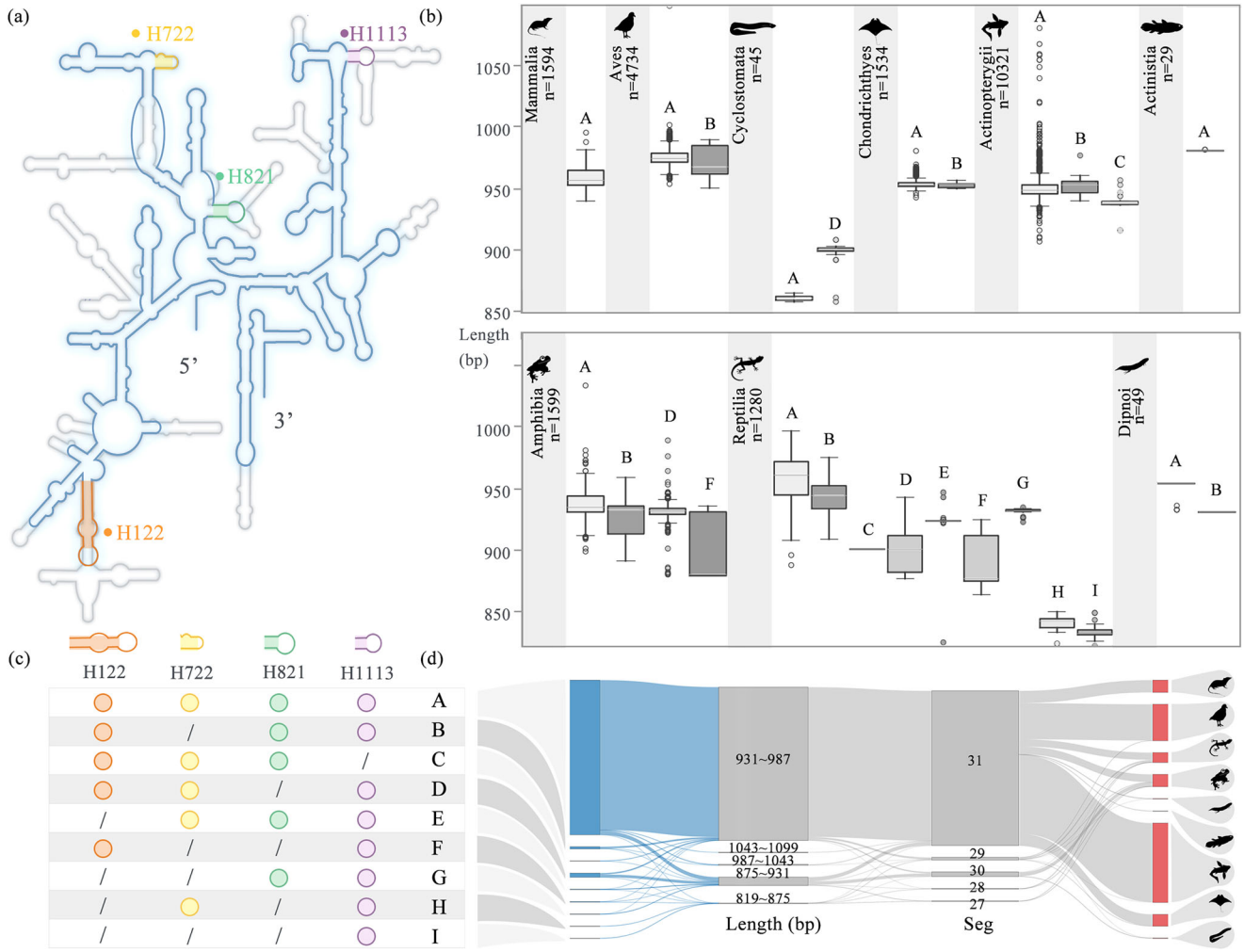

FIGURE 2 | Vertebrate mitochondrial 12S rRNA model and structural diversity. (a) Consensus secondary structure of vertebrate mt 12S rRNA mapped onto *Escherichia coli* SSU rRNA (gray skeleton; J01695).Conserved domains (dark blue) and variable regions (colored) follow *E. coli* stem-loop numbering. (b) Distribution of nine structural types (A–I) across nine major vertebrate groups. Box colors correspond to structural types and the *y*-axis represents sequence length (bp). (c) Key characteristics of structural types A–I. (d) Sankey diagram quantifying relationships, from left to right: structural types, sequence length, number of segments yielded by bpRNA, and vertebrate groups.

