Gene 686 (2019) 164–170

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Mitogenomics of Central American weakly-electric fishes

Celestino Aguilar[a,b,c], Matthew J. Miller[d], Jose R. Loaiza[a,c], Rüdiger Krahe[e], Luis F. De León[a,f,*]

[a] Centro de Biodiversidad y Descubrimiento de Drogas, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT AIP), P. O. Box 0843-01103, Panama, Panama
[b] Department of Biotechnology, Acharya Nagarjuna University, Guntur, India
[c] Smithsonian Tropical Research Institute, P.O. Box 0843-03092, Balboa Ancón, Panama, Panama
[d] Sam Noble Oklahoma Museum of Natural History and Department of Biology, University of Oklahoma, Norman, OK, USA
[e] Institut für Biologie, Humboldt-Universität zu Berlin, Berlin, Germany
[f] Department of Biology, University of Massachusetts Boston, Boston, MA, USA

ARTICLE INFO

Keywords:
Gymnotiformes
Intergenic spacer
Mitogenome
Next generation sequencing
Panama

ABSTRACT

Electric fishes are a diverse group of freshwater organisms with the ability to generate electric organ discharges (EODs) that are used for communication and electrolocation. This group (ca. 200 species) has originated in South America, and six species colonized the Central American Isthmus. Here, we assembled the complete mitochondrial genomes (mitogenomes) for three Central American electric fishes (i.e. *Sternopygus dariensis*, *Brachyhypopomus occidentalis*, and *Apteronotus rostratus*), and, based on these data, explored their phylogenetic position among Gymnotiformes. The three mitogenomes show the same gene order, as reported for other fishes, with a size ranging from 16,631 to 17,093 bp. We uncovered a novel 60 bp intergenic spacer (IGS) located between the *COII* and $tRNA^{Lys}$ genes, which appears to be unique to the Apteronotidae. Furthermore, phylogenetic relationships supported the traditional monophyly of Gymnotiformes, with the three species positioned within their respective family. In addition, the genus *Apteronotus* belongs to the early diverging lineage of the order. Finally, we found high sequence divergence (13%) between our *B. occidentalis* specimen and a sequence previously reported in GenBank, suggesting that the prior mitogenome of *B. occidentalis* represents a different South American species. Indeed, phylogenetic analyses using *Cytochrome b* gene across the genus placed the previously reported individual within *B. bennetti*. Our study provides novel mitogenome resources that will advance our understanding of the diversity and phylogenetic history of Neotropical fishes.

1. Introduction

Electric fishes (Teleostei, Gymnotiformes) are a highly diverse group of freshwater organisms that originated in South America (Albert, 2001). One of the defining features of these fishes is their ability to produce electric organ discharges (EODs) that are used for communication and electrolocation (Moller, 1995; Bullock et al., 2005). EODs are species-specific electric signals that can be divided into pulse-type and wave-type, depending on the shape and regularity of the discharge. In addition, there is evidence for reproductive character displacement in EOD waveform in this group (Crampton et al., 2011), which has over 200 currently described species.

Electric fishes are widely distributed in lowland freshwater habitats throughout South America (Albert and Crampton, 2005; Hulen et al., 2005). In Central America, however, only 6 species and five genera have been reported so far, including *Apteronotus rostratus*, *Brachyhypopomus occidentalis*, *Eigenmannia humboldtii*, *Gymnotus carapo*, *G. panamensis* and *Sternopygus dariensis* (Reis et al., 2003; Albert and Crampton, 2005; Alda et al., 2013). Despite the high diversity of Neotropical electric fishes, limited genomic resources are currently available for the group, particularly for Central American species. For instance, to date, only nine mitochondrial genomes of Gymnotiformes have been deposited in GenBank (Saitoh et al., 2003; Nakatani et al., 2011; Lavoué et al., 2012; Elbassiouny et al., 2016), but none of these mitogenomes

Abbreviations: ATP6 and *ATP8*, ATPase subunit 6 and 8 genes; *CYTB*, cytochrome *b* gene; *COI-III*, cytochrome oxidase subunits I-III genes; NCR, non-coding region; *ND1–6* and *ND4L*, NADH dehydrogenase subunits 1–6 and 4 L genes; ML, maximum likelihood; BI, Bayesian inference; *rRNA*, ribosomal RNA; *16S* and *12S*, large and small subunits of ribosomal RNA genes; *tRNA*, transfer RNA; PCG, protein coding gene; mitogenome, mitochondrial genome; *Ala*, alanine; *Arg*, arginine; *Asn*, asparagine; *Aps*, aspartic acid; *Cys*, cysteine; *Gln*, glutamine; *Glu*, glutamic acid; *Gly*, glycine; *His*, histidine; *Ile*, isoleucine; *Leu*, leucine; *Lys*, lysine; *Met*, methionine; *Phe*, phenylalanine; *Pro*, proline; *Ser*, serine; *Thr*, threonine; *Trp*, tryptophan; *Tyr*, tyrosine; *Val*, valine

* Corresponding author at: Department of Biology, University of Massachusetts Boston, Boston, MA, USA.
E-mail address: luis.deleonreyna@umb.edu (L.F. De León).

https://doi.org/10.1016/j.gene.2018.11.045
Received 19 June 2018; Received in revised form 5 November 2018; Accepted 15 November 2018
Available online 16 November 2018
0378-1119/ Published by Elsevier B.V.

