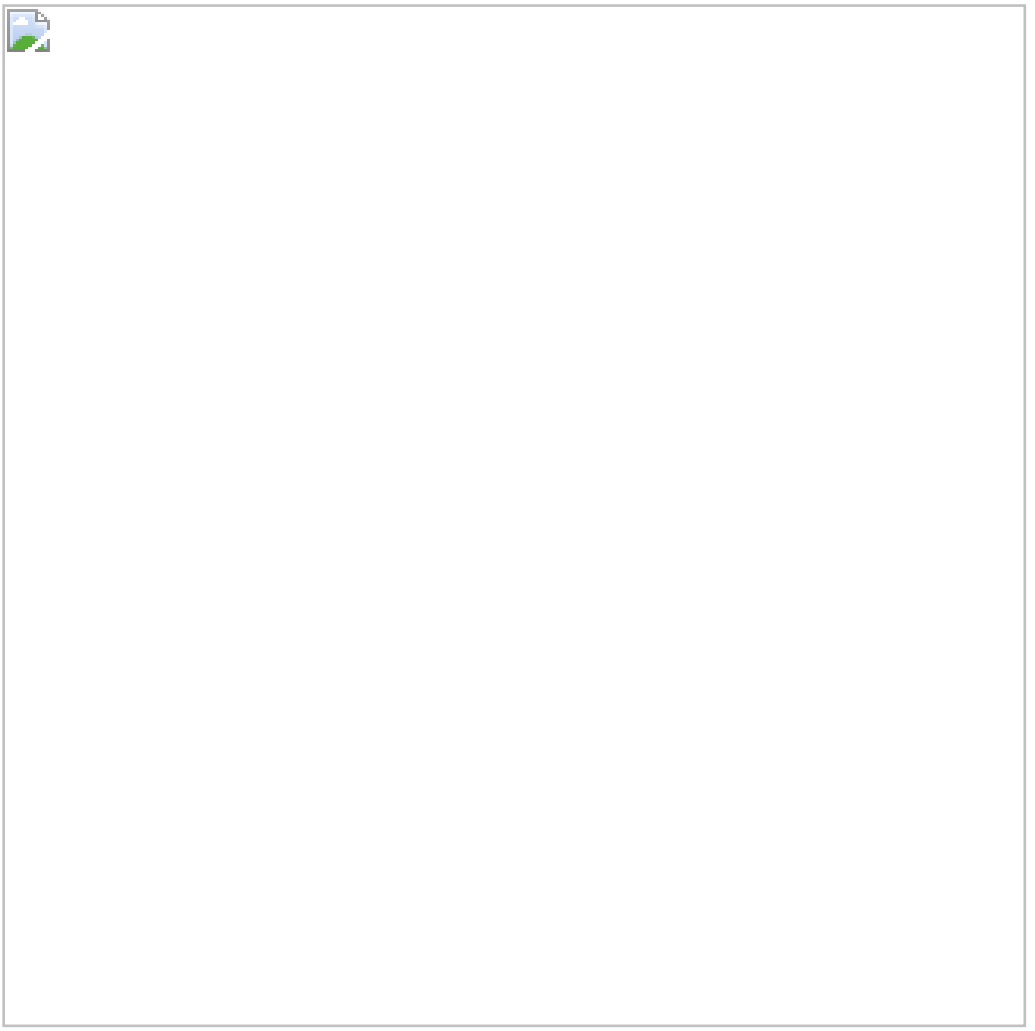

Mitochondrial genomes for the Afrobatrachian Hyperolius substriatus (Anura: Hyperoliidae) obtained from museum specimens

Authors: Bittencourt-Silva, Gabriela B., Kamouyiaros, Maria, Hartigan, Ashlie, and Okamura, Beth

Source: Journal of Vertebrate Biology, 73(24009)

Published By: Institute of Vertebrate Biology, Czech Academy of Sciences

URL: https://doi.org/10.25225/jvb.24009

Journal of Vertebrate Biology

Open Access

J. Vertebr. Biol. 2024, 73: 24009 DOI: 10.25225/jvb.24009

RESEARCH PAPER

Mitochondrial genomes for the Afrobatrachian *Hyperolius substriatus* (Anura: Hyperoliidae) obtained from museum specimens

Gabriela B. BITTENCOURT-SILVA*, Maria KAMOUYIAROS, Ashlie HARTIGAN and Beth OKAMURA