3 | Results and Discussion

3.1 | Refined Models and Structural Diversity

Previous studies have established foundational insights into the secondary structural features of mt 12S rRNA in birds and fish. For example, avian mt 12S rRNA exhibits 40 stems, 20 internal loops, and 17 external loops across four canonical domains, with 423 invariant nucleotide positions, which highlights the profound level of evolutionary conservation within this group (De Los Monteros 2003). In contrast, piscine mt 12S rRNA displays greater structural complexity, featuring 43 stems and elevated nucleotide conservation in core pairing regions (Wang and Lee 2002). While both groups share conserved functional domains and critical nucleotide positions, key differences emerge: Birds exhibit hypervariable regions enriched with insertions, potentially linked to environmental adaptation, whereas fish demonstrate diversified stem architectures, possibly reflecting ecological niche diversity. However, methodological limitations—such as alignment biases, model oversimplifications, and incomplete taxonomic sampling—constrain the resolution of structural variability, particularly in hypervariable regions and recombination-prone zones. These discrepancies underscore the need for refined analytical frameworks to disentangle lineage-specific structural dynamics.

Through systematic data cleaning and reannotation, this study screened the raw sequences of 27 772 vertebrate mitochondrial genomes and obtained 22 179 standardized records. After a four-tier validation system (gene colocalization, annotation cross-validation, length threshold filtering, and exclusion of control region-proximal sequences) and manual verification, 21 900 high-confidence mt 12S rRNA sequences were retained. Further length filtering (800–1300 bp) and sequence identity–based deduplication generated a non-redundant dataset of 14 048 *rrnS* sequences, covering major vertebrate groups. Notably, reptiles and amphibians exhibited the highest filtration rates (both exceeding 10%), while birds, ray-finned fish, and cartilaginous fish showed filtration rates below 3%, primarily due to sequence extraction failures (4.7%) or length thresholds (e.g., <900 bp for