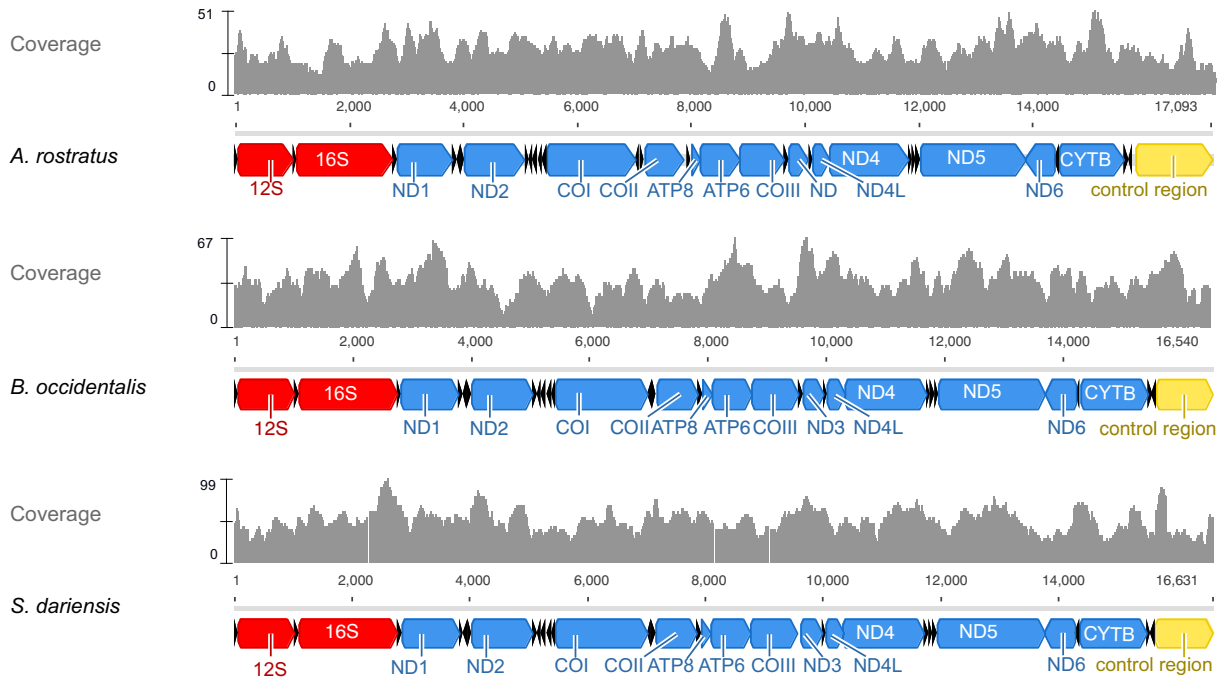

Fig. 1. Structure and read coverage of the assembled mitogenomes of *A. rostratus*, *B. occidentalis* and *S. dariensis*. Gray (Coverage) plots show the read depth across the complete mitogenomes. Direction of gene transcription for each species is indicated by the arrows: rRNA genes (red), PCGs (blue), tRNA genes (black) and the control region (yellow). PCG and rRNA genes are shown with standard abbreviations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

were generated for Central American taxa. In the case of *B. occidentalis*, it is difficult to determine if the individual from South America previously reported by Lavoué et al. (2012) corresponds to the Central American species, particularly, because *B. occidentalis* presents a wide geographic distribution in South and Central America (Crampton et al., 2016b), and species-level divergence is likely to exist across the species' range (Bermingham and Martin, 1998; Picq et al., 2014). This knowledge gap is important because the dynamic history of the Central American Isthmus has led to a complex evolutionary and phylogeographic history among electric fishes as well as other freshwater fishes (Bermingham and Martin, 1998; Picq et al., 2014). Thus, generating molecular datasets – including complete mitogenomes – for Central American species of electric fishes is valuable to improve our understanding of diversification in Neotropical environments.

Here, we assemble for the first time complete mitogenome sequences for three Central American weakly-electric fishes: two wave-type species – *Apteronotus rostratus* and *Sternopygus dariensis* – and the pulse-type *Brachyhypopomus occidentalis*. We also compile currently available mitogenomic data to assess the phylogenetic position of the three species within Gymnotiformes. In addition, we estimate genetic distances across complete mitochondrial genomes of three *Brachyhypopomus* individuals available in GenBank, and examine species diversity within this genus. Our study provides novel genomic resources that could facilitate further work on the conservation genetics, phylogenetics, and evolution of Central American Gymnotiformes as well as other freshwater fishes.

2. Materials and methods

2.1. Study site and sampling protocol

We collected three individuals from each of the following species: *A. rostratus, S. dariensis* and *B. occidentalis* in La Hoya stream, which flows into the Chucunaque River in the Darien Province, eastern Panama (N 8.2536, W -77.7189). Fish were detected using wire-electrodes connected to a mini-amplifier (Radio/Shack, Fort Worth, TX), and collected using a hand-net. Fish were euthanized with an overdose of eugenol ($C_{10}H_{12}O_2$) derived from clove oil, and specimen were deposited in Neotropical Fish Collection at the Smithsonian Tropical Research Institute. Our collecting protocol was authorized by Panama's Ministry of the Environment (MiAmbiente; permit number SE/A-100-14) and approved by the Institutional Animal Care and Use Committee (IACUC-16-001) at the Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT AIP).

2.2. Sequencing

Mitochondrial genomes were obtained as the byproduct of Next Generation Sequencing (NGS) of Ultraconserved Elements (UCEs; Faircloth et al., 2012), as part of ongoing studies on the population genomics of the weakly-electric fish *B. occidentalis*. We prepared UCEs libraries following a standard protocol (available from http://ultraconserved.org) using the 500 loci Actinopterygii probe set (Actinopterygians 0.5Kv1; Faircloth et al., 2013). Libraries were sequenced using 300 bp paired-end Illumina MiSeq platform (Illumina, San Diego, CA) at the Smithsonian Tropical Research Institute (STRI) Naos Molecular Lab in Panama City, Panama.

2.3. Mitogenome assembly and annotation

We followed Aguilar et al. (2016) to generate mitogenomes from UCE sequencing reads. Briefly, we used Illumiprocessor (Faircloth, 2013), which employs Trimmomatic (Bolger et al., 2014) to clean and trim reads. We then assembled all reads using Trinity (Grabherr et al., 2011). Complete mitogenomes were identified by performing BLAST (Basic Local Alignment Search Tool; https://blast.ncbi.nlm.nih.gov/Blast.cgi) searches of sequence similarity on contigs larger than 15,000 bp.

To circularize each mitogenome, we examined the overlap of each contig's edge, using the software Geneious version 11.1.4 (http://www.

Table 1
Characteristics of the mitochondrial genomes of Central American electric fishes. The table shows the mitogenome structure of three species in the following order: *S. dariensis*, *B. occidentalis* and *A. rostratus*.