Natural History Museum, London, United Kingdom; e-mail: g.bittencourt@nhm.ac.uk

Received 19 January 2024; Accepted 4 March 2024; Published online 19 April 2024

Abstract. We report the first mitochondrial genomes of the Afrobatrachian frog species *Hyperolius substriatus*, providing insights into the gene arrangements, duplications, and evolutionary dynamics within this taxonomically complex genus. Assemblies were gained from two museum specimens and included all 37 typical vertebrate mitogenomic genes. In both cases, the mitogenome of *H. substriatus* shows duplications of tRNAs (*trnL*, *trnT* and *trnP*). The WANCY region exhibits the characteristic Afrobatrachian pattern but with unique intergenic spacers. Both copies of duplicated non-coding and control regions show high sequence similarity within and between specimens. The distinct gene order and homology of duplications observed can be explained under the tandem duplication and random loss model of mitochondrial evolution. Furthermore, a comparison with *Hyperolius marmoratus* highlights differences in gene copy number and synteny, and the presence of a duplication of *trnM* in the latter appears to be a derived condition. Together, these findings provide valuable insights into the taxonomy and evolution of *Hyperolius*, contributing to understanding interspecific gene reorganisation in this diverse group and suggesting that gene arrangements could help resolve some of the taxonomic problems of this complex genus.

Key words: Amphibia, East African reed frog, Africa, mitogenome, WANCY region

Introduction

Most animals have circular mitochondrial genomes (referred to here as mitogenomes) containing 37 genes: 13 protein-coding (PCGs), two ribosomal RNAs (rRNAs), and 22 transfer RNAs (*trns*) (Boore 1999). In addition to these, vertebrate mitogenomes typically also feature a large non-coding region called the control region (CR) and a triple-stranded displacement loop (D-loop) region (Kasamatsu et al. 1971).

Vertebrate mitogenomes generally have conserved gene organisation. However, the WANCY region, which includes the origin of replication of the light strand (O_L) and tRNAs in the order: *trnW*, *trnA*, *trnN*, O_L, *trnC*, and *trnY* (Seutin et al. 1994), is considered a hotspot of gene rearrangement. Different arrangements observed in this region have been explained by the tandem duplication-random loss (TDRL) model, which presupposes that novel gene orders arise from tandem repeats, followed by random deletions of each pair of the duplicated genes (Boore 1999, Moritz et al. 2017).

Kurabayashi & Sumida (2013) propose that the WANCY region, denoted as WNO_LACY, is a synapomorphy of Afrobatrachia, which comprises 426 species of frogs distributed in the families