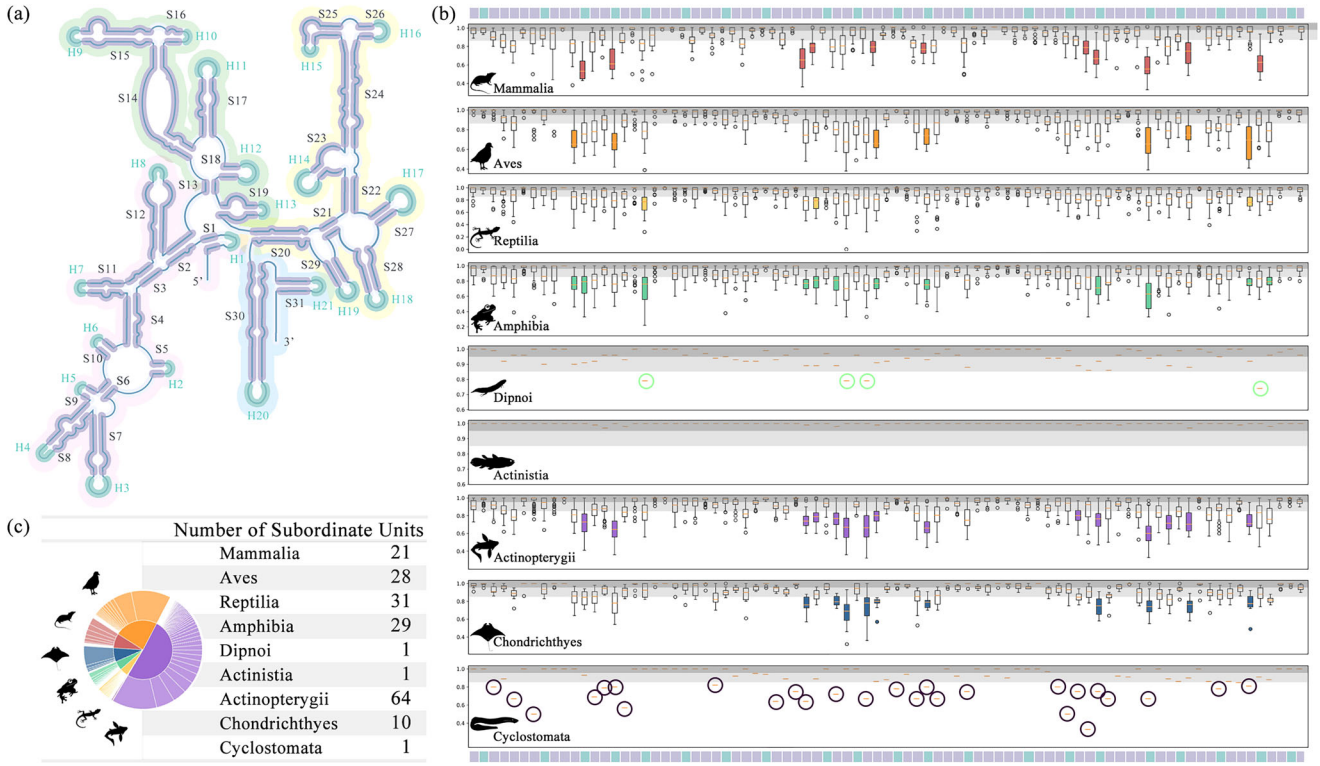

FIGURE 3 | Structural dissection and phylogenetic consistency of Type A. (a) Domain segmentation by bpRNA: The Type A structure is split into 83 substructures, with 21 hairpin loops (green) and 31 paired stem segments (blue-purple), each analyzed independently. (b) Consistency index across taxonomic groups: Sequence datasets grouped by family were aligned to calculate consistency indices. Boxplots summarize these indices across orders within each vertebrate class, highlighting conserved and variable regions. (c) Taxonomic representation: A pie chart quantifies the number of sequences from major taxonomic groups included in the analysis.

teleosts and <800 bp for other groups; Supporting Information S2).

Using a phylogenetically stratified iterative template refinement strategy, 21 185 *rrnS* secondary structures were successfully predicted after quality control based on five criteria: absence of 12S, insufficient length, prediction failure, excessive "N" content, and structural incompleteness. These structures were classified into nine types (A–I) based on domain features (Figure 2). Type A (95% of total) matched the canonical human mt 12S rRNA architecture, while types B–I exhibited independent or combinatorial losses of specific stem regions (H122, H722, H821, and H1113). Specifically, Type B lacked H722, Type C lacked H1113, Type D lacked H821, Type E lacked H122, Type F lacked H821+H722, Type G lacked H122+H722, Type H lacked H122+H821, and Type I lacked three stems. Type D (n = 556) and Type B (n = 291) were the most frequent non-A types; the other types ranged from 9 to 76 sequences, indicating sporadic occurrence and specificity to narrow phylogenetic branches. Taxonomic distribution analysis revealed that Type A was universally present across vertebrates, with mammals and coelacanths exclusively retaining this type. Reptiles (n = 1580) and amphibians (n = 1777) exhibited the highest structural diversity. Type C was restricted to Actinopterygii; Type D occurred in amphibians, reptiles, and cyclostomes; Types E and G were reptile-specific; and Type F was observed in amphibians and reptiles (Supporting Information S5). Furthermore, this study identified 12 Type A sequences with exceptional lengths (1020–1099 bp). Given their sparse representation, they were not classified as a separate subgroup. However, as their variations were confined to hypervariable regions, future studies should investigate their structural characteristics within a denser phylogenetic samplong framework.