Genomic region	Start	Stop	Size	Spacer (+) or overlap (−)	Direction	Start codon	Stop codon
$tRNA^{Phe}$	36,892	68/69/69	68/69/69	0/0/0	F		
12S rRNA	69/70/70	1020/1018/1016	952/949/947	0/0/0	F		
$tRNA^{Val}$	1021/1019/1017	1092/1089/1088	72/71/72	0/0/0	F		
16S rRNA	1093/1090/1089	2766/2750/2760	1674/1661/1672	0/0/0	F		
$tRNA^{Leu}$	2767/2751/2761	2841/2825/2835	75/75/75	1/0/2	F		
ND1	2843/2826/2838	3817/3785/3806	975/960/969	2/4/5	F	ATG/ATT/ATT	TAA/TAA/TAA
$tRNA^{Ile}$	3820/3790/3812	3891/3862/3883	72/73/72	-2/−2/−2	F		
$tRNA^{Gln}$	3890/3861/3882	3960/3931/3952	71/71/71	−1/−1/−1	R		
$tRNA^{Met}$	3960/3931/3952	4029/4000/4020	70/70/70	0/0/0	F		
ND2	4030/4001/4021	5074/5045/5070	1045/1045/1050	0/0/4	F	ATG/ATG/ATA	T−/T−/TAA
$tRNA^{Trp}$	5075/5046/5075	5144/5117/5145	70/72/71	0/1/1	F		
$tRNA^{Ala}$	5145/5119/5147	5213/5186/5215	69/68/69	1/1/1	R		
$tRNA^{Asn}$	5215/5188/5217	5287/5260/5289	73/73/73	31/29/30	R		
$tRNA^{Cys}$	5319/5290/5320	5384/5356/5385	66/67/66	0/1/0	R		
$tRNA^{Tyr}$	5385/5358/5386	5454/5426/5454	70/69/69	1/1/1	R		
COI	5456/5428/5456	7021/6987/7015	1566/1560/1560	−5/2/−5	F	GTG/GTG/GTG	AGA/TAA/AGA
$tRNA^{Ser}$	7017/6990/7011	7087/7060/7081	71/71/71	4/4/5	R		
$tRNA^{Asp}$	7092/7065/7087	7160/7133/7155	69/69/69	14/13/11	F		
COII	7175/7147/7167	7866/7837/7847	692/691/681	0/0/60	F	ATG/ATG/ATG	TA−/T−/TAA
$tRNA^{Lys}$	7867/7838/7908	7940/7912/7981	74/75/74	1/1/1	F		
ATP8	7942/7914/7983	8109/8078/8144	168/165/162	−10/−7/−4	F	ATG/ATG/GTG	TAA/TAG/TAG
ATP6	8100/8072/8141	8782/8754/8823	683/683/683	0/0/0	F	ATG/ATG/ATA	TA−/TA−/TA-
COIII	8783/8755/8824	9566/9538/9607	784/784/784	0/1/0	F	ATG/ATG/ATG	T−/T−/T-
$tRNA^{Gly}$	9567/9540/9608	9637/9609/9678	71/70/71	0/0/1	F		
ND3	9638/9610/9680	9987/9959/10030	350/350/351	0/0/6	F	ATG/ATA/ATC	TA−/TA−/TAA
$tRNA^{Arg}$	9988/9960/10037	10,057/10029/10106	70/70/70	0/0/0	F		
ND4L	10,058/10030/10107	10,354/10326/10403	297/297/297	−7/−7/−7	F	ATG/ATA/ACG	TAA/TAA/TAA
ND4	10,348/10320/10397	11,728/11700/11777	1381/1381/1381	0/0/0	F	ATG/ATG/ATG	T−/T−/T-
$tRNA^{His}$	11,729/11701/11778	11,797/11769/11846	69/69/69	0/0/0	F		
$tRNA^{Ser}$	11,798/11770/11847	11,865/11836/11913	68/67/67	1/1/0	F		
$tRNA^{Leu}$	11,867/11838/11914	11,939/11910/11986	73/73/73	0/0/0	F		
ND5	11,940/11911/11987	13,775/13713/13819	1836/1803/1833	−5/−5/−5	F	ATG/ATG/ATC	TAA/TAA/TAA
ND6	13,771/13709/13815	14,292/14230/14336	522/522/522	0/0/0	R	ATG/ATG/GTG	AGG/AGG/AGA
$tRNA^{Glu}$	14,293/14231/14337	14,361/14299/14405	69/69/69	2/5/6	R		
CYTB	14,364/14305/14412	15,504/15441/15552	1141/1137/1141	0/4/0	F	ATG/ATG/ATG	T−/TAG/T-
$tRNA^{Thr}$	15,505/15446/15553	15,576/15518/15623	72/73/71	−1/−2/−2	F		
$tRNA^{Pro}$	15,576/15517/15622	15,645/15585/15691	70/69/70	0/0/0	R		
control region D-loop	15,646/15586/15692	16,631/16540/17093	986/955/1402	0/0/0	F		

geneious.com; Kearse et al., 2012). We also verified the assemblage of each mitogenome, using the genome-based assembler software NOVOPlasty 2.7.2 (Dierckxsens et al., 2016). For this assembly we used the raw UCEs sequencing reads as well as reference sequences of the *COI* mtDNA gene from closely related species available in GenBank: *A. albifrons* (Accession no. AB054132), *B. occidentalis* (Accession no. AP011570) and *S. arenatus* (Accession no. KX058571). The k-mer length was set to 39 bp, which resulted in mtDNA circular contigs. Lastly, raw reads were mapped onto their respective generated mitochondrial contig using Geneious.

We annotated genes using the MitoFish and MitoAnnotator (Iwasaki et al., 2013) and also inspected alignments manually, by comparing them with GenBank reference mitogenomes from *A. albifrons* (Accession no. AB054132), *S. arenatus* (Accession no. KX058571), *B. verdii* (*B. n.*sp. VERD – Accession no. KX058570) and *B. occidentalis* (Accession no. AP011570) in Geneious. Nucleotide base composition was also calculated in Geneious. Finally, strand asymmetry was estimated using the following formulas: AT-skew = (A–T)/(A + T) and GC-skew = (G–C)/(G + C), which represent an estimate of the nucleotide compositional difference between complete mitogenomes (Perna and Kocher, 1995).