* *Corresponding Author*

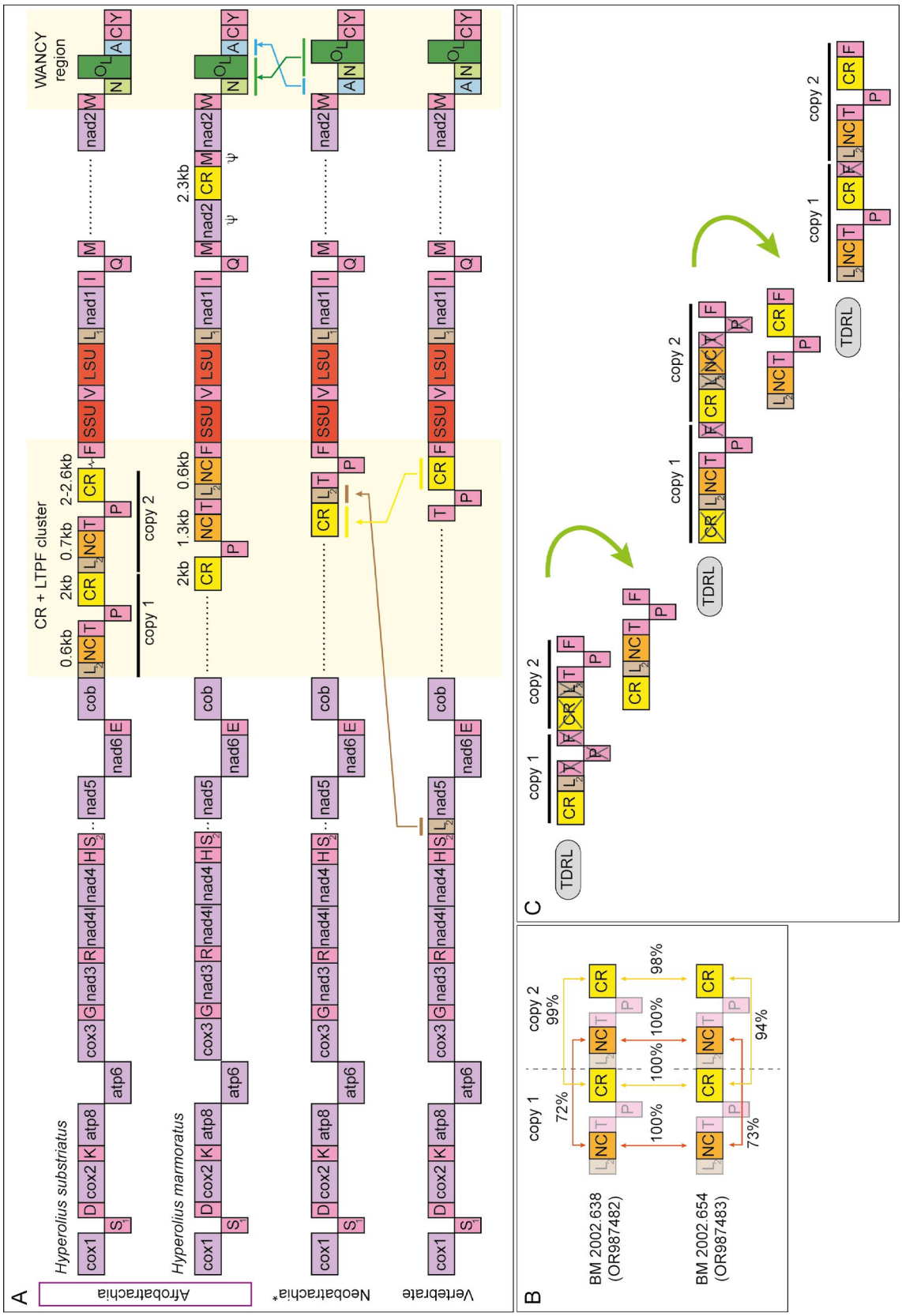

Fig. 1. Mitochondrial genome organisation. Protein-coding genes, rRNAs (SSU and LSU), light-strand origin of replication (O_L), and non-coding regions are represented and labelled in boxes. All tRNA genes are labelled based on their relevant single-letter amino acid codes. L_1 and L_2 indicate Leucine (CUN) and Leucine (UUR), respectively. S_1 and S_2 indicate Serine (AGY) and Serine (UCN), respectively. The letter 'ψ' denotes a reported pseudogene. The 'WANCY' region and 'CR + tRNA LTPF cluster' are highlighted. Boxes above represent genes encoded by the heavy strand and the ones below by the light strand. A) Gene order and arrangement; B) percentage of sequence similarity between the two copies of the CR + LTPF cluster of the two specimens of *Hyperolius substriatus*; C) potential TDRL events that explain the observed pattern of gene order.

Arthroleptidae, Brevicipitidae, Hemisotidae, and Hyperoliidae (*sensu* Frost et al. 2006). The family Hyperoliidae Laurent, 1943 comprises over 200 species distributed in 17 genera in sub-Saharan Africa. *Hyperolius* Rapp, 1842 is the most speciose genus with 144 species. Six complete (Kurabayashi & Sumida 2013, Hemmi et al. 2020) and five near-complete (Zhang et al. 2013) mitogenomes are available for Afrobatrachians. To date, there is only one complete and one partial mitogenome available for the genus *Hyperolius* (*Hyperolius marmoratus* (GenBank: AB777218) and *H. ocellatus*, (GenBank: JX564872)). Here, we present two near-complete but uncircularised mitogenomes of the East African reed frog *Hyperolius substriatus* Ahl, 1931.