3.2 | Evolutionary Insights from Structural Diversity

When mt 12S rRNA sequences were aligned within a taxonomic framework, conserved regions, variable regions, and phylogenetically informative sites became discernible. Since only mammalian mt 12S rRNA structures strictly conformed to Type A, sequences were partitioned into distinct regions based on structural features. Multi-sequence alignments of color-coded regions and taxonomic group-specific identity percentages were visualized as heatmaps, with full sequence identity values provided in Supporting Information S6.

Consistency indices for 83 substructures were calculated to construct an evolutionary structural map of vertebrate mitochondrial rRNA (consistency detailed in Supporting Information S7 and Figure 3). Site-level analysis revealed that the red region comprised 5′ segments (seg1–seg12) and hairpin loops (H1–H8). Central domain helices (H1, H6, H8, seg6, seg9) showed ≥95% consistency across all taxa, indicating their role as highly conserved core functional modules. Hypervariable sites clustered in three regions: mammalian-specific sites (30–40, 60–67, 80–90,

130–140, 160–170, and 230–240); the green region (seg13–seg19 and H9–H13) included hypervariable sites at 300–310, 330–340, 360–370, 380–390, and 470–480; the yellow region (seg20–seg29 and H14–H19) spanned 580–590, 640–650, and 750–770; and the blue region (seg30–seg31 and H20–H21) exhibited hypervariability at 855–865 and 885–895. Similar hypervariable site distributions across taxa suggest tight coupling between RNA sequence regions and structural stability, implying that structural variations may directly influence functional constraints.

3.3 | Technical Limitations and Annotation Biases

When separating the mt 12S rRNA structures into 52 functional domains (31 stems, 21 loops) using bpRNA, inherent limitations arose: The tool only recognizes multibranch loops and hairpin loops, while internal loops were merged into adjacent stems. Manual validation confirmed that Type A structures comprised 31 stems, 21 hairpin loops, and 42 multibranch loops. Additionally, bpRNA's de novo numbering system hindered direct HC. For example, types B–H shared 29 stems but differed in domain configurations. Structural predictions based solely on sequence length were unreliable (Figure 2d).

Unexpected prediction errors were disproportionately high in amphibians and reptiles. For instance, H9 stem loss in Viperidae, Testudinidae, Agamidae, and Crocodylidae correlated with systematic annotation biases. Re-annotation revealed a highly conserved tRNA-Phe (F) sequence 300 bp upstream of the misannotated mt 12S rRNA 5′ end (identity ≥98%), indicating erroneous boundary assignments. By analyzing the entire genomic region encompassing both F and mt 12S rRNA, rather than relying solely on flawed rRNA annotations, the secondary structure prediction pipeline successfully recovers missing stem-loop domains (e.g., H9).

To resolve sequencing ambiguities and mitigate annotation-driven prediction failures, we recommend adopting long-read sequencing (e.g., PacBio HiFi) to generate high-quality data. Quality control pipelines should then integrate structural conservation checks (e.g., H1–H8 stem verification) and tRNA/rRNA colocalization analysis. Furthermore, mitochondrial genome annotation standards should incorporate secondary structure parameters (e.g., stem-loop topology, pseudoknot density, and non-canonical base pairs). Structured metadata from these analyses could enable multidimensional insights into gene boundary definition, adaptive evolution, and disease-related mutations.

Acknowledgments

We sincerely thank Professor Xie Qiang of Sun Yat-sen University for the invaluable insights he generously shared during the conceptualization of this research. We also thank Professor Dennis V. Lavrov of Iowa State University for his steadfast encouragement throughout this project. We are grateful to all team members for their dedicated work, particularly to colleagues at the Lab of Insect Systematics and Evolutionary Biology (LISEB), Jiangxi Normal University for providing technical expertise. We also acknowledge the anonymous reviewers for their thoughtful critiques that strengthened this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The raw data utilized in this study were sourced from the publicly accessible GenBank database (https://www.ncbi.nlm.nih.gov/genbank/), with specific retrieval commands detailed in Supporting Information S1. All processed data files, including preprocessed datasets, statistical analysis outputs, and associated metadata, have been deposited in the figshare scientific data repository (https://figshare.com/projects/Vertebrate-12S/235928). These curated datasets are permanently archived and publicly available to support research reproducibility.