2.4. DNA sequence alignment and data partitioning scheme

Thirteen mitogenome sequences were analyzed, including the outgroup *Astyanax paranae* (GenBank accession no. KX609386). For these analyses, we excluded the control region, the heterogeneous base composition ND6 gene, as well as the 22 tRNA genes. We separately aligned each gene [2 RNAs and the 12 protein coding genes (PCGs)] with MAFFT (Katoh and Standley, 2013), using the alignment with translation option in Geneious. Stop codons were removed manually. Finally, all 14 genes were concatenated for a final alignment of 13,508 bp. The same alignment was used for phylogenetic reconstruction under two inference methods (see Phylogenetic analysis).

To infer the optimal partitioning strategy for our analyses, we used the PartitionFinder2 on XSEDE (Lanfear et al., 2016) as implemented in CIPRES Science Gateway v3.3 cluster (Miller et al., 2010). Simultaneously, Bayesian Information Criterion (BIC) was used to select the best model for each nucleotide dataset, following the greedy algorithm. All data blocks were defined by gene types (each of the 12 PCGs and two rRNAs), and PCGs were divided by each one of the three codon positions. Partition schemes and best-fit models are provided in Table S1.

2.5. Phylogenetic analysis

We used two inference methods, Bayesian inference (BI) and Maximum Likelihood (ML), for phylogenetic reconstruction. BI on MrBayes 3.2.6 (Ronquist et al., 2012), was performed through the online CIPRES Science gateway with two independent runs using 100,000,000 Markov Chain Monte Carlo (MCMC) iterations, following the models suggested by the PartitionFinder, with four simultaneous chains, and sampling every 1000 generations. Support for nodes (Bayesian Posterior Probability = BPP) and parameter estimates were derived from a majority rule consensus of the last 5000 trees sampled after convergence. For the ML method we used the same partitions in IQ-TREE version 1.6.8 (Nguyen et al., 2015). Nodal support was

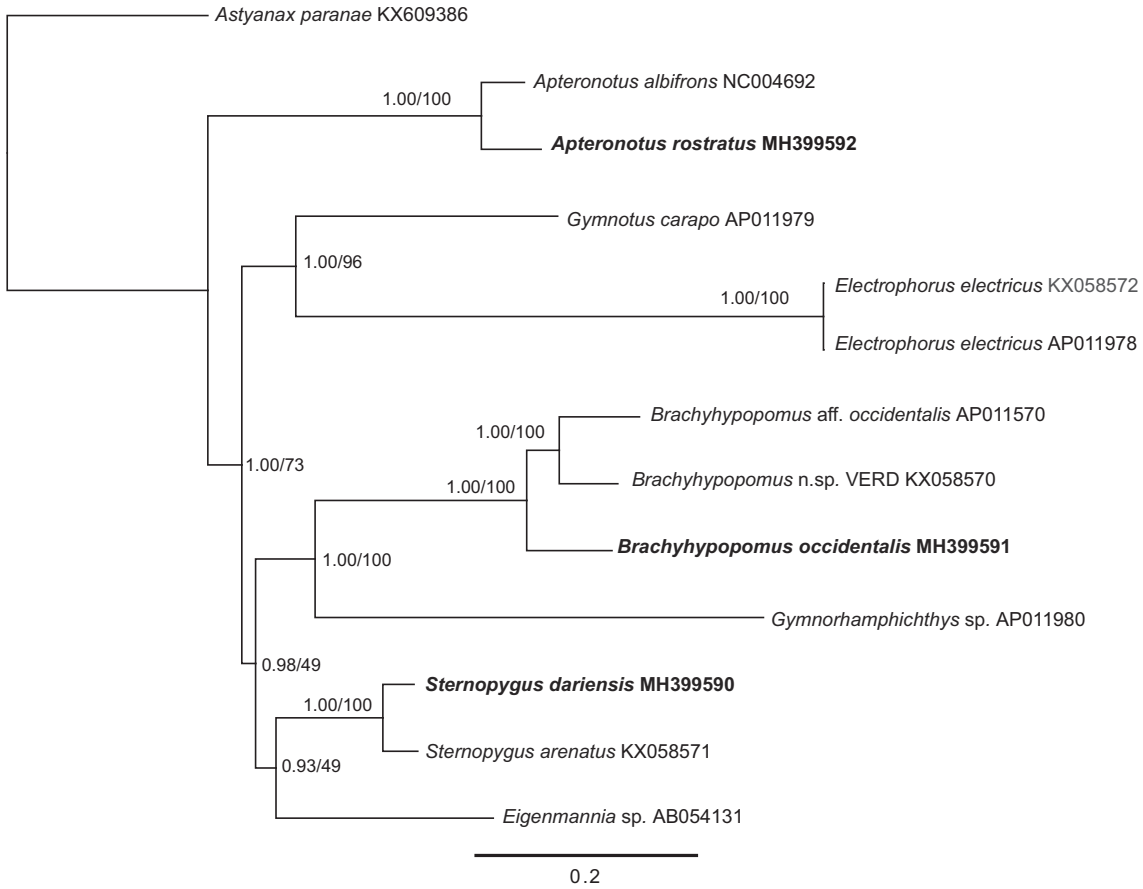

Fig. 2. Phylogenetic relationships of Gymnotiformes based on complete mitochondrial genomes. The alignment represents 12 protein-coding and 2 ribosomal RNA genes, concatenated. Numbers at nodes are statistical support values for MrBayes Bayesian posterior probability (BPP) and IQ-TREE bootstraps (BS). The lengths of the branches follow the phylogram of the BI tree.

Table 2
Genetic divergence between *Brachyhypopomus* individuals. The data represent differences in the number of base substitutions per site for both mitochondrial protein-coding genes and complete mitogenomes between *B. occidentalis* (MH399591; reported here) and the previously reported *B. occidentalis* (AP011570) and *B. n.*sp. VERD (KX058570). *P*-distance refers to an uncorrected pairwise distance.