Material and Methods

The two specimens of *H. substriatus* (BMNH 2002.638 and BMNH 2002.654) from the Nilo Nature Forest Reserve (–4.93333, 38.65000), northwest of the East Usambara Mountains in Tanzania, are part of the collection of the Natural History Museum in London, UK. These specimens, collected in 2002, were fixed and preserved in 70% ethanol. A small fragment of liver was dissected from each specimen using sterile forceps. The samples were lysed overnight in 180 µl Qiagen® ATL buffer and 10 µl Proteinase K at 56 °C. Total DNA was extracted from each lysate using the Qiagen® DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) per the manufacturer's protocol. Genomic libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® enzymatic fragmentation (Illumina, San Diego) and sequenced on a NovaSeq PE150 (Novogene Europe, Cambridge). Low-quality ends and adapters were trimmed using Trim Galore v. 0.6.10 (https://github.com/FelixKrueger/TrimGalore), applying a Phred score threshold of 20 and stringency of 4. NOVOPlasty v.4.2 (Dierckxsens et al. 2017) was used to independently *de novo* assemble the mitogenomes using k-mer size 33, and a cytochrome *c* oxidase subunit I sequence from *H. substriatus* used as a seed (GenBank: KY177140). Contigs generated by NOVOPlasty were then visualised in Geneious 9.0.5 (https://www.geneious.com) for final assembly.

For mitogenome annotation, we used MITOS (Bernt et al. 2013) and tRNAscan-SE 2.0 (Lowe & Chan 2016). Both mitogenomes and respective annotations were manually inspected in Geneious and edited in some cases (see Results). Potential homologies between the two copies of the NC region and the two CRs were investigated for each specimen by aligning these sequences using MAFFT v7.309 (Katoh et al. 2002) (Auto algorithm; 200PAM/k = 2; offset value = 0.123; gap opening penalty = 1.53) in Geneious. To confirm the taxonomic position of our mitogenomes within Afrobatrachians, we used the same algorithm to align rRNAs (12S and 16S) + *trnV* of both *H. substriatus* specimens with other ranoid taxa, including representatives of all families of Afrobatrachians, plus a proximate (Microhylidae) and a more distant (Natatanura) outgroup for rooting. Poorly aligned regions were masked using GBlocks (Castresana 2000) applying the default parameters. A maximum likelihood tree was built using RAxML-NG v. 1.1.0 (Kozlov et al. 2019). The best substitution model (TIM2 + I + G4m) was calculated using Modeltest-NG, and the run mode was set to ML tree search + bootstrapping (autoMRE) to a maximum of 1,000 replicates. Complete mitogenomes of Afrobatrachians and outgroups were downloaded from GenBank (RefSeq versions): *Breviceps adspersus* (NC023379), *Hemisus marmoratus* (NC023380), *Hyperolius marmoratus* (NC023381), *Trichobatrachus robustus* (NC023382), *Kaloula rugifera* (NC029409) and *Rana coreana* (NC068259).

Results

Two near-complete assemblies for specimens BMNH 2002.638 and BMNH 2002.654 were produced (GenBank accessions: OR987482 and OR987483) using NOVOPlasty. Of the two libraries used, only the mitogenome for specimen BMNH 2002.654 resulted in a circularised mitogenome; however, upon manual inspection, one of the four final contigs

Table 1. Size and gene types of mitogenomes of *Hyperolius substriatus** and *Hyperolius marmoratus***.

Species	ID	Size bp	PCGs	rRNA	tRNAs	Duplicated genes
H. substriatus	BMNH 2002.638	21,244	13	2	25	*trnL*(*CUN*), *trnP*, *trnT*
H. substriatus	BMNH 2002.654	21,840	13	2	25	*trnL*(*CUN*), *trnP*, *trnT*
H. marmoratus	AB777218	22,595	13	2	23	*trnM*, NAD2

*This study; **Kurabayashi & Sumida (2013).

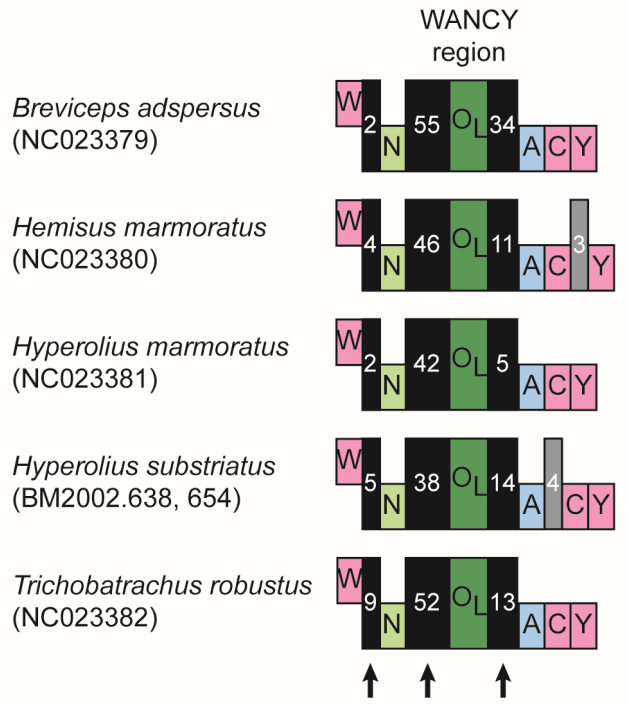

Fig. 2. Afrobatrachian gene order and arrangement of the WANCY region. Arrows indicate intergenic spacers common to all species and numbers denote the base pair length of each spacer.