Fengxia Li
Gengyun Niu
Renzeng Du
Kun Wan
David Roy Smith
Meicai Wei
Haihe Shi

References

Andrews, R. J., L. Baber, and W. N. Moss. 2017. "RNAStructuromeDB: A Genome-Wide Database for RNA Structural Inference." *Scientific Reports* 7: 17269.

Andronescu, M., V. Bereg, H. H. Hoos, and A. Condon. 2008. "RNA STRAND: The RNA Secondary Structure and Statistical Analysis Database." *BMC Bioinformatics [Electronic Resource]* 9: 340.

Bernt, M., A. Braband, B. Schierwater, and P. F. Stadler. 2013. "MITOS: Improved De Novo Metazoan Mitochondrial Genome Annotation." *Molecular Phylogenetics and Evolution* 69: 313–319.

Blair, C. 2023. "Organellar DNA Continues to Provide a Rich Source of Information in the Genomics Era." *Molecular Ecology* 32: 2144–2150.

Bugnon, L. A., A. A. Edera, S. Prochetto, et al. 2022. "Secondary Structure Prediction of Long Noncoding RNA: Review and Experimental Comparison of Existing Approaches." *Briefings in Bioinformatics* 23: bbac205.

Cannone, J. J., S. Subramanian, M. N. Schnare, et al. 2002. "The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and Other RNAs." *BMC Bioinformatics [Electronic Resource]* 3: 2.

Danaee, P., M. Rouches, M. Wiley, et al. 2018. "bpRNA: Large-Scale Automated Annotation and Analysis of RNA Secondary Structure." *Nucleic Acids Research* 46: 5381–5394.

De Los Monteros, A. E. 2003. "Models of the Primary and Secondary Structure for the 12S rRNA of Birds: A Guideline for Sequence Alignment." *DNA Sequence* 14: 241–256.

Ganser, L. R., M. L. Kelly, D. Herschlag, and H. M. Al-Hashimi. 2019. "The Roles of Structural Dynamics in the Cellular Functions of RNAs." *Nature Reviews Molecular Cell Biology* 20: 474–489.

Gutell, R. R. 2015. "rRNA—The Evolution of That Magic Molecule." *RNA* 21: 627–629.

Gutell, R. R., J. C. Lee, and J. J. Cannone. 2002. "The Accuracy of Ribosomal RNA Comparative Structure Models." *Current Opinion in Structural Biology* 12: 301–310.

Kalvari, I., E. P. Nawrocki, N. Ontiveros-Palacios, et al. 2021. "Rfam 14: Expanded Coverage of Metagenomic, Viral and MicroRNA Families." *Nucleic Acids Research* 49: D192–D200.

Kerpedjiev, P., S. Hammer, and I. L. Hofacker. 2015. "Forna (Force-Directed RNA): Simple and Effective Online RNA Secondary Structure Diagrams." *Bioinformatics* 31: 3377–3379.

Lott, S. C., M. Wolfien, K. Riege, et al. 2017. "Customized Workflow Development and Data Modularization Concepts for RNA-Sequencing and Metatranscriptome Experiments." *Journal of Biotechnology* 261: 85–96.

Lukić-Bilela, L., D. Brandt, N. Pojskić, et al. 2008. "Mitochondrial Genome of *Suberites domuncula*: Palindromes and Inverted Repeats Are Abundant in Non-Coding Regions." *Gene* 412, no. 1–2: 1–11.

Mallatt, J., C. W. Craig, and M. J. Yoder. 2010. "Nearly Complete rRNA Genes Assembled from Across the Metazoan Animals: Effects of More Taxa, a Structure-Based Alignment, and Paired-Sites Evolutionary Models on Phylogeny Reconstruction." *Molecular Phylogenetics and Evolution* 55, no. 1: 1–17.