DNA sequence	Alignment size (bp)	*p*-Distance (%)	
		AP011570	KX058570
ATP6	681	16.7	18.4
ATP8	159	12.6	13.8
COI	1551	14.8	14
COII	678	12.2	13.7
COIII	774	13.8	13.7
Cyt B	1128	14.5	13.3
ND1	957	16.5	16
ND2	1044	16.8	17.5
ND3	348	17.2	16.7
ND4	1380	15.7	15.8
ND4L	294	19.9	18.2
ND5	1833	16.3	15.9
ND6	516	15.6	18.1
Complete mtDNA	16,584	13.1	13.2

assessed through 10,000 ultrafast bootstraps (BS) (Hoang et al., 2018) and 1000 Shimodaira-Hasegawa-like approximate likelihood ratio test (Guindon et al., 2010) replicates. We also built a ML tree in RAxML v. 8.1.21 (Stamatakis, 2014) as implemented in the Cipres gateaway web server. For this analysis we used the GTR + Γ + I model with joint branch length optimization, and 1000 bootstrap samples to calculate support values. The resulting tree topologies were similar, but IQ-TREE consistently had higher bootstrap support. This difference is likely due to the fact that IQ-TREE algorithm allows for the specification of multiple substitution models, but the RAxML method is limited to the GTR family of models. For this reason, we used IQ-TREE for construction of the final ML tree.

2.6. Genetic divergence in Brachyhypopomus mitogenomes

To assess genetic distances among the three complete *Brachyhypopomus* mitogenomes, we estimated the proportion of nucleotide differences (uncorrected *p*-distance; Nei and Kumar, 2000). Standard errors were calculated using 500 bootstrap replicates in MEGA v7 (Kumar et al., 2016). We also performed a BLAST search using the 548 bp *COI* sequence fragment (e.g., the barcoding gene; Hebert et al., 2003). Finally, to confirm the species identity of available *Brachyhypopomus* mitogenomes, we built a phylogenetic tree using a dataset of 83 *CYTB* sequences representing 26 species previously reported by Crampton et al. (2016a), using RAxML under the fast bootstrap method with1000 replications. The resulting RAxML tree was used as input for species delimitation analysis via multi-rate Poisson tree processes (mPTP; Kapli et al., 2017), using the Markov chain Monte Carlo (MCMC) algorithm with two chains and the Likelihood Ratio Test set to 0.01. To assess the confidence of our ML species delimitation, we ran four MCMC chains of 50 million steps each. Overall support for our ML estimate was obtained by calculating average support values (ASV) across all MCMC runs, following Kapli et al. (2017). ASV values close to one indicate high support for the ML species delimitation. Finally, we

assessed convergence of each MCMC chain by calculating the average standard deviation of delimitation support values (ASDDSV). The ASDDSV is computed by averaging the standard deviation of per-node delimitation support values across all MCMC chains (Ronquist et al., 2012; Kapli et al., 2017). This quantity approaches zero when multiple MCMC runs converge on the same delimitation distribution.

3. Results and discussion

3.1. Mitogenome structure

Using both Trinity and the de novo assembly approach of NOVOPlasty, we recovered a single contig covering the full-length of the reference mitogenome for each of the three Gymnotiforms species. The three mitogenome sequences were similar in both methods, except that Trinity was unable to recover the control region of *A. rostratus*. This indicates that the software NOVOPlasty provides a better method for mitogenome assembly than Trinity. Mapping of the raw data to their respective Gymnotiforms mitogenomes resulted in deep, complete coverage of the genomes (Fig. 1). The read coverage varied within and among the different mitochondrial genomes (Fig. 1), but on average it was moderate, ranging from 39 to 43 reads per nucleotide. The mitogenome size was 16,631 bp for *S. dariensis* (GenBank accession no. MH399590), 16,540 bp for *B. occidentalis* (GenBank accession no. MH399591) and 17,093 bp for *A. rostratus* (GenBank accession no. MH399592). All three mitogenomes contained 2 ribosomal genes (12S and 16S), 22 tRNAs, 13 protein-coding genes (PCGs), as well as a control region (Table 1 and Fig. 1). Overall, the recovered circular mitogenomes were congruent with the gene count and organization of other Gymnotiformes mitogenomes (Nakatani et al., 2011; Lavoué et al., 2012; Elbassiouny et al., 2016).

The nucleotide composition showed a strand bias consistent with the strand asymmetry observed in other fishes (Cheng et al., 2012; Hao et al., 2016). Specifically, the A + T content was 58.6%, 54.1% and 57.1% for *A. rostratus*, *B. occidentalis* and *S. dariensis*, respectively. The average AT-skew was 0.08, ranging from 0.05 in *B. occidentalis* to 0.09 in *S. dariensis*. The average GC-skew was −0.32, ranging from −0.34 in *S. dariensis* to −0.28 in *B. occidentalis*. The 13 PCGs encompassed ~69% (11,440 bp) of the total mitogenome; twelve of which were on the forward strand, while ND6 was on the reverse strand. Overall, we found ~3791 codons, excluding stop codons, that were predicted for codon usage across the three mitogenomes.

The start codon in *S. dariensis* and *B. occidentalis* was a typical ATN codon, but the start codon in the *COI* gene was GTG for all three species, which is consistent with other fish mitogenomes (Satoh et al., 2016; Shi et al., 2016). However, *A. albifrons* mitogenome shows alternative start codons in three genes: GTG for *ATP8* and *ND6*, and ACG for *ND4L*. In fishes, the ACG start codon has been only reported in the *A. albifrons ND4L* gene (Satoh et al., 2016) and in the *ND1* gene of the perciform *Otolithes ruber* (Guo et al., 2017). The three species shared the TAA stop codon in 3 PCGs (*ND1, ND4L and ND5*), while AGA and AGG stop codons were present in *COI* (except in *B. occidentalis*, TAA) and *ND6* (except in *A. rostratus*, AGA), respectively. The remaining PCGs (*ND, COII, ATP6, COIII, ND3, ND4* and *CYTB*) had TAG, TAA or the incomplete stop codons TA/T, which are presumably completed during post- transcriptional polyadenylation (Ojala et al., 1981). This pattern of stop codons is also common in other fishes (Kim et al., 2006; Nakatani et al., 2011). In addition, there were the typical 22 tRNAs predicted by Mitofish and tRNAscan, with a length ranging from 66 bp to 75 bp and including two $tRNA^{Leu}$ and two $tRNA^{Ser}$. The two rRNA genes were located between $tRNA^{Phe}$ and $tRNA^{Leu}$ and were separated by $tRNA^{Val}$.