comprised a 21,853 bp repeated motif. As such, this contig was excluded. While neither final assembly was circularised by NOVOPlasty, both resulted in a single contiguous sequence with the same gene order and contained all expected vertebrate mitochondrial genes. This outcome allowed for the first characterisation of the *H. substriatus* mitogenome and further comparative analyses. Table 1 summarises the size and gene types of mitogenomes of *H. substriatus* and *Hyperolius marmoratus* (GenBank: AB777218).

The mitogenome of *H. substriatus* contains a total of 40 genes because there is a duplication of tRNAs *trnL* (*CUN*), *trnT*, and *trnP* (Fig. 1A). Thirty genes are encoded on the H-strand and ten on the L-strand, in accord with the typical vertebrate arrangement. Nucleotide composition (average between both specimens) is 32% A, 33% T, 22% C, and 13% G.

As expected, the WANCY region of *H. substriatus* follows the Afrobatrachia pattern (Fig. 1A). Additionally, a short intergenic spacer (IGS) of 5 bp is found between *trnW* and *trnN*, followed by two longer ones of 38 and 14 nucleotides that are respectively found before and after the O_L. Presumed homologues of these IGSs (same position and similar sizes) are also present but previously unreported in other Afrobatrachians (*B. adspersus*, *Hemisus marmoratus*, *Hyperolius marmoratus* and *T. robustus*). Finally, a fourth IGS (4 bp) is observed between *trnA* and *trnC*, a pattern which is only similarly observed in *Hemisus marmoratus* (Fig. 2).

The cytochrome *b* (cyt*b*) gene of *H. substriatus* lacks a complete stop codon, presumably generated upon post-transcriptional polyadenylation (see comments in Discussion). The MITOS annotation of this gene