Morandi, E., M. J. van Hemert, and D. Incarnato. 2022. "SHAPE-Guided RNA Structure Homology Search and Motif Discovery." *Nature Communications* 13: 1722.

Nawrocki, E. P., and S. R. Eddy. 2013. "Infernal 1.1: 100-Fold Faster RNA Homology Searches." *Bioinformatics* 29: 2933–2935.

Noller, H. F., J. P. Donohue, and R. R. Gutell. 2022. "The Universally Conserved Nucleotides of the Small Subunit Ribosomal RNAs." *RNA* 28: 623–644.

Page, R. D. 2000. "Comparative Analysis of Secondary Structure of Insect Mitochondrial Small Subunit Ribosomal RNA Using Maximum Weighted Matching." *Nucleic Acids Research* 28: 3839–3845.

Petrov, A. S., E. C. Wood, C. R. Bernier, et al. 2019. "Structural Patching Fosters Divergence of Mitochondrial Ribosomes." *Molecular Biology and Evolution* 36: 207–219.

Pett, W., J. F. Ryan, K. Pang, et al. 2011. "Extreme Mitochondrial Evolution in the Ctenophore *Mnemiopsis leidyi*: Insights from mtDNA and the Nuclear Genome." *Mitochondrial DNA* 22: 130.

Ramrath, D. J. F., M. Niemann, M. Leibundgut, et al. 2018. "Evolutionary Shift Toward Protein-Based Architecture in Trypanosomal Mitochondrial Ribosomes." *Science* 362: eaau7735.

Satoh, T. P., M. Miya, K. Mabuchi, and M. Nishida. 2016. "Structure and Variation of the Mitochondrial Genome of Fishes." *BMC Genomics [Electronic Resource]* 17: 719.

Siebert, S., and R. Backofen. 2005. "MARNA: Multiple Alignment and Consensus Structure Prediction of RNAs Based on Sequence Structure Comparisons." *Bioinformatics* 21: 3352–3359.

Szymanski, M., A. Zielezinski, J. Barciszewski, et al. 2016. "5SRNAdb: An Information Resource for 5S Ribosomal RNAs." *Nucleic Acids Research* 44: D180–D183.

Vicens, Q., and J. S. Kieft. 2022. "Thoughts on How to Think (and Talk) About RNA Structure." *Proceedings of the National Academy of Sciences* 119: e2112677119.

Wang, H. Y., and S. C. Lee. 2002. "Secondary Structure of Mitochondrial 12S rRNA Among Fish and Its Phylogenetic Applications." *Molecular Biology and Evolution* 19: 138–148.

Xie, Q., Y. Tian, L. Zheng, and W. Bu. 2008. "18S rRNA Hyper-Elongation and the Phylogeny of Euhemiptera (Insecta: Hemiptera)." *Molecular Phylogenetics and Evolution* 47: 463–471.

Zhang, S., J. Li, and S. J. Chen. 2024. "Machine Learning in RNA Structure Prediction: Advances and Challenges." *Biophysical Journal* 126: S0006349524000675.

Zhang, Y., M. Lang, J. Jiang, et al. 2024. "Multiple Sequence Alignment-Based RNA Language Model and Its Application to Structural Inference." *Nucleic Acids Research* 52: e3.

Zhao, Q., Z. Zhao, X. Fan, et al. 2021. "Review of Machine Learning Methods for RNA Secondary Structure Prediction." *PLoS Computational Biology* 17: e1009291.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Information S1: inz270039-sup-0001-tableS1.xls **Suppporting Informaiton S2**: inz270039-sup-0002-tableS2.xls **Supporting Information S3**: inz270039-sup-0003-SuppMat.txt **Supporting Information S4**: inz270039-sup-0004-tableS4.xls **Supporting Information S5**: inz270039-sup-0005-tableS5.xls **Supporting Information S6**: inz270039-sup-0006-tableS6.xls **Supporting Information S7**: inz270039-sup-0007-tableS7.xls **Supplementary Material**: inz270039-sup-0008-SuppMat.doc