3.2. Non-coding regions, intergenic spacers and overlapping

We found small intergenic spacers (IGS) ranging in size from 1 to 60 bp, and totalling 58 bp in *S. dariensis*, 68 bp in *B. occidentalis* and 134 bp in *A. rostratus* (Table. 1). These IGS regions were similar across species and represent a common feature of Gymnotiformes. One of these spacers, with a size of 29 to 31 bp, represents the origin of L-strand replication (OL), and is located between $tRNA^{Asn}$ and $tRNA^{Cys}$. We also found a large IGS of 1402 bp that belongs to the Control Region (D-loop) in all three species. We also observed 8 gene overlaps with a total of 31 bp; the two longest of which contained 10 bp (between *ATP8* and *ATP6*) and 7 bp (between *ND4L* and *ND4*) (Table 1).

3.3. Novel COII/tRNA-Lys intergenic spacer in Apteronotus

We uncovered a novel 60 bp IGS located between *COII* and $tRNA^{Lys}$ in the mitogenome of *A. rostratus*. This IGS was supported by a sequence coverage of > 27× (Fig. S1). Interestingly, we also identified the same 60 bp IGS in the previously reported mitogenome of *A. albifrons* from South America (Saitoh et al., 2003; GenBank: AB054132). However, this IGS was not found in any of the other available gymnotiform mitogenomes (Fig. S2). The presence of a similar spacer in two *Apteronotus* mitogenomes that were sequenced independently suggests that this IGS represents a unique feature of the genus *Apteronotus*. To our knowledge, this is the first report of a long IGS occurring between the genes *COII* and $tRNA^{Lys}$ in the order Gymnotiformes. In other fishes, the presence of unique IGS has been reported between the genes $tRNA^{Thr}$ and $tRNA^{Pro}$ in Gadiformes (Bakke et al., 1999; Jørgensen et al., 2014), including walleye pollock (*Theragra chalcogramma*; Poulsen et al., 2013), the whiting (*Merlangius merlangus*) and haddock (*Melanogrammus aeglefinushiting* Roques et al., 2006).

Currently, the origin of these IGS and their apparent absence in other Gymnotiformes is not well understood. If we accept a basal phylogenetic position of *Apteronotus* (Fig. 2), one possibility is that purifying selection on non-coding regions (Rand, 1993) led to reduction in mitogenome size during the evolutionary history of Gymnotiformes. However, further work is necessary to determine the biological implications of these IGS in Gymnotiformes. Overall, we suggest that comparative studies of this unique mitogenomic feature across species could help elucidate the phylogenetic history of the group. In addition, further work should explore the use of these IGS as genetic markers for the genus *Apteronotus* or the entire Apteronotidae.

3.4. Phylogenetic relationships and taxonomic implications

Phylogenetic analyses based on BI and ML of 12 PCGs +2 rRNAs recovered similar tree topology. Nodal support was relatively high, and this was notably so for the phylogenetic tree based on the BI method. For the ML method, two nodes (*Sternopygus* and its relationship with *Brachyhypopomus* and *Gymnotus* showed the lowest support, but the position of the species within each clade remained identical for the two methods. However, further analyses with larger sample sizes are required to solve these nodes. Overall, however, several important features of the tree topology were shared across the analyses (Fig. 2). For instance, the traditional monophyly of Gymnotiformes was strongly supported, and the five recognized groups within Gymnotiformes (Apteronotidae, Sternopygidae, Gymnotidae, Hypopomidae and Rhamphichthydae) were recovered by both analyses (Fig. 2).

Furthermore, the family Apteronotidae was placed as the basal group to all the Gymnotiformes, in agreement with a recent mitogenomic study (Elbassiouny et al., 2016), but in contrast to previous phylogenic studies (Alves-Gomes et al., 1995; Tagliacollo et al., 2016). Our result was also consistent with a previous morphologically-based phylogeny by Triques (1993), suggesting a wave-type species as basal to the diversification of Gymnotiformes.

Finally, the monophyly of families and genera was also supported by both analyses. In particular, the sister relationship within Gymnotidae (*Gymnotus* and *Electrophorus*), the phylogenetic positions of the Hypopomidae (*Brachyhypopomus* species) and Rhamphichthydae

(*Gymnorhamphichthys*), and the monophyly of Sternopygidae (*Sternopygus dariensis, S. arenatus* and *Eigenmannia* sp.) was highly supported by both phylogenetic methods (BPP ≥ 0.93; BS ≥ 73, excluding Sternopygidae, Fig. 2), in agreement with previous taxonomic and phylogenetic studies (Triques, 1993; Albert, 2001; Arnegard et al., 2010; Elbassiouny et al., 2016; Tagliacollo et al., 2016). In addition, *S. dariensis* and *S. arenatus*, *A. rostratus* and *A. albifrons* as well as *B. occidentalis*, *B. verdii* (*B.* n.sp. VERD - GenBank KX058570) and *B. occidentalis* (GenBank AP011570) showed a monophyletic relationship with respect to each genus (Fig. 2). These results thus confirm the phylogenetic position of Central American electric fishes within Gymnotiformes as a whole (Arnegard et al., 2010; Elbassiouny et al., 2016; Tagliacollo et al., 2016).

Although our results are broadly consistent with previous phylogenetic analyses of Gymnotiformes, multiple aspects of the biogeography and diversification of this group remain to be explored. We suggest that harnessing mitogenomic data across the tree of life of Gymnotiformes is important to further our understanding of this diverse group of freshwater fishes.