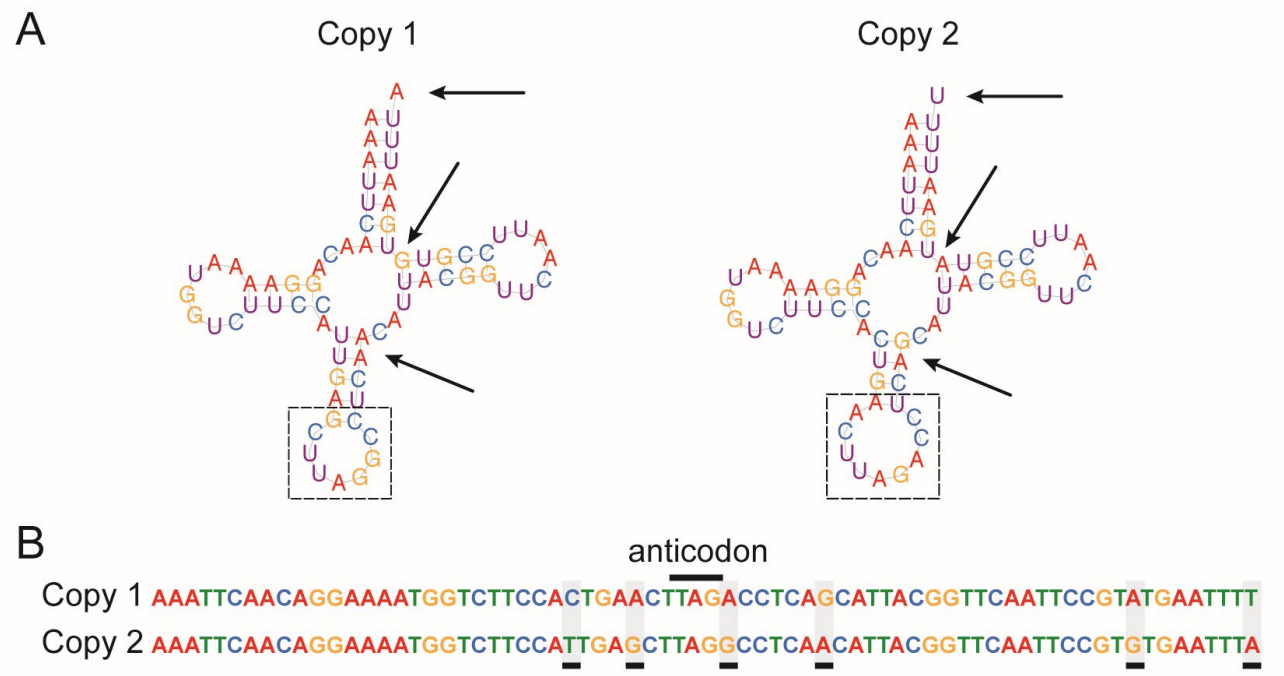

Fig. 3. Secondary structure and DNA sequences of two copies of *trnL* (CUN). A) Arrows and boxes indicate mismatches between copies. B) DNA sequence differences in bases (highlighted in grey and underscored).

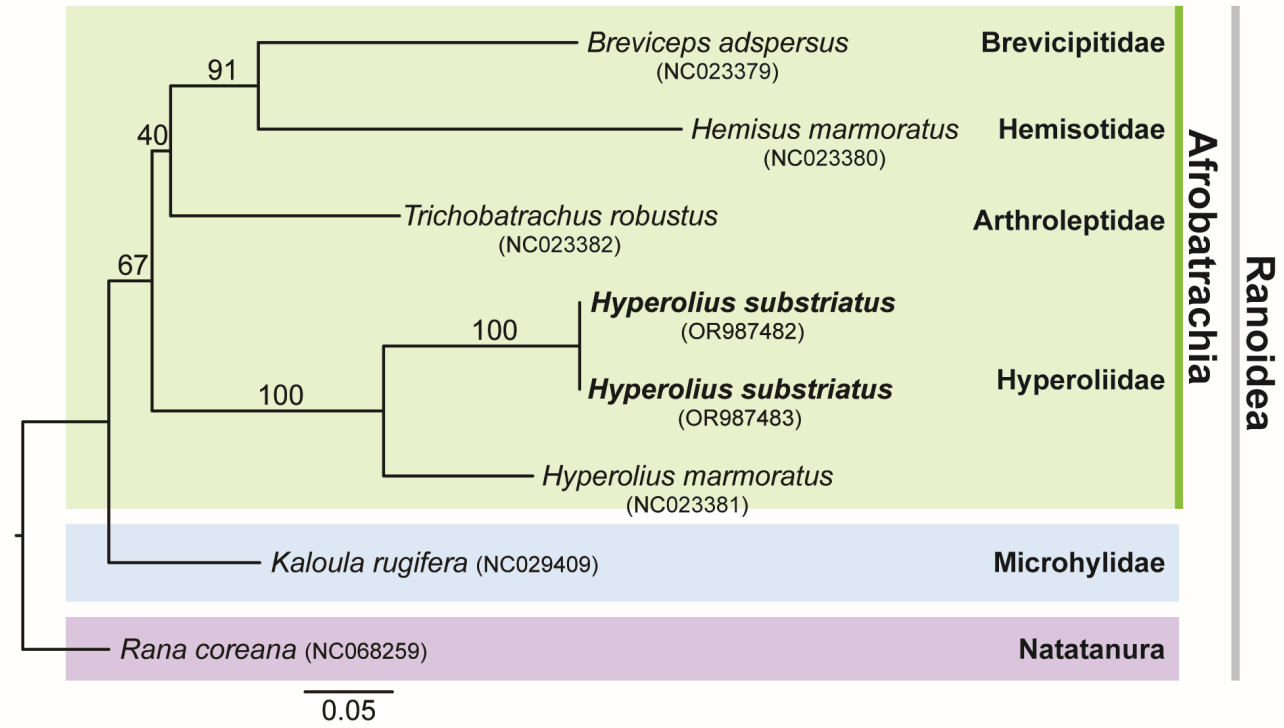

Fig. 4. Maximum likelihood phylogeny based on the two ribosomal RNAs (12S and 16S) and *trnV*. Values above the branch indicate bootstrap support. GenBank accession numbers are shown in parentheses.

was shorter than expected (1,134 bp; 378 amino acids), so it was manually annotated extending a further nine base pairs, resulting in 1,143 bp and 381 amino acids (Table 2).