3.5. *Brachyhypopomus occidentalis* phylogenetic relationships

We found over 13% sequence divergence between our complete *B. occidentalis* mitogenome, and the one conspecific mitogenome available in GenBank (Lavoué et al., 2012; AP011570). For individual genes, our analysis of *p*-distances between mitogenomes revealed values ranging from 12.2% in *COII* to 19.9% in *ND4L*. For PCGs, average divergence was 15.6% (Table 2). We also found 99.8% similarity between our sequence and two previously reported sequences of *COI* (548 bp fragments) for *B. occidentalis* from Panama (Accession no. KX766510 and KX766511; Crampton et al., 2016a, 2016b), but only 90.6% similarity with the *B. occidentalis* mitogenome reported by Lavoué et al. (2012; Accession no. AP011570). These results indicate that our mitogenome represents the correct sequence for *B. occidentalis*. Indeed, our mPTP species-delimitation analysis of 26 species of *Brachyhypopomus* placed the previously reported individual within the *B. bennetti* clade from South America rather than within *B. occidentalis* from Central America (Fig. S3).

Interestingly, the species-delimitation analysis recovered only 22 out of the 26 putative species of *B. occidentalis* reported by Crampton et al. (2016a, 2016b) (Fig. S3), and these results showed high average support values (ASV ≥ 0.94), with all MCMC chain converging on the same delimitation distributions (ASDDSV ≤ 0.000039). These results suggest that species diversity within *Brachyhypopomus* is sometimes overestimated, particularly when using a limited number of molecular markers. In other cases, however, species diversity might be underestimated, as suggested by the presence of two distinct geographic clades, corresponding to Middle America and the Pacific of South America.

Changes in the known species identity of *Brachyhypopomus* are likely due to several factors: (i) the presence of cryptic species within traditionally-recognized morphological species; (ii) the existence of geographical and/or population-level variation within a single species that has been traditionally considered as multiple species; (iii) the incomplete sampling of species across the range of genus; and (iv) perhaps the erroneous identification of individuals in the field. Thus, despite important efforts to study taxonomic diversity in Gymnotiformes (Crampton et al., 2016b; Hulen et al., 2005; Tagliacollo et al., 2016), our results suggest that further work is needed to elucidate the taxonomic status and species diversity in both *Brachyhypopomus* and Gymnotiformes in general.

Overall, our study expands our understanding of the evolution and structure of mitochondrial genomes in Central American freshwater fishes and generates novel molecular data that can be used to solve the taxonomic status as well as the phylogenetic history of Neotropical electric fishes. However, further research integrating morphological and novel molecular data is necessary to disentangle the patterns of species diversity in Gymnotiformes.

Disclosure statement

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the anonymous reviewers for valuable comments and suggestions. We also thank the Ecological and Evolutionary Genomics Laboratory (EEGL) at STRI, Panama, and the Biology Department at the University of Massachusetts Boston for support and training during the development of this project. This work was funded by the Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT, Panamá) in the form of an international internship to CA, and grants (No. ITE12-002, FID16-116) to LFD. Additional support was provided by Instituto para la Formación y Aprovechamiento de los Recursos Humanos (IFARHU-SENACYT) in the form of a doctoral fellowship to CA, and the University of Massachusetts Boston to LFD. JRL was also supported by the Sistema Nacional de Investigación (SNI;157-2017).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2018.11.045.

References

Aguilar, C., De Léon, L.F., Loaiza, J.R., McMillan, W.O., Miller, M.J., 2016. Extreme sequence divergence between mitochondrial genomes of two subspecies of white-breasted wood-wren (*Henicorhina leucosticta*, Cabanis, 1847) from western and Central Panamá. Mitochondrial DNA 27, 956–957.

Albert, J.S., 2001. Species diversity and phylogenetic systematics of American knifefishes (Gymnotiformes, Teleostei). Museum of Zoology, University of Michigan, Ann Arbor, MI.

Albert, J.S., Crampton, W.G.R., 2005. Diversity and phylogeny of neotropical electric fishes (Gymnotiformes). In: Bullock, T.H., Hopkins, C.D., Popper, A.N., Fay, R.R. (Eds.), Electroreception, Springer Handbook of Auditory Research. Springer New York, pp. 360–409.

Alda, F., Picq, S., De León, L.F., González, R., Walz, H., Bermingham, E., Krahe, R., 2013. First record of *Gymnotus henni* (Albert, Crampton and Maldonado, 2003) in Panama: Phylogenetic position and electric signal characterization. Check List 9, 655–659.

Alves-Gomes, J.A., Orti, G., Haygood, M., Heiligenberg, W., Meyer, A., 1995. Phylogenetic analysis of the south-American electric fishes (order Gymnotiformes) and the evolution of their electrogenic system – a synthesis based on morphology, electrophysiology, and mitochondrial sequence data. Mol. Biol. Evol. 12, 298–318.

Arnegard, M.E., Zwickl, D.J., Lu, Y., Zakon, H.H., 2010. Old gene duplication facilitates origin and diversification of an innovative communication system twice. Proc. Natl. Acad. Sci. U. S. A. 107, 22 172–22 177. https://doi.org/10.1073/pnas.1011803107.

Bakke, I., Shields, G.F., Johansen, S., 1999. Sequence characterization of a unique intergenic spacer in Gadiformes mitochondrial DNA. Mar. Biotechnol. 1, 411–0415.

Bermingham, E., Martin, A.P., 1998. Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. Mol. Ecol. 7, 499–517.

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

Bullock, T.H., Fay, R.R., Hopkins, C.D., Popper, A.N., 2005. Electroreception. Springer Handbook of Auditory Research. Springer, New York.

Cheng, J., Ma, G.Q., Song, N., Gao, T.X., 2012. Complete mitochondrial genome sequence of bighead croaker *Collichthys niveatus* (Perciformes, Sciaenidae): a mitogenomic perspective on the phylogenetic relationships of Pseudosciaeniae. Gene 491, 210–223.

Crampton, W.G.R., Lovejoy, N.R., Waddell, J.C., 2011. Reproductive character displacement and signal ontogeny in a sympatric assemblage of electric fish. Evolution 65, 1650–1666. https://doi.org/10.1111/j.1558-5646.2011.01245.x.

Crampton, W.G.R., de Santana, C.D., Waddell, J.C., Lovejoy, N.R., 2016a. Phylogenetic systematics, biogeography, and ecology of the electric fish genus *Brachyhypopomus* (Ostariophysi: Gymnotiformes). PLoS One 11, e0161680.