In *H. substriatus*, the CR + LTPF tRNA cluster (*trnL* (*CUN*)-*trnT-trnP-trnF*) differs from the typical Neobatrachian arrangement (Kurabayashi & Sumida 2013) in having two copies of *trnL* (*CUN*)-*trnT-trnP*, with a non-coding region (NC) inserted between *trnL* (*CUN*) and *trnT* (Fig. 1A,C). The two copies of *trnT* and *trnP* have identical nucleotide sequences, suggesting that both are functional (i.e. neither has become a pseudogene). The two copies of the *trnL* (*CUN*) have the same length and similar structure but differ by six nucleotides (Fig. 3). A further distinguishing feature is that the two copies of the CR are arranged after the *trnP*. The first copy of the CR has a similar length in both specimens of *H. substriatus*, whereas the second copy varies substantially in length (BMNH 2002.638 = 2,023 bp; BMNH 2002.654 = 2,695 bp) (Fig. 1A). These copies of the CR are incomplete and probably are connected to the *trnF* (Fig. 1A), which is adjacent to the SSU (12S).

Comparisons of the two copies of the NC region and the two CRs were conducted to explore homologies further. Pairwise sequence similarity was calculated for the NC regions using 654 bp (89% of the original 728 bp). The first NC region was identical in both

Table 2. Size and number of encoded amino acids of the cyt*b* gene of some Afrobatrachians.

GenBank ID	Species	Base pairs	Amino acids	Stop codon
NC023379	*Breviceps adspersus*	1140	380	AGA
LC498571	*Breviceps mossambicus*	1140	380	AGA
LC498572	*Breviceps poweri*	1137	379	TAA
NC023380	*Hemisus marmoratus*	1152	384	TAA
NC023381	*Hyperolius marmoratus*	1140	380	TAA
OR987482	*Hyperolius substriatus*	1143	381	TA*
OR987483	*Hyperolius substriatus*	1143	381	TA*
NC023382	*Trichobatrachus robustus*	1146	382	TAA

*Incomplete stop codon.

specimens, and the second was revealed to vary from 72 to 73% (Fig. 1B). Comparisons of the CRs were based on 1,993 bp (73% of the original 2,703 bp). The first CR is identical in both specimens, whereas the second is 98% similar. A comparison of each copy within the specimen revealed 94-99% similarity (Fig. 1B).

The phylogeny confirms that *H. substriatus* groups with *Hyperolius marmoratus* (Fig. 4). Low branch support for other clades is likely due to the small selection of genes used to infer the phylogeny. However, the close relationship between Afrobatrachia and Microhylidae and their sister relationship with Natatanura is well supported elsewhere (see Kurabayashi & Sumida 2013).

Discussion

Here, we provide the first mitochondrial genome records for the Afrobatrachian *H. substriatus*. Both assemblies are similar in length to the complete mitogenome available for *Hyperolius marmoratus* (Table 2), suggesting non-circularisation is due to incomplete cover of the CR. Because the available mitogenome of *H. ocellatus* is only partial (9,457 bp) and does not include the complete WANCY region or CR + LTPF cluster, no further comparisons can be made now.

A single copy of *trnM* was found in *H. substriatus*, while two were found in *Hyperolius marmoratus*. Because typical Neobatrachia and vertebrate mitogenomes only have one copy of *trnM*, the condition in *Hyperolius marmoratus* is a derived feature. The organisation of the WANCY region and IGSs is consistent in the other five species of Afrobatrachians compared here, except for one extra IGS in *H. substriatus* and *Hemisus marmoratus* (see Fig. 2). These small fragments of non-coding DNA are possibly remnants of gene duplication and are expected under the TDRL model. However, at least three TDRL events would be necessary to explain the pattern observed in Afrobatrachians (Fig. 2).

It is not uncommon for the cyt*b* of vertebrates to lack or have incomplete stop codons (e.g. Görtz & Feldmann 1982, Murray et al. 1994, McKnight & Shaffer 1997, Glenn et al. 2002). The cyt*b* gene usually encodes 380 amino acids in vertebrates, but deviations from this number are not unusual. For instance, within the Afrobatrachians analysed here, the number of amino acids varies between 379 and 384 (Table 2).