Crampton, W.G.R., de Santana, C.D., Waddell, J.C., Lovejoy, N.R., 2016b. A taxonomic revision of the Neotropical electric fish genus *Brachyhypopomus* (Ostariophysi: Gymnotiformes: Hypopomidae), with descriptions of 15 new species. Neotrop. Ichthyol. 14. https://doi.org/10.1590/1982-0224-20150146.

Dierckxsens, N., Mardulyn, P., Smits, G., 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucl. Acids Res. 45, e18.

Elbassiouny, A.A., Schott, R.K., Waddell, J.C., Kolmann, M.A., Lehmberg, E.S., Nynatten, A.V., Crampton, W.G.R., Chang, B.S.W., Lovejoy, N.R., 2016. Mitochondrial genomes

of the South American electric knifefishes (Order Gymnotiformes). In: Mitochondrial DNA Part B. 1. pp. 401–403.
Faircloth, B.C., 2013. Illumiprocessor: A Trimmomatic Wrapper for Parallel Adapter and Quality Trimming. (doi:10.6079).
Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.
Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of Ultraconserved Elements (UCEs). PLoS One 8, e65923.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Guo, C.-C., Liu, M., Lin, J.-J., Dai, F.-Q., 2017. Complete mitochondrial genome and the phylogenetic position of the tigertooth croaker *Otolithes ruber* (Perciformes: Sciaenidae). In: Mitochondrial DNA Part B. 2. pp. 132–133.
Hao, G., Wu, Q., Zhong, H., Zhou, Y., 2016. Complete mitochondrial genome of *Pterophyllum scalare* (Perciformes, Cichlidae). Mitochondrial DNA A 27, 4215–4216.
Hebert, P.D.N., Cywinska, A., Ball, S.L., Dewaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. Biol. Sci. 270, 313–321.
Hoang, D.T., Chernomor, O., Haeseler, Avon, Minh, B.Q., Le, S.V., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522.
Hulen, K.G., Crampton, W.G.R., Albert, J.S., 2005. Phylogenetic systematics and historical biogeography of the Neotropical electric fish *Sternopygus* (Teleostei: Gymnotiformes). Syst. Biodivers. 3, 407–432.
Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T.P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., Nishida, M., 2013. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Mol. Biol. Evol. 30, 2531–2540.
Jørgensen, T.E., Bakke, I., Ursvik, A., Andreassen, M., Moum, T., Johansen, S.D., 2014. An evolutionary preserved intergenic spacer in gadiform mitogenomes generates a long noncoding RNA. BMC Evol. Biol. 14, 182.
Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., Flouri, T., 2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics 33, 1630–1638.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kim, B.-C., Kang, T., Kim, M., Kim, C.-B., 2006. The complete mitogenome of *Rhodeus uyekii* (Cypriniformes, Cyprinidae). DNA Seq. 17, 181–186.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msw260.
Lavoué, S., Miya, M., Arnegard, M.E., Sullivan, J.P., Hopkins, C.D., Nishida, M., 2012. Comparable ages for the independent origins of electrogenesis in African and South American weakly electric fishes. PLoS One 7, e36287.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE), IEEE, pp. 1–8.
Moller, P., 1995. Electric Fishes: History and Behavior. Chapman & Hall, London.
Nakatani, M., Miya, M., Mabuchi, K., Saitoh, K., Nishida, M., 2011. Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. BMC Evol. Biol. 11, 177.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. 25371430.
Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470–474.
Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–358.
Picq, S., Alda, F., Krahe, R., Bermingham, E., 2014. Miocene and Pliocene colonization of the central American Isthmus by the weakly electric fish *Brachyhypopomus occidentalis* (Hypopomidae, Gymnotiformes). J. Biogeogr. 41, 1520–1532.
Poulsen, J.Y., Byrkjedal, I., Willassen, E., Rees, D., Takeshima, H., Satoh, T.P., Shinohara, G., Nishida, M., Miya, M., 2013. Mitogenomic sequences and evidence from unique gene rearrangements corroborate evolutionary relationships of myctophiformes (Neoteleostei). BMC Evol. Biol. 13, 111.
Rand, D.M., 1993. Endotherms, ectotherms, and mitochondrial genome-size variation. J. Mol. Evol. 37, 281–295.
Reis, R.E., Kullander, S.O., Ferraris, C.J., Junior, C., 2003. Checklist of the Freshwater Fishes of Central and South America. EDIPUCRS Porto Alegre Brasil Brasil.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Roques, S., Fox, C.J., Villasana, M.I., Rico, C., 2006. The complete mitochondrial genome of the whiting, *Merlangius merlangus* and the haddock, *Melanogrammus aeglefinus*: a detailed genomic comparison among closely related species of the Gadidae family. Gene 383, 12–23.
Saitoh, K., Miya, M., Inoue, J.G., Ishiguro, N.B., Nishida, M., 2003. Mitochondrial genomics of ostariophysan fishes: perspectives on phylogeny and biogeography. J. Mol. Evol. 56, 464–472. https://doi.org/10.1007/s00239-002-2417-y.
Satoh, T.P., Miya, M., Mabuchi, K., Nishida, M., 2016. Structure and variation of the mitochondrial genome of fishes. BMC Genomics 17, 719.
Shi, X., Tian, P., Lin, R., Huang, D., Wang, J., 2016. Characterization of the complete mitochondrial genome sequence of the globose head whiptail *Cetonurus globiceps* (Gadiformes: Macrouridae) and its phylogenetic analysis. PLoS One 11, e0153666.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Tagliacollo, V.A., Bernt, M.J., Craig, J.M., Oliveira, C., Albert, J.S., 2016. Model-based total evidence phylogeny of Neotropical electric knifefishes (Teleostei, Gymnotiformes). Mol. Phylogenet. Evol. 95, 20–33.
Triques, M.L., 1993. Filogenia dos generos de Gymnotiformes (Actinopterygii, Ostariophysi), com base em caracteres esqueleticos. In: Comun. Museu Ciencias PUCRS Serie Zool. 6. pp. 85–130.