Duplicated NC regions and CRs of similar sizes to those found in *H. substriatus* also occur in *Hyperolius marmoratus* but with different synteny (see Fig. 1A). Given the high similarity between the two copies of the NC regions and CRs observed in *H. substriatus*, we speculate that the two copies corresponding to the CR + LTPF tRNA cluster are homologous and could be explained by three TDRL events (Fig. 1C). The arrangement of this gene cluster in *H. substriatus* is a derived condition from Neobatrachia.

This study provides the first mitogenome-wide comparison of gene orders and arrangements for the genus *Hyperolius*, showing significant interspecific reorganisation and duplication-loss events. The unique variation within *Hyperolius* mitogenomes revealed here contributes to growing research on interspecific gene reorganisation and improves taxonomic resolution within this speciose lineage.

Acknowledgements

Our research was funded by The Leverhulme Trust (RPG-2019-322). We thank Diego San Mauro and Mark Wilkinson for their valuable comments that improved the quality of the paper.

Author Contributions

The authors confirm their contribution to the paper as follows: study conception and design: G.B. Bittencourt Silva, B. Okamura, A. Hartigan; data acquisition: A. Hartigan, G.B. Bittencourt Silva; analysis and interpretation of results: G.B. Bittencourt Silva, M. Kamouyiaros; draft manuscript preparation: G.B. Bittencourt Silva. All authors contributed to the writing, revised the manuscript critically for intellectual content, agree to be accountable for all aspects of the work, and approved the final version for publication.

Data Availability Statement

The mitogenomes generated for this study are available in GenBank: OR987482 and OR987483.

Literature

Bernt M., Donath A., Jühling F. et al. 2013: MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol. 69: 313–319.*

Boore J.L. 1999: Animal mitochondrial genomes. *Nucleic Acids Res. 27: 1767–1780.*

Castresana J. 2000: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol. 17: 540–552.*

Dierckxsens N., Mardulyn P. & Smits G. 2017: NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res. 45: 18.*

Frost D.R., Grant T., Faivovich J. et al. 2006: The amphibian tree of life. *Bull. Am. Mus. Nat. Hist. 297: 1–291.*

Glenn T.C., Staton J.L., Vu A.T. et al. 2002: Low mitochondrial DNA variation among American alligators and a novel non-coding region in crocodilians. *J. Exp. Zool. 294: 312–324.*

Görtz G. & Feldmann H. 1982: Nucleotide sequence of the cytochrome *b* gene and adjacent regions from rat liver mitochondrial DNA. *Curr. Genet. 5: 221–225.*

Hemmi K., Kakehashi R., Kambayashi C. et al. 2020: Exceptional enlargement of the mitochondrial genome results from distinct causes in different rain frogs (Anura: Brevicipitidae: *Breviceps*). *Int. J. Genomics 2020: 6540343.*

Kasamatsu H., Robberson D.L. & Vinograd J. 1971: A novel closed-circular mitochondrial DNA with properties of a replicating intermediate. *Proc. Natl. Acad. Sci. U. S. A. 68: 2252–2257.*

Katoh K., Misawa K., Kuma K. & Miyata T. 2002: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res. 30: 3059–3066.*

Kozlov A.M., Darriba D., Flouri T. et al. 2019: RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics 35: 4453–4455.*

Kurabayashi A. & Sumida M. 2013: Afrobatrachian mitochondrial genomes: genome reorganisation, gene rearrangement mechanisms, and evolutionary trends of duplicated and rearranged genes. *BMC Genomics 14: 633.*

Lowe T.M. & Chan P.P. 2016: tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res. 44: W54–W57.*

McKnight M.L. & Shaffer H.B. 1997: Large, rapidly evolving intergenic spacers in the mitochondrial DNA of the salamander family Ambystomatidae (Amphibia: Caudata). *Mol. Biol. Evol. 14: 1167–1176.*

Moritz C., Dowling T. & Brown W. 2017: Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst. 18: 269–292.*

Murray B.W., McGillivray W.B., Barlow J.C. et al. 1994: The use of cytochrome *b* sequence variation in estimation of phylogeny in the Vireonidae. *Condor 96: 1037–1054.*

Seutin G., Lang B.F., Mindell D.P. & Morais R. 1994: Evolution of the WANCY region in amniote mitochondrial DNA. *Mol. Biol. Evol. 11: 329–340.*

Zhang P., Liang D., Mao R.-L. et al. 2013: Efficient sequencing of Anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. *Mol. Biol. Evol. 30: 1899–1915